STRATEGIC COMMUNICATION WITH LYING COSTS

NAVIN KARTIK

Columbia University and University of California, San Diego

E-mail address: nkartik@gmail.com.

Abstract. I study a model of strategic communication between an informed Sender and an uninformed Receiver. The Sender bears a cost of lying, or more broadly, of misrepresenting his private information. The main results show that inflated language naturally arises in this environment, where the Sender (almost) always claims to be a higher type than he would under complete information. Regardless of the intensity of lying cost, there is incomplete separation, with some pooling on the highest messages. The degree of language inflation and how much information is revealed depends upon the intensity of lying cost. The analysis delivers a framework to span a class of cheap talk and verifiable disclosure games, unifying the polar predictions they make under large conflicts of interest. I apply the model to highlight how the degree of manipulability of information can affect the tradeoff between delegation and communication.

Date: December 12, 2008 (First version: November 2003).

Key words and phrases. Cheap Talk, Signaling, Misreporting, Exaggeration, Disclosure, Persuasion, Delegation.

This paper subsumes portions of an earlier manuscript circulated as "Information Transmission with Almost-Cheap Talk." The other parts of that manuscript are now superseded by "Selecting Cheap-Talk Equilibria," with Ying Chen and Joel Sobel.

This research grew out of my Ph.D. dissertation, and I am indebted to my advisors, Doug Bernheim and Steve Tadelis, for their generous support and advice. I am also grateful to Vince Crawford and, especially, Joel Sobel for encouragement and numerous conversations. I thank David Ahn, Nageeb Ali, Yeon-Koo Che, Wouter Dessein, Peter Hammond, Cristóbal Huneeus, Maxim Ivanov, Jon Levin, Mikko Packalen, Ilya Segal, Jeroen Swinkels, Bob Wilson, a number of seminar and conference audiences, the Editor (Andrea Prat) and two anonymous referees for helpful comments. Chulyoung Kim provided able research assistance. I gratefully acknowledge financial support from a John M. Olin Summer Research Fellowship, a Humane Studies Fellowship, the National Science Foundation, and the Institute for Advanced Study at Princeton. I also thank the Institute for its hospitality.

1. Introduction

The strategic transmission of private information plays an important role in many areas of economics and political science. Most theoretical studies of direct communication can be classified into the following two categories. One approach, initiated by Grossman (1981) and Milgrom (1981), assumes that information is verifiable and agents can withhold information but not lie. These are referred to as games of "persuasion" or "verifiable disclosure." The second approach, pioneered by Crawford and Sobel (1982) and Green and Stokey (2007, circulated in 1981), assumes that information is unverifiable and an agent can lie arbitrarily without direct costs, typically referred to as games of "cheap talk."

These are two extremes in a continuum of possibilities. Rather than lying — synonymous in this paper with misrepresentation or misreporting — being completely costless or prohibitively costly, reality usually lies in between: individuals can and do misrepresent private information, but bear some cost in doing so. These costs arise for various reasons. First, there may be probabilistic ex-post state verification that results in penalties if misreporting is detected, such as random audits on taxpayers (e.g. Allingham and Sandmo, 1972). Second, there may be costs of manipulating information or falsifying a message, such as when a manager has to spend time "cooking the numbers" or otherwise expend costly resources in order to report higher profits (e.g. Lacker and Weinberg, 1989; Maggi and Rodriguez-Clare, 1995). Third, recent experimental work suggests that people have an intrinsic aversion to lying, even though messages are *prima facie* cheap talk (Gneezy, 2005; Hurkens and Kartik, 2008; Sánchez-Pagés and Vorsatz, 2006).

This paper studies the classic strategic communication setting of Crawford and Sobel (1982) (hereafter CS), with the innovation that some messages entail exogenous or direct lying costs for the Sender. I use the term "lying costs" to broadly capture a class of misrepresentation costs, such as those discussed above. To fix ideas, here is a simplified version of the model presented in Section 2: a Sender is privately informed about his type, $t \in [0,1]$, and must send a message to the Receiver, who then takes a decision which affects both players' payoffs. There is a conflict of interest because the Sender wants the Receiver to believe that his type is higher than it actually is. I suppose that for each Sender type, t, there is a set of messages, M_t, with the interpretation that any message $m \in M_t$ has the literal or exogenous meaning "my type is t." In this framework, the Sender is said to be lying when he sends a message $m \in M_{t'}$ for some $t' \neq t$, and the magnitude of a lie can be measured by the distance between his true type and the type he claims to be. When type t sends a message $m \in M_{t'}$, he bears a cost $kC(t',t)$, where k is a scalar that parameterizes the intensity of lying cost. If $k = 0$, messages are cheap talk, and the setting reduces to that of CS. This paper is concerned with $k > 0$. Assume that for any t, it is cheapest to tell the truth, and moreover, the marginal cost of a bigger lie is increasing in the magnitude of the lie.

Lying costs transform the CS cheap-talk model into one of costly signaling, although the induced structure is different from traditional "monotonic" signaling games (Cho and Sobel, 1990). The main results I provide concern the amount of information revealed by the Sender in equilibrium, and the language — the equilibrium mapping from types to messages — through which information is transmitted. For most of the paper, I focus on a tractable and appealing class of Perfect Bayesian equilibria, described in Section 3.[1] The structure of these equilibria is simple: low Sender types *separate*, revealing themselves through their equilibrium message, whereas high types segment into one or more connected *pools*. The equilibria feature *language inflation* in the sense that (almost) every type t sends a message $m \in M_{t'}$ for some $t' > t$. That is, the Sender (almost) always claims to be a higher type than he truly is, even though he suffers a cost of lying. Of course, the Receiver is not systematically deceived, because she recognizes that the Sender is inflating his messages and adjusts her decision accordingly. While this state of affairs may appear paradoxical at first glance, the key point is that if the Sender were to tell the truth (i.e. send a message $m \in M_t$ when his type is t), the Receiver would infer that his type is in fact *lower* than it actually is, since she expects messages to be inflated. There is thus an inescapable inefficiency for the Sender in equilibrium, reminiscent of unproductive education signaling in Spence (1973) or dissipative advertising in Milgrom and Roberts (1986a).

Language inflation is difficult to explain in a framework of pure cheap talk, due to the indeterminacy of language in any cheap-talk game. Yet, the phenomenon occurs in numerous settings of communication with conflicts of interest. For example, analysts at brokerage firms that have an underwriting relationship with a stock tend to provide inflated recommendations on that recommendation (e.g. Lin and McNichols, 1998; Michaely and Womack, 1999). The literature discusses two basic forces that determine an affiliated analyst's incentives: first, the existence of an underwriting relationship creates a pressure to induce clients to buy the stock; second, there are present compensation and future reputation rewards for more accurate recommendations. In terms of my model, the more an analyst's compensation is determined by accurate assessments, the higher is the intensity of lying costs, k.[2] I establish some intuitive and testable comparative statics with respect to this intensity. At the extreme case, as $k \to \infty$, equilibria converge to full separation with almost no language inflation, i.e. there is close to truth-telling. On the other hand, as $k \to 0$, there is a lot of pooling with extremely inflated language.

In Section 4, I analyze in detail a canonical parametrization of the model, based on CS's influential "uniform-quadratic" example. This permits closed form solutions for equilibria in a wide range of lying costs, and delivers additional comparative statics to those mentioned above. I then briefly apply these results to revisit the question of optimal allocation of authority in organizations between an informed agent and uninformed decision-maker. Dessein (2002) showed

[1] Although the possibility of other equilibria is not ruled out, these are the only ones that satisfy a refinement criterion proposed by Bernheim and Severinov (2003).

[2] Of course, there may be an additional component to lying costs stemming from regulatory concerns.

that a decision-maker prefers cheap-talk communication to unconstrained delegation if and only if the conflict of interest is sufficiently large. In contrast, I find that whenever the lying cost intensity is not too small in the leading example, communication can dominate delegation regardless of the conflict of interest. More generally, my results indicate that the tradeoff between communication and delegation will depend upon how costly it is for the agent to manipulate information.

Section 5 proposes an extension to span the cheap talk and verifiable disclosure approaches to strategic communication. As previously noted, the verifiable disclosure literature assumes that the Sender cannot lie about his type, but can withhold information. Not being able to lie may be viewed as the limit of my model as $k \to \infty$. To capture the ability to withhold information, I augment the model so that the Sender can make vague statements about his type, including being silent altogether at no cost. The main results are extended to this setting. One consequence is that when the conflict of interest between Sender and Receiver is large, the current framework unifies the dichotomous results of *unraveling* (full information revelation) in verifiable disclosure games and *babbling* (no information revelation) in cheap-talk games.

To summarize: the analysis here shows how lying costs can significantly affect the outcome of strategic information transmission. Consequently, direct communication may have a greater scope to reduce the costs of asymmetric-information-based incentive problems than otherwise suggested. This theoretical finding is consistent with experimental evidence, where "overcommunication" is often documented (e.g. Cai and Wang, 2006). Section 6 concludes with some comments on economic implications and potential applications of the current framework.

1.1. **Related Literature.** This paper contributes to the literatures on cheap talk, costly signaling, and verifiable disclosure. Most closely related to my work are the following contemporaneous papers: Chen (2007), Chen et al. (2008), Kartik et al. (2007), and Ottaviani and Squintani (2006). I now discuss the connections in some detail, followed by a briefer mention of other related work.

Kartik et al. (2007) also study the notion of costly lying and language inflation. However, the focus and analysis is quite different, because they study an *unbounded* type space setting. They prove the existence of separating or fully-revealing equilibria for arbitrarily small intensity of lying cost (and other costly signaling structures).[3] This result relies essentially on the unboundedness of the type space, because the Sender is then able to inflate his message without restraint. In contrast, in this paper I study the canonical *bounded* type space model of CS, and

[3]Austen-Smith and Banks (2000) analyze a model of cheap talk and costly signaling, where the costly signal takes the form of "burned money." That is, the cost of the signaling instrument does not vary with the Sender's type. Their focus is on how the presence of this additional instrument can enlarge the set of CS equilibria, with a key result that so long as the Sender's ability to impose costs on himself is sufficiently large, separating equilibria exist. Kartik (2007) shows that as the ability to burn money shrinks to 0, the set of equilibria in Austen-Smith and Banks (2000) converges to the entire set of CS equilibria. Kartik (2005) discusses the differences between burned money and costly lying.

show that separating equilibria do not exist regardless of the lying cost intensity.[4] Intuitively, given that messages must be inflated, the upper bound on the type space acts as a barrier to full separation. Consequently, the attention here is necessarily on partially-pooling equilibria. Aside from requiring a more complex analysis, this also permits me to address issues that cannot be meaningfully answered in an unbounded type space model, such as large conflicts of interest and comparative statics of partial-separation.

In a bounded type space, Ottaviani and Squintani (2006) study the "uniform-quadratic" specification of CS. They introduce the possibility that the Receiver may be naive or non-strategic. This transforms the game from one of cheap talk into a costly signaling structure that can be subsumed by the current framework (see Section 2.1). For some parameterizations, they construct an equilibrium that has similar features to the equilibria studied here. Their parametric restrictions require the conflict of interest between Sender and Receiver to be sufficiently large relative to other parameters, which is not required here. Ottaviani and Squintani (2006) provide some comparative statics that are complementary to those in this paper. There is no analog in their work to the analysis I undertake in Section 5 about withholding information.

Chen (2007) furthers the non-strategic approach by modeling not only Receiver naivety, but also Sender naivety or mechanical honesty. Specifically, she posits that with some probability the Sender just reports his type truthfully, and with some independent probability the Receiver takes the message to be truthful. Using the uniform-quadratic specification of CS, she shows that there is a unique equilibrium in the cheap-talk extension (Manelli, 1996) of her model that satisfies a monotonicity requirement. Language is inflated in her equilibrium, for reasons similar to this paper, since Receiver naivety induces a particular form of lying costs. A benefit of assuming Sender naivety is that it ensures that all messages are used in equilibrium, obviating the need to consider off-the-equilibrium-path beliefs.[5] A strength of the approach here is that both the underlying cheap-talk model and the costly signaling structure are more general.

Chen et al. (2008) propose a refinement called *No Incentive to Separate* (NITS) for cheap-talk games like CS. Among other justifications for the criterion, they show that only CS equilibria satisfying NITS can be the limit of certain equilibria in a class of games with lying costs, as the intensity of lying costs shrink to 0. Although their model does not strictly subsume the current setting, their arguments can be adapted. The analysis here is entirely complementary, because I study arbitrary intensities of lying cost, rather than just the cheap-talk limit as the costs vanish. *Inter alia*, this paper provides an explicit characterization of a set of equilibria for any cost intensity.

[4]This statement requires some caveats when the setting is more general than that described in this introduction; see Theorem 1 and fn. 10.

[5]The specification of Sender naivety imposes particular restrictions on beliefs when a message is observed that is not sent by a strategic Sender. In effect, it substitutes for restrictions on off-the-equilibrium-path beliefs in a fully rational model.

With regards to the broader signaling literature, I build upon techniques developed by Mailath (1987) to study costly signaling with a continuum of types. Whereas his analysis concerns separating equilibria, the main results in this paper concern partially-pooling equilibria, as already noted. Cho and Sobel (1990) were the first to derive incomplete separation in the manner obtained here: pooling at the top of the type space using the highest messages and separation at the bottom. Their analysis is for a class of signaling games that have a finite type space and satisfy a single-crossing property that does not apply to the current setting. A more closely related model is that of Bernheim and Severinov (2003). Although they study a topic in public finance and the issues we each focus on are rather different, the formal structures have similarities. My equilibrium characterization builds upon their work.

Lastly, I should also briefly mention two other literatures that touch upon related issues. First, there is a small literature on contracting and mechanism design with misrepresentation costs, initiated by Green and Laffont (1986). Generally, the costs are taken to be either zero or infinite, but see Deneckere and Severinov (2007) for an exception. Second, there are some papers in accounting that propose signaling theories of earnings management, e.g. Stein (1989). They also motivate costly misreporting for technological or legal reasons, but to my knowledge only consider unbounded type spaces and special functional forms.

2. The Model

There are two players, a Sender (S) and a Receiver (R). The Sender has private information summarized by his *type* $t \in T = [0,1]$, which is drawn from a differentiable probability distribution $F(t)$, with strictly positive density for all $t \in T$. After privately observing his type t, the Sender sends the Receiver a *message*, $m \in M$, where M is a Borel space of available messages. After observing the message, R takes an *action*, $a \in \mathbb{R}$.

I assume that $M = \bigcup_t M_t$, with $|M_t| = \infty$ for all t and $M_t \cap M_{t'} = \emptyset$ if $t \neq t'$. The interpretation is that any $m \in M_t$ carries the literal or exogenous meaning "my type is t." Since $\{M_t\}_{t \in T}$ is a partition of M, any m can be associated with a unique t such that $m \in M_t$; denote this mapping by $\Psi(m)$. Thus, if $\Psi(m) > \Psi(m')$, the Sender is claiming in literal terms to be a higher type when he sends m rather than m', and I will say that m is higher or larger than m' (only weakly, if $\Psi(m) \geq \Psi(m')$).

The payoff for R is given by $U^R(a,t)$, and the payoff for S is given by $U^S(a,t) - kC(\nu(\Psi(m)),t)$. Here, $U^R : \mathbb{R} \times T \to \mathbb{R}$ and $U^S : \mathbb{R} \times T \to \mathbb{R}$ represent the two players' preferences over type-action pairs, whereas $C : \mathbb{R} \times T \to \mathbb{R}$ is a lying cost function, $k > 0$ is a scalar that parameterizes its intensity, and $\nu : T \to \mathbb{R}$ is a function that will be further discussed below. Note that m is payoff-relevant only for S. All aspects of the game except the value of t are common knowledge.

I now introduce two sets of assumptions on players' preferences. The first set is identical to Crawford and Sobel (1982), while the latter defines the costs of lying.

Preferences over actions: For $i \in \{R, S\}$, the function $U^i(\cdot,\cdot)$ is twice continuously differentiable on $\mathbb{R} \times T$.[6] Using subscripts to denote derivatives as usual, $U^i_{11} < 0 < U^i_{12}$ for $i \in \{R, S\}$, so that both players' payoffs are concave in R's action and supermodular in (a, t). For any t, there exists $a^R(t)$ and $a^S(t)$ respectively such that $U^R_1\left(a^R(t), t\right) = U^S_1\left(a^S(t), t\right) = 0$, 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. These assumptions imply that for $i \in \{R, S\}$, $\frac{da^i(\cdot)}{dt} > 0$, i.e. both players prefer higher actions when the Sender's type is higher.

Preferences over messages: The function $\nu(\cdot)$ is continuously differentiable on T and strictly increasing. The function $C(\cdot,\cdot)$ is twice continuously differentiable on $\mathbb{R} \times T$, with $C_{11} > 0 > C_{12}$. These assumptions imply that there is a continuous and weakly increasing function, $r^S : T \to T$, such that for any t, $r^S(t) = \arg\min_{t' \in T} C(\nu(t'), t)$. In words, $r^S(t)$ is the type that t would claim to be in order to minimize message costs. Equivalently, any $m \in M_{r^S(t)}$ is a cost-minimizing message for t. I assume that $r^S(1) = 1$, but nothing else about the function r^S. The assumptions on C and ν imply that when $\Psi(m)$ is farther from $r^S(t)$, m is more costly for t, and moreover, the marginal change in cost from claiming to be a slightly higher type than $\Psi(m)$ is larger in absolute value.

2.1. **Discussion.** Here are some comments on the model's assumptions and interpretations.

Message space. As mentioned, each message is associated with some type, with the interpretation that every statement the Sender can make has a literal meaning about the exact value of his type.[7] An alternative interpretation is that the Sender must submit some evidence or data to the Receiver, and his type represents the true realization of this evidence, which is costly to falsify. The assumption that each M_t contains an infinite number of messages may be viewed as assuming a *rich language*: there are a large number of ways in which the Sender can make a statement with the literal content that his type is t, for example, "my type is t," or "you should believe that my type is t," and so forth.[8] It is worth noting that a special case of the model is when the Sender can make a two-dimensional statement about his type: on one dimension he reports an exact value about his type and suffers a cost of misreporting the truth, while on the other, he just makes a cheap-talk statement from some arbitrary (infinite) set. For example, firms can

[6] More precisely, there exists an open set T' containing T and two twice continuously differentiable functions on $\mathbb{R} \times T'$ which are respectively identical to U^S and U^R on $\mathbb{R} \times T$. Analogous statements apply to subsequent assumptions about differentiability of ν and C.

[7] Section 5 shows how the model can be extended to cases where the Sender is permitted to make "vague" statements about his type (e.g., "my type lies in $[0.1, 0.3]$").

[8] An infinite number of messages in each M_t is a stronger assumption than necessary; a large enough finite number would always do. Some of the results only require $|M_t| \geq 1$.

take costly actions to manipulate earnings reports that must be submitted, but they can also simultaneously provide costless explanations or justifications.

Lying costs. If $k = 0$, the model reduces to the cheap-talk model of CS. Since $k > 0$, messages are discriminatory signals in the sense that the cost of a message varies with the Sender's type. An important example to keep in mind throughout the analysis is as follows: $\nu(t) = t$, and $C_1(t,t) = 0$ for all t (e.g., $C(x,t) = (x-t)^2$). In this case, $r^S(t) = t$, i.e. it is cheapest to tell the literal truth.

The assumption $r^S(1) = 1$ means that the highest type minimizes cost by claiming to be itself. Obviously, this assumption is satisfied when it is cheapest for each type to tell the truth. It also holds in an application discussed below, even though truth-telling is not cost-minimizing for all types there. In any case, the assumption is made primarily for expositional convenience; all the analysis can be generalized to settings where $r^S(1) < 1$, as pointed out in fn. 10.

Let me emphasize that the Sender's costs of lying in this model depend only upon his type and the message he uses. Plainly, this is appropriate when the costs stem from potential auditing penalties or technological/opportunity costs of falsifying information. In terms of capturing psychological aversion to lying, the interpretation is that the Sender is averse to making statements that would be considered untruthful under a literal or otherwise exogenous interpretation of statements.[9] The assumption that costs are convex in the type that the Sender claims to be (holding fixed his true type) is a substantive but natural one. For example, it would arise if there is a fixed probability of misreporting being detected and penalties are convex in the magnitude of misreporting.

Naive Receivers. Admitting the possibility that the Receiver may be naive or credulous can induce costs on the Sender that are similar to exogenous lying costs. To illustrate, suppose as in Chen (2007) or Ottaviani and Squintani (2006) that messages are cheap talk, but with probability $q \in (0,1)$, the Receiver is (perceived to be) credulous, in which case she takes message m at face value, responding with action $a^R(\Psi(m))$. Now, when the Sender is of type t, the expected utility from sending message m and receiving action a from a strategic Receiver is given by

$$(1-q)U^S(a,t) + qU^S(a^R(\Psi(m)),t).$$

Setting $k = \frac{q}{1-q}$, $\nu(t) = a^R(t)$, and $C(x,t) = -U^S(x,t)$, the above payoff expression can be normalized and rewritten as $U^S(a,t) - kC(\nu(\Psi(m)),t)$. The properties of U^S and U^R imply that C and ν thus defined satisfy my assumptions. Note that in this case,

$$r^S(t) = \begin{cases} (a^R)^{-1}(a^S(t)) & \text{if } a^S(t) < a^R(1) \\ 1 & \text{if } a^S(t) \geq a^R(1), \end{cases}$$

[9]One might suggest that psychological costs of lying should take into account the Receiver's endogenous interpretation of a message. While this is interesting, I leave it for future research.

and is therefore not strictly increasing. Indeed, if the conflict of interest between Sender and Receiver is sufficiently large ($a^S(0) \geq a^R(1)$), r^S is a constant function. Nevertheless, r^S is weakly increasing and satisfies $r^S(1) = 1$.

2.2. **Strategies and Equilibrium.** A pure strategy for the Sender strategy is a (measurable) function $\mu : T \to M$, so that $\mu(t)$ is the message sent by type t. Denote the posterior beliefs of the Receiver given a message m by the cumulative distribution $G(t \mid m)$. A pure strategy for the Receiver is a function $\alpha : M \to \mathbb{R}$, so that $\alpha(m)$ is the action taken when message m is observed. The solution concept is Perfect Bayesian Equilibrium (PBE), which requires the Sender to maximize utility for each type given the Receiver's strategy, the Receiver to maximize utility given his beliefs after every message, and beliefs to satisfy Bayes' rule wherever it is well-defined.

For most of the analysis, I only consider pure strategy PBE, which are referred to as just "equilibria" hereafter. Moreover, I typically restrict attention to equilibria where the Sender's strategy is weakly increasing, in the sense that if $t > t'$, then $\Psi(\mu(t)) \geq \Psi(\mu(t'))$. This property is called *message monotonicity*, and an equilibrium which satisfies the property is a *monotone equilibrium*. In a monotone equilibrium, the Sender makes a weakly higher claim about his type when his type is truly higher. Attention to such equilibria seems reasonable because higher types intrinsically prefer higher messages due to the cost of lying structure (formally, $r^S(t)$ is weakly increasing).

3. Signaling and Inflated Language

3.1. **Impossibility of Full Separation.** The basic signaling problem is that the Sender wants the Receiver to infer that he is a higher type than he actually is. Say that a type t is *separating* if $\{t' : \mu(t') = \mu(t)\} = \{t\}$, i.e. the message sent by t reveals that the Sender's type is t. A *separating equilibrium* is one where all types are separating. Naturally, the first question is whether such equilibria exist.

It is generally more convenient to work with the mapping $\rho := \Psi \circ \mu$ directly, rather than just μ. Note that $\rho : T \to T$, and $\rho(t)$ is interpreted as the type that t claims to be.

Definition 1. A type t is said to be using *inflated language* if $\rho(t) > r^S(t)$. A type t is *telling the truth* if $\rho(t) = r^S(t)$.

The above definition is appropriate because it compares what a type claims to be in equilibrium relative to what it would claim under complete information, i.e. if the Receiver knew the Sender's type, which is $r^S(\cdot)$ by definition. Thus, *truth-telling* refers to what a type would say under complete information; whether it satisfies $\rho(t) = t$ depends on the properties of C and ν. Since under complete information the Sender would play a strategy such that $\rho(\cdot) = r^S(\cdot)$,

any deviation in the equilibrium $\rho(\cdot)$ from $r^S(\cdot)$ stems from the signaling problem. The following result identifies a key property of ρ in any separating region of types.

Lemma 1. *If types (t_l, t_h) are separating in a monotone equilibrium, then for each $t \in (t_l, t_h)$, $\rho(t) > r^S(t)$ and*

$$\rho'(t) = \frac{U_1^S\left(a^R(t), t\right) \frac{da^R}{dt}(t)}{kC_1\left(\nu(\rho(t)), t\right) \nu'(\rho(t))}. \tag{DE}$$

(Note: all proofs are in Appendix A.) The Lemma says that in a monotone equilibrium, ρ must solve (DE) on any open interval of types that are separating. This is straightforward *given* differentiability of ρ, since in that case, (DE) is the first order condition of the following maximization problem for each $t \in (t_l, t_h)$:

$$\arg\max_{\tilde{t} \in (t_l, t_h)} U^S(a^R(\tilde{t}), t) - kC(\nu(\rho(\tilde{t})), t). \tag{1}$$

The proof that ρ must be differentiable is based on Mailath (1987). Observe that the sign of ρ' in (DE) is determined by the sign of $C_1(\cdot, \cdot)$, since all the other terms on the right hand side are positive. In turn, the sign of $C_1(\nu(\rho(t), t))$ depends on whether $\rho(t)$ is greater or less than $r^S(t)$. Since a monotone equilibrium requires ρ to be weakly increasing, (DE) implies that *all types in the interior of a separating interval must be using inflated language*, i.e. $\rho(t) > r^S(t)$.

Consider now whether there can be complete separation in a monotone equilibrium. If so, Lemma 1 implies that all types in $(0, 1)$ must be using inflated langauge. However, since $r^S(1) = 1$, the highest type cannot be using inflated language. This intuition underlies the next result.

Theorem 1. *There is no separating monotone equilibrium.*

The formal logic is as follows: a separating monotone equilibrium requires a continuous function to solve an initial value problem defined by the differential equation (DE) together with some boundary condition $\rho(0) \geq r^S(0)$. The proof of Theorem 1 shows that there is no solution on the domain $t \in [0, 1]$, regardless of the choice of boundary condition. Intuitively, each type must inflate its claim in order to separate itself from lower types, but one eventually "runs out" of claims that can be made. More precisely, the solution to the initial value problem hits the upper bound of 1 at some $t < 1$.[10]

[10]Here is where the assumption that $r^S(1) = 1$ simplifies matters. If $r^S(1) < 1$, then it may be possible to have a separating monotone equilibrium, depending on other parameters; in particular, for any $r^S(1) < 1$, then will be a separating monotone equilibrium if k is large enough. However, the Theorem can be generalized to show that even in these cases, a separating monotone equilibrium will not exist if k is not too large. Intuitively, as k decreases, the slope of the solution to the relevant initial value problem increases, such that one eventually runs out of messages again. It can also be shown that there is no separating equilibrium in the class of *all* PBE, including non-monotone equilibria in pure or mixed strategies, if either of two conditions hold: $r^S(0) = 0$ (which is satisfied when truth-telling is cost-minimizing for all types) or k is sufficiently small.

Theorem 1 contrasts with the main result of Kartik et al. (2007), who show that when the type space is not bounded above, a separating monotone equilibrium exists for any $k > 0$. The reason is that when there is no upper bound on the type space, the solution to the differential equation (DE) can be extended over the whole domain.[11]

3.2. **Pooling Equilibria.** Theorem 1 implies that any monotone equilibrium involves some pooling. The logic that the Sender eventually runs out of types to mimic for separation suggests a class of pooling equilibria to study, where *low types separate and high types pool on the highest messages*. In what follows, I first define this class precisely, then discuss the structure and existence of such equilibria, followed by a justification for focusing on this class.

Let a *separating function* be a continuous and increasing function that solves (DE) with the initial condition $\rho(0) = r^S(0)$. This choice of initial condition is motivated by the usual "Riley condition" of least costly separation in signaling games. The Appendix (Lemma A.1) establishes that there is a unique separating function, denoted ρ^* hereafter, which is well-defined on some interval $[0, \bar{t}]$, where $\bar{t} < 1$ is such that $\rho^*(\bar{t}) = 1$.

Definition 2. A Sender's strategy μ is a LSHP (Low types Separate and High types Pool) strategy if there exists $\underline{t} \in [0, \bar{t}]$ such that:

(1) for all $t < \underline{t}$, $\mu(t) \in M_{\rho^*(t)}$ (i.e., $\rho(t) = \rho^*(t)$),
(2) for all $t \geq \underline{t}$, $\mu(t) \in M_1$ (i.e., $\rho(t) = 1$).

An equilibrium (μ, α) is a LSHP equilibrium if μ is an LSHP strategy.

A LSHP equilibrium features a *cutoff type*, $\underline{t}$, such that any type $t < \underline{t}$ separates by playing $\rho^*(t)$, whereas all types above $\underline{t}$ send messages in M_1. The structure of such equilibria is elucidated by several observations. A key point is that the following indifference condition must hold if $\underline{t} > 0$:

$$U^S(\alpha(\mu(\underline{t})), \underline{t}) - kC(\nu(1), \underline{t}) = U^S(a^R(\underline{t}), \underline{t}) - kC(\nu(\rho^*(\underline{t})), \underline{t}). \tag{2}$$

That is, type $\underline{t} > 0$ must be indifferent between sending $\mu(\underline{t}) \in M_1$ and inducing $\alpha(\mu(\underline{t}))$ versus sending any $m \in M_{\rho^*(\underline{t})}$ and inducing $a^R(\underline{t})$. This follows from the fact that any $t < \underline{t}$ plays $\rho^*(t)$ and induces action $a^R(t)$, combined with the equilibrium requirement that type $\underline{t}$ should not prefer to mimic some $t < \underline{t}$, and conversely, no $t < \underline{t}$ should prefer to mimic $\underline{t}$. Equation (2)

[11] It is also appropriate here to compare with Bernheim and Severinov (2003). Translating their model — specifically, their main case of "imitation towards the center" — to the current notation, separating equilibria exist in their setting whenever k is large enough. The reason is that they have different assumptions on U^S and U^R. In particular, they assume that $a^S(t) = a^R(t)$ for $t \in \{0, \frac{1}{2}, 1\}$. This contrasts with the current setting, where $a^S(t) > a^R(t)$ for all t. Moreover, in their model, the signal space is constrained, so that sufficiently low (resp. high) types necessarily inflate (resp. deflate) language. These differences have subtle but significant consequences on the analysis.

implies that for any $\underline{t} > 0$, there is a unique $\alpha(\mu(\underline{t}))$. This imposes a restriction on which types can be sending $\mu(\underline{t})$, because $\alpha(\mu(\underline{t}))$ must be optimal for the Receiver against $\mu(\underline{t})$.

Since $\mu(t) \in M_1$ for all $t \geq \underline{t}$, it is equally expensive in terms of direct message costs for a given type to mimic any $t \geq \underline{t}$. Consequently, the cheap-talk constraints of CS apply within $[\underline{t}, 1]$. To state this precisely, define (with an abuse of notation), for any $t'' < t'$,

$$a^R(t'', t') = \arg\max_a \int_{t''}^{t'} U^R(a, t)\, dF(t)$$

as the optimal action for the Receiver if the only information she has is that the Sender's type lies in $[t'', t']$. Given a cutoff type $\underline{t} \geq 0$, let a *partial-partition* refer to partition of $[\underline{t}, 1]$ into intervals, denoted by a strictly increasing sequence, $\langle t_0 = \underline{t}, t_1, \ldots, t_J = 1\rangle$. CS's logic dictates that there must be a partial-partition such that for all $j \in \{1, \ldots, J-1\}$,

$$U^S(a^R(t_{j-1}, t_j), t_j) - U^S(a^R(t_j, t_{j+1}), t_j) = 0. \tag{3}$$

For $j = 1, \ldots, J$, every type $t \in (t_{j-1}, t_j)$ sends the same message, call it $m_j \in M_1$, and induces the same action, $a^R(t_{j-1}, t_j)$. Of course, $m_j \neq m_k$ for $k \neq j$, and a boundary type t_j for $j = 1, \ldots, J-1$ can send either m_j or m_{j+1}, while type 1 must send m_J.

The above discussion pins down the Sender's strategy in any LSHP equilibrium with cutoff type $\underline{t} > 0$, modulo payoff-irrelevant permutations of messages that carry the same literal meaning. It is important to underscore that although all types above $\underline{t}$ are incurring the same direct message cost ($\rho(t) = 1$ for all $t \geq \underline{t}$), they need not be sending the same message. Thus, there can be *multiple pools* in an LSHP equilibrium, but all pools use messages in M_1.[12] The possibility of sustaining multiple pools with messages of the same literal meaning is ensured by the rich language assumption ($|M_t|$ is large for each t), and plays a key role in assuring existence of LSHP equilibrium when costs of lying are small relative to the conflict of interest.

On the Receiver's side, it is clear what she must play on the equilibrium path: $\alpha(\mu(t)) = a^R(t)$ for all $t < \underline{t}$, and $\alpha(m_j) = a^R(t_{j-1}, t_j)$ for any m_j such that $\mu(t) = m_j$ for all $t \in (t_{j-1}, t_j)$. As usual, optimality of the Sender's strategy places some constraints on what responses the Receiver can take off the equilibrium path.

The preceding remarks have identified a set of necessary conditions for a LSHP equilibrium with $\underline{t} > 0$. The following result establishes that they are also sufficient, and provides an existence result taking into account the possibility that there may be LSHP equilibria where no types separate ($\underline{t} = 0$).

[12]Throughout, a pool refers to a maximal set of types that use the same message, and thus induce the same action from the Receiver.

Theorem 2. *In any LSHP equilibrium, there is a cutoff type, $\underline{t} \in [0, \overline{t}]$, and a partial-partition, $\langle t_0 = \underline{t}, t_1, \ldots, t_J = 1 \rangle$, such that*

$$U^S(a^R(t_{j-1}, t_j), t_j) - U^S(a^R(t_j, t_{j+1}), t_j) = 0 \qquad \forall j \in \{1, \ldots, J-1\}, \tag{4}$$

$$U^S(a^R(\underline{t}, t_1), \underline{t}) - kC(\nu(1), \underline{t}) = U^S(a^R(\underline{t}), \underline{t}) - kC(\nu(\rho^*(\underline{t})), \underline{t}) \quad \text{if } \underline{t} > 0. \tag{5}$$

Conversely, given any cutoff type and partial-partition that satisfy (4), (5), and

$$U^S(a^R(\underline{t}, t_1), 0) - kC(\nu(1), 0) \geq U^S(a^R(0), 0) - kC(\nu(r^S(0)), 0) \quad \text{if } \underline{t} = 0, \tag{6}$$

there is a corresponding LSHP equilibrium.

For any $k > 0$, there is an LSHP equilibrium that satisfies (4)–(6). If $r^S(0) < r^S(1)$ and k is sufficiently large, there is an LSHP equilibrium with $\underline{t} > 0$.

Inequality (6) says that if the cutoff type is 0, then type 0 weakly prefers pooling with $[0, t_1]$ by claiming to be type 1 to revealing itself with a least-costly message. Existence is established by constructively showing that (4), (5), and (6) can jointly be satisfied.[13] The last part of the Theorem is intuitive: it says that so long as not all types have the same cost-minimizing messages, if the intensity of lying cost is sufficiently large, then at least some types can be separating in equilibrium.

Figure 1 illustrates a LSHP equilibrium. It is drawn for a case where $a^R(t) = r^S(t) = t$, so that sequential rationality requires the Receiver to only take actions in $[0, 1]$. The Sender's strategy is represented by the solid curve, which is $\rho(t)$. The separating function, ρ^*, coincides with ρ on $[0, \underline{t})$, and then extends as the dotted curve up to $\overline{t}$. There are two pools in this equilibrium: every $t \in [\underline{t}, t_1]$ plays $\mu(t) = m_1 \in M_1$, whereas every $t \in (t_1, 1]$ plays $\mu(t) = m_2 \in M_1 \setminus \{m_1\}$. The Receiver's strategy is depicted in the dashed blue line, which shows the correspondence $\beta : T \rightrightarrows \mathbb{R}$ defined by $\beta(t) = \cup_{m \in M_t} \alpha(m)$. That is, $\beta(t)$ is the set of all actions taken by the Receiver in response to messages where the Sender claims to be type t. Note that up until $\rho(\underline{t})$, β is is the mirror image of ρ around the 45° line, which verifies that for the region of separating types, the Receiver inverts the Sender's message. Since there are two pools, $|\beta(1)| = 2$.

Since the separating function ρ^* has $\rho^*(t) > r^S(t)$ for all $t \in (0, \overline{t}]$, *every LSHP equilibrium displays inflated language*, so long as $r^S(0) < 1$. The inflation of language is inefficient and costly for every type $t < \underline{t}$, since such types are revealing themselves in equilibrium, yet not minimizing their message costs. Telling the truth — or more generally, reducing the magnitude of inflation — is not profitable, however, because it would lead to adverse inferences from the Receiver, who expects the equilibrium degree of language inflation and thus rationally deflates (seen in Figure 1 by $\beta(\cdot)$ being strictly below the 45° line).

[13] Although (6) is not guaranteed to hold in an arbitrary LSHP equilibrium, one can show that it is necessary if the Receiver's strategy is to be weakly increasing in the sense that $\alpha(m) \geq \alpha(m')$ if $\Psi(m) \geq \Psi(m')$, a property that is appealing given the monotonicity of the Sender's strategy.

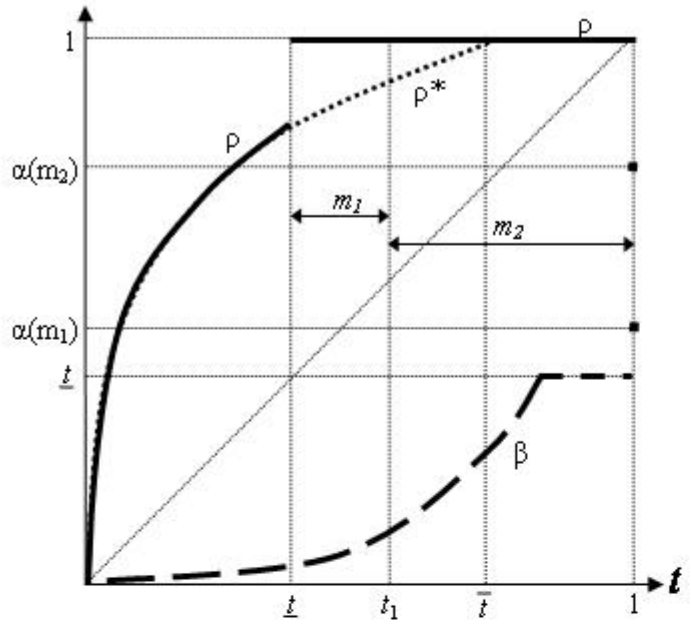

Figure 1 – A LSHP equilibrium: solid curve represents Sender's strategy via $\rho(t) = \Psi(\mu(t))$; dotted curve is the separating function, ρ^*; dashed curve represents Receiver's strategy via $\beta(t) = \cup_{m \in M_t} \alpha(m)$.

As mentioned, LSHP equilibria are an intuitively natural class of pooling equilibria to focus on: they feature message monotonicity, which is appealing given the preference of higher types for higher messages; low types separate in the cheapest way possible consistent with message monotonicity; and all pooling occurs on messages in the set M_1, which is effectively the barrier to complete separation. For these reasons, the structure of LSHP equilibria is reminiscent of the "D1" equilibria of Cho and Sobel (1990), the "central pooling" equilibria of Bernheim and Severinov (2003), and, in a more distant setting, the "variable bribes equilibria" of Esö and Schummer (2004).[14] Indeed, Cho and Sobel (1990) established that in standard signaling games where complete separation is not possible due to a bound on the signal space, refinements based on Kohlberg and Mertens's (1986) notion of *strategic stability* (Cho and Kreps, 1987; Banks and Sobel, 1987) select equilibria where low types separate in a least costly manner and high types pool on the maximal costly signal. Because the current model has some differences with standard signaling games, these refinements have limited power per se. Nevertheless, Bernheim and

[14]A difference with Cho and Sobel (1990) is that there cannot be multiple pools in their D1 equilibria, unlike here where this plays an important role. The difference arises because in the current model, not all types generally wish to be perceived as the highest type, an assumption made by Cho and Sobel (cf. Proposition 1). This is also the case in Esö and Schummer (2004). Multiple pools do arise in Bernheim and Severinov's (2003) central pooling equilibria; a distinction with their analysis is that it is important here to allow for the possibility that $\underline{t}$ may equal 0, because there generally will not exist LSHP equilibria with positive $\underline{t}$ when k is small. Bernheim and Severinov's equilibria always have $\underline{t} > 0$; the difference stems from the preference distinctions noted in fn. 11.

Severinov (2003) proposed a modification of the D1 refinement, called *monotonic D1 equilibrium*, which can be shown to rule out any non-LSHP equilibrium.[15] Moreover, there is always an LSHP equilibrium which satisfies the refinement criterion. Therefore, focussing on this class of equilibria can be formally justified. Appendix B provides complete details.

The remainder of this section looks at various special cases of the model which provide additional insights into the structure of communication.

3.3. **Large Conflicts of Interest.** LSHP equilibria are particularly simple if the conflict of interest between Sender and Receiver is large in the following precise sense.

Definition 3. The Sender has (or there is) a *large bias* if $a^S(0) \geq a^R(1)$.

When there is a large bias, the Sender would always like to be perceived as high a type as possible, which is the usual assumption in costly signaling games. This implies that there cannot be multiple pools in an LSHP equilibrium, since no type would be willing to send a message $m \in M_1$ if there exists some $m' \in M_1$ such that $\alpha(m') > \alpha(m)$. Consequently, all types above $\underline{t}$ must induce action $a^R(\underline{t}, 1)$ in equilibrium, and condition (3) becomes vacuous. This yields a simplification of condition (2) and the following result.

Proposition 1. *Assume the Sender has a large bias. There is a single pool in any LSHP equilibrium. It is necessary and sufficient for an LSHP equilibrium with $\underline{t} > 0$ that*

$$U^S(a^R(\underline{t}, 1), \underline{t}) - U^S(a^R(\underline{t}), \underline{t}) = k(C(\nu(1), \underline{t}) - C(\nu(\rho^*(\underline{t})), \underline{t})) \tag{7}$$

have a solution in the domain $(0, \bar{t})$. It is necessary and sufficient for an LSHP equilibrium with $\underline{t} = 0$ (where all types pool) that

$$U^S(a^R(0, 1), 0) - U^S(a^R(0), 0) \geq k(C(\nu(1), 0) - C(\nu(r^S(0)), 0)). \tag{8}$$

The Proposition implies that if there is large bias, then for any t, there is at most one message in M_t that is used in an LSHP equilibrium. Therefore, the rich language assumption that $|M_t|$ be large for each t is unnecessary; the analysis would apply just as well if instead $|M_t| = 1$ for all t. Note also that it is simple to directly establish existence of an LSHP equilibrium under large bias.[16] This is because the only binding incentive constraint is for type $\underline{t}$, unlike in the general bias case where incentive constraints can also bind for boundary types between pools.

[15] i.e., any equilibrium where the Sender's strategy is not an LSHP strategy for a positive measure of types.

[16] The argument is as follows: if (8) does not hold, then there must be at least one solution to (7), since each side of (7) is continuous, and at $\underline{t} = \bar{t}$ the right-hand side is 0 whereas the left-hand side is strictly positive (by $\bar{t} < 1$ and large bias).

3.4. **No Conflict of Interest.** I have assumed that $a^S(t) > a^R(t)$ for all t, which was crucial for Theorem 1. At the other extreme, consider $a^S(t) = a^R(t)$ for all t, so that there is no conflict of interest. Intuitively, one would expect complete separation with truth-telling in this case. Formally, although (DE) does not apply, it is clear that the appropriate separating function becomes $\rho^*(t) = r^S(t)$ for all t. There is a separating equilibrium where each t sends any $m \in M_{r^S(t)}$, with the caveat that any t and $t' \neq t$ use distinct messages if $r^S(t) = r^S(t')$. Since $\rho(t) = \rho^*(t)$ in such equilibria, these are indeed LSHP equilibria.

3.5. **Almost-Cheap Talk.** Return now to the general case of conflict of interest. Recall that when $k = 0$, the setting reduces to the CS model of cheap talk. An interesting question then is what happens when the intensity of lying cost is small, $k \approx 0$. It must be that $\underline{t} \approx 0$ in any LSHP equilibrium, because by (DE), the separating function ρ^* cannot be extended very far, i.e. $\bar{t} \approx 0$. Hence, as the cost of lying vanish, the type space is almost entirely partitioned into pools, with extremely inflated language.[17] The following result says that in fact, there generally is an LSHP equilibrium where no type separates.

Proposition 2. *Assume that there is no CS cheap-talk equilibrium, (α, μ), with $U^S(\alpha(\mu(0)), 0) = U^S(a^R(0), 0)$. Then, for all k sufficiently small, there is a LSHP equilibrium with $\underline{t} = 0$, where types partition as in a CS cheap-talk equilibrium.*

The assumption made in the proposition is mild: intuitively, it holds "generally" in the sense that if a given constellation of U^S, U^R, and F (the prior on types) has a cheap-talk equilibrium with $U^S(\alpha(\mu(0)), 0) = U^S(a^R(0), 0)$, then small perturbations to any of the three functions would lead to the assumption being satisfied.[18] The proof uses a result from Chen et al. (2008) that there is always a cheap-talk equilibrium where $U^S(\alpha(\mu(0)), 0) > U^S(a^R(0), 0)$, under the assumption. Pick any one of these; if k is sufficiently small, then it is an LSHP equilibrium for all types to pool just as in the cheap-talk equilibrium, using messages only in M_1, with the Receiver responding to any $m \notin M_1$ with $\alpha(m) = a^R(0)$.

3.6. **Very Costly Talk.** Consider what happens as the intensity of lying cost, k, gets large. A simple observation is that the possible gains from lying for any type are bounded in a PBE, since the range of the Receiver's strategy in a PBE is contained in $[a^R(0), a^R(1)]$. Hence, for any t and

[17] Since (3) must hold for every boundary type in the partition of $[\underline{t}, 1]$, and $\underline{t} \approx 0$, the partition of $[\underline{t}, 1]$ must be close to some CS partition in any sensible metric. Chen et al. (2008) discuss which cheap-talk equilibria can be limits of a class of equilibria in a model of costly lying as the lying costs vanish. Although their model does not nest the current setting, and the class of equilibria they consider does not include all LSHP equilibria (they require monotonicity of the Receiver's strategy, which need not hold in an LSHP equilibrium), one can nevertheless show that if $r^S(t)$ is strictly increasing, any cheap-talk equilibrium which is the limit of a sequence of LSHP equilibria as $k \to 0$ must satisfy Chen et al.'s (2008) *No Incentive to Separate* property.

[18] For example, in the uniform-quadratic case of CS, which is parameterized by a single number $b > 0$ that measures the magnitude of conflict of interest, the assumption only excludes $\left\{b : b = \frac{1}{2K(K+1)} \text{ for some } K \in \mathbb{N}^*\right\}$, which is a nowhere-dense set of zero Lebesgue measure.

any $\varepsilon > 0$, any PBE satisfies $\rho(t) \in (r^S(t) - \varepsilon, r^S(t) + \varepsilon)$ if k is sufficiently large. Since the type space is compact, it follows that when k is sufficiently large, *any PBE is close to truth-telling*. Consequently, in a LSHP equilibrium, $\underline{t}$ becomes arbitrarily close to $\inf\{t : r^S(t) = 1\}$ as $k \to \infty$. This implies that when all types have distinct cost-minimizing messages (so that r^S is strictly increasing), there is just a single pool in an LSHP equilibrium when k is sufficiently large, since all types sufficiently close to 1 wish to be perceived as high a type as possible.

4. An Example and Application

To complement the somewhat abstract analysis thus far, this section works out a canonical example, providing some further analysis that may be useful for applications. I then apply the results to the question of communication versus delegation (Dessein, 2002).

4.1. **Example.** The example is based on the uniform-quadratic setting of CS that has been the workhorse for applied work: $F(t) = t$, $U^R(a,t) = -(a-t)^2$, and $U^S(a,t) = -(a-t-b)^2$ for some bias parameter $b > 0$. For the lying cost structure, assume that $\nu(t) = t$ and $C(x,t) = -(x-t)^2$.[19] In other words, the Sender's lying cost is quadratic in the distance between his true type and the type he claims to be. This specification satisfies all the maintained assumptions and, as will be seen, is tractable. It implies that $a^S(t) = t + b$, $a^R(t) = t$, $a^R(t,t') = \frac{t+t'}{2}$, and $r^S(t) = t$.

The quadratic lying cost permits a closed-form solution for the separating function, ρ^*. To derive it, first observe that the differential equation (DE) can be simplified to

$$\rho'(t) = \frac{b}{k(\rho(t) - t)}. \tag{9}$$

The reader may verify that the family of solutions to (9) is given by

$$-\rho(t) - \frac{b}{k} \ln\left(\frac{b}{k} + t - \rho(t)\right) = c, \tag{10}$$

where c is a constant to be determined. Since the separating function has the initial condition $\rho^*(0) = r^S(0) = 0$, it follows from (10) that $c = -\frac{b}{k}\ln\left(\frac{b}{k}\right)$. Substituting this value of c into (10) and manipulating terms yields the following solution for the separating function:

$$e^{-\frac{k}{b}\rho^*(t)} = 1 - \frac{k}{b}(\rho^*(t) - t). \tag{11}$$

The upper bound on types that can separate using ρ^*, $\bar{t}$ defined by $\rho^*(\bar{t}) = 1$, can be solved for explicitly from (11) as

$$\bar{t} = 1 - \frac{b}{k}\left(1 - e^{-\frac{k}{b}}\right). \tag{12}$$

[19] It is not important what the message space is, so long as it satisfies the maintained assumptions. For concreteness, one can suppose that for each t, $M_t = \{t\} \times \mathbb{N}$, so that $M = [0,1] \times \mathbb{N}$. As discussed in Section 2.1, this has the interpretation that the Sender's message has two components: he reports an exact value about his type (which bears a lying cost) and a cheap-talk message.

It is straightforward to check that because $k > 0$ and $b > 0$, $\bar{t} \in (0, 1)$, as predicted by Theorem 1. Moreover, $\bar{t}$ is increasing in $\frac{k}{b}$ and converges to 0 (resp. 1) as $\frac{k}{b} \to 0$ (resp. ∞). This confirms the intuition that the maximal feasible set of separation, $[0, \bar{t}]$, is larger when the cost of lying is larger relative to the conflict of interest.

The necessary conditions for LSHP equilibrium, (2) for the boundary types between pools and (3) for $\underline{t} > 0$, can be simplified for the current setting as follows:

$$t_{j+1} - t_j = t_1 - t_0 + 4jb \qquad \forall j \in \{1, \ldots, J-1\}, \tag{13}$$

$$-\left(\frac{t_1 - \underline{t}}{2} - b\right)^2 + b^2 = k\left((1-\underline{t})^2 - (\rho^*(\underline{t}) - \underline{t})^2\right) \qquad \text{if } \underline{t} > 0. \tag{14}$$

Note that condition (13) is the well-known form that CS's arbitrage condition takes under uniform prior and quadratic preferences. Theorem 2 also implies that there is an LSHP equilibrium that satisfies

$$-\left(\frac{t_1}{2} - b\right)^2 + b^2 \geq k \qquad \text{if } \underline{t} = 0. \tag{15}$$

This condition says when there are no separating types, type 0 weakly prefers pooling with $[0, t_1]$ and incurring message cost k (the cost of sending any message in M_1 for type 0) to revealing itself at no cost. Conversely, given any solution to (13)–(15), there is a corresponding LSHP equilibrium.

The next result focuses on LSHP equilibria with a single pool, for these are particularly tractable and permit clear comparative statics. Condition (13) is vacuous in this case, and conditions (14) and (15) are simplified by setting $t_1 = 1$.

Proposition 3. *In the leading example:*

(a) *A single-pool LSHP equilibrium exists if and only if* $\bar{t} \geq 1 - 4b$, *or equivalently,*

$$e^{\frac{k}{b}}(4k - 1) \geq -1. \tag{16}$$

In particular, a single-pool LSHP equilibrium exists if either $k \geq \frac{1}{4}$ *or* $b \geq \frac{1}{4}$.

(b) *A single-pool LSHP equilibrium with* $\underline{t} = 0$ *(i.e., an uninformative LSHP equilibrium) exists if and only if* $b \geq k + \frac{1}{4}$.

(c) *All single-pool LSHP equilibria are essentially equivalent, i.e. they have the same cutoff type* $\underline{t}$.

(d) *If* $b \geq \frac{1}{4}$, *all LSHP equilibria have a single pool and are essentially equivalent.*

(e) *Fix any* $b > 0$ *and let* $\underline{t}(k)$ *denote the cutoff in single-pool LSHP equilibria as a function of* k, *defined on some interval* $K \subseteq \mathbb{R}_{++}$ *given by* (16). *Then* $\underline{t}(\cdot)$ *is continuous and increasing, strictly increasing at any* k *with* $\underline{t}(k) > 0$, *and differentiable at all* $k \in \mathit{Interior}(K) \setminus \{b - \frac{1}{4}\}$.

The Proposition has a number of interesting implications. Key is part (c), which assures that all single-pool LSHP equilibria have the same cutoff, and hence are essentially or outcome equivalent. CS showed that under cheap talk, an informative equilibrium — one where no Receiver action has ex-ante probability one — exists if and only if $b < \frac{1}{4}$. Part (d) of the Proposition implies that when cheap talk cannot be informative, all LSHP equilibria are essentially equivalent and have a single pool, although they are not necessarily uninformative. Indeed, combining parts (a), (b), and (d), we see that for any positive lying cost intensity, there is a set of biases $b \in [\frac{1}{4}, k + \frac{1}{4})$ for which there would be no informative cheap-talk equilibrium but there is an informative and essentially unique LSHP equilibrium. Part (b) of the Proposition also implies that if there is an informative cheap-talk equilibrium ($b < \frac{1}{4}$), all LSHP equilibria are informative. Furthermore, part (a) of the Proposition implies that when cheap talk can be informative, there is a threshold $k^*(b) < \frac{1}{4}$ such that for all $k \geq k^*(b)$, there is a single-pool LSHP equilibrium, whereas for all $k < k^*(b)$, every LSHP equilibrium has multiple pools.[20] The last part of the Proposition provides a comparative static on single-pool LSHP equilibria: in particular, for any $b > 0$, the unique cutoff is strictly increasing in k when the cutoff is strictly positive. Hence, when the lying cost intensity increases, single-pool LSHP equilibria become more informative (when they exist) in the strong sense of refining the Receiver's information partition.

4.2. **Application: Communication versus Delegation.** Dessein (2002) raised the question of whether the Receiver can do better by delegating decision-making to the Sender, without any constraints, rather than communicating via cheap talk with him. The tradeoff is one of "loss of information" under communication versus "loss of control" under delegation. Dessein proved that which arrangement is better for the Receiver depends upon the magnitude of conflict: delegation is superior to cheap-talk communication if and only if the agent's bias is sufficiently small. In particular, in the uniform-quadratic setting of CS that is the focus of this section, delegation is preferred whenever cheap talk can be informative, i.e. whenever the bias is $b < \frac{1}{4}$.

The results derived in this paper so far suggest that when manipulating information may have some costs, whether delegation dominates communication or not will depend upon the cost intensity. The discussion in Section 3.6 implies that for any given conflict of interest, communication strictly dominates delegation when k is large enough, even though there is always some loss of information under communication. The more subtle and interesting question, though, is whether for every finite intensity of lying cost, delegation will dominate communication for all small enough conflicts of interest. The issue is important for understanding whether the central

[20] It can also be shown that for any $b > 0$, if k is small enough, every LSHP equilibrium has exactly $N(b)$ pools, where $N(b)$ is the maximal number of actions that can be induced in a cheap-talk equilibrium with bias b. CS showed that $N(b)$ is the smallest integer greater than or equal to $-\frac{1}{2} + \frac{1}{2}\left(1 + \frac{2}{b}\right)^{\frac{1}{2}}$.

organizational tradeoff identified by Dessein — loss of information versus control — remains relevant so long as information is at all manipulable.[21]

For the parametric setting of this section, I am able to provide a sharp negative answer. Following Dessein (2002), say that communication is superior to or dominates delegation if there is an equilibrium under communication in which the ex-ante welfare of the Receiver is strictly higher than under delegation.

Proposition 4. *In the leading example, there is a finite $\hat{k}$ such that for any $k \geq \hat{k}$, communication is superior to delegation for all $b > 0$. In particular, if $k \geq \frac{1}{4}$ and $b \in (0, \frac{3}{16})$, communication is superior to delegation.*

To gain some intuition for the result, note that by the earlier discussion, the key issue is whether there exists some cost intensity for which communication dominates delegation for all *small* enough biases. The answer is *a priori* unclear because, for any fixed level of lying cost (including cheap talk), both communication and delegation approach the first best as conflicts vanish. Dessein (2002) showed that delegation is an order of magnitude superior to cheap-talk communication when conflicts are small. This implies that since LSHP equilibria for any $k > 0$ have one or more interval pools on the highest messages, delegation will be superior to communication if the set of pooling types is too large relative to the bias. Therefore, for communication to dominate delegation requires that the region of separating types be large enough relative to the bias. The proof of Proposition 4 shows that this is the case in single-pool LSHP equilibria for the relevant parameters.

How large exactly must lying costs be in order for communication to dominate delegation for all biases? Numerical analysis reveals that $k \geq \frac{1}{4}$ is sufficient; recall from Proposition 3 that this is precisely the threshold cost intensity that guarantees existence of single-pool LSHP for all b. Figure 2 graphs the difference in the Receiver's ex-ante utility from communication over delegation as a function of b for different values of k. The figure shows that for any $k \geq \frac{1}{4}$, communication is superior to delegation for any $b > 0$. As $k \to \infty$, communication becomes close to fully revealing (which yields ex-ante utility 0 to the Receiver), hence the Receiver's welfare difference between communication and delegation converges to $-b^2$ (which is the Receiver's utility under delegation).

5. Withholding Information and Verifiable Disclosure

The cheap-talk approach to strategic communication assumes that it is costless for the Sender to lie arbitrarily. An alternative approach, initiated by Grossman (1981) and Milgrom (1981), is to assume that that the Sender can tell the truth or withhold information at no cost, but cannot lie. A justification is that "there are penalties for perjury, false advertising or warranty

[21] If information is not manipulable at all, or equivalently there are infinite lying costs, then as Dessein (2002, p. 831) noted, communication will be fully revealing and hence dominate delegation for all levels of bias.

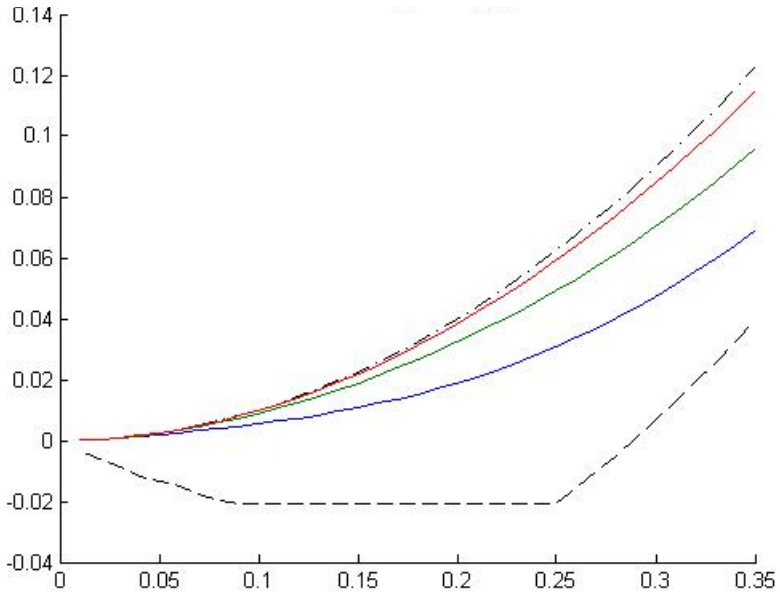

Figure 2 – Receiver's ex-ante welfare gain from communication over delegation as a function of the bias, b. Highest curve is b^2, corresponding to full-revelation under $k = \infty$ (verifiable disclosure); next three are for single-pool LSHP equilibrium with $k = 1$, $k = 0.5$, and $k = 0.25$ respectively; lowest curve is for most-informative equilibrium of cheap talk.

violation that are sufficiently sure and heavy that false reporting never pays" (Milgrom and Roberts, 1986b, p. 19). In the literature, this is often referred to as *verifiable disclosure*, or persuasion or partial-provability. The classic result is that there is complete separation in every PBE (Seidmann and Winter, 1997), due to a well-known "unraveling argument."[22] This is in striking contrast with the cheap-talk prediction of no information revelation when there is a large enough conflict of interest between Sender and Receiver (Crawford and Sobel, 1982, Corollary 1).

Costly lying provides a framework to unify these divergent results. However, the basic model I have studied has a limitation in this respect: it does not permit the Sender to withhold information freely, and to this extent, the model does not perfectly correspond to verifiable disclosure when $k \to \infty$. Accordingly, this section develops an extension to accommodate the costless withholding of information.[23]

5.1. **A Simple Form of Withholding Information.** For this section, assume that $\nu(t) = t$ and that $r^S(t) = t$, to provide a clear comparison with the standard verifiable disclosure setting. Moreover, to avoid some tedious complications that do not add much insight, assume that the

[22] In general, unraveling requires enough structure on preferences — satisfied by the CS assumptions on U^S and U^R — and that each type be able to verifiably disclose itself if it wishes to.

[23] I am grateful to Yeon-Koo Che for prompting this extension.

Sender always wishes to be thought of as the highest type, i.e. there is large bias. This is the standard assumption in verifiable disclosure models, and as noted above, implies that cheap talk and verifiable disclosure generate polar predictions.

Augment the message space to now be $M^* = M \cup M_\phi$, where M is as before, and M_ϕ is a new set of one or more messages (i.e., $|M_\phi| \geq 1$ and $M_\phi \cap M = \emptyset$). Assume that for any t and $m \in M_\phi$, $C(m,t) = C(t,t)$,[24] so that all messages in $M_\phi \cup M_t$ are equally costly for the Sender. The rest of the model remains as before. The set M_ϕ can be viewed as a set of pure cheap-talk messages (it is without loss of generality to normalize $C(m,t) = 0$ for all $m \in M_\phi$ and t), or as the Sender being silent, or otherwise *freely withholding information*.[25] In this formulation, withholding information takes a particularly stark and simple form: at the limit as $k \to \infty$, the Sender is effectively choosing between either credibly revealing his true type, or sending a message that is costless to all types. While this structure will be generalized shortly, it is convenient to start with this simple formulation, which I call the *withholding model*.

Consider first whether there can be full separation for any k. Suppose there is such an equilibrium. Any message in M_ϕ can not be used by any $t > 0$, for all lower types would mimic it. This implies that separation of all types greater than zero must be achieved with non-withholding messages. However, the argument of Theorem 1 can be extended to show that this is impossible. Consequently, there must be pooling in equilibrium. As it is not my goal to characterize every pooling equilibrium, some of which can be implausible,[26] I only show that LSHP equilibria can be extended in an obvious way to the current environment.

To state this precisely, start with any LSHP equilibrium strategy profile when withholding is not allowed. Now consider a strategy profile in the withholding model where the Receiver responds to any withholding message, $m \in M_\phi$, with a maximally skeptical inference and plays $\alpha(m) = a^R(0)$; in other respects, strategies are identical to the LSHP profile. Call such a strategy profile an *augmented LSHP profile*.

Proposition 5. *An augmented LSHP profile is an equilibrium of the withholding model.*

The reason is simple: given the Receiver's play, telling the truth is always at least as good (strictly better if $t > 0$) as withholding, since it costs the same and leads to a weakly preferred (strictly, if $t > 0$) action. Hence, for any lying cost intensity, arbitrarily large or small, there are equilibria with the following properties: there is lying on the equilibrium path (since a positive

[24] Writing $C(m,t)$ involves an abuse of notation, but is a convenient shorthand that should not cause any confusion.

[25] If one instead assumes that $C(t,t) > C(m,t)$ for any t and $m \in M_\phi$, then the Sender's cost of sending any $m \notin M_\phi$ has fixed and marginal components. Although this formulation is not analyzed here, I note that when combined with "$k = \infty$," it yields a model akin to *costly disclosure* (e.g., Verrecchia, 1983), which has also been studied recently by Esö and Galambos (2008).

[26] For example, there is always an uninformative equilibrium where all types send the same $m \in M_\phi$, and the Receiver responds to every message with $a^R(0,1)$. This can be ruled out by standard refinements such as Cho and Kreps's (1987) D1 criterion, for any $k > 0$ (and even weaker criteria for large k).

measure of types play $m \in M_1$ in any LSHP equilibrium); there exists some $\underline{t} \geq 0$ such that types in $[\underline{t}, 1]$ form a single pool, claiming to be the highest type; all types below $\underline{t}$ separate, but claim to be higher types than they truly are; withholding does not occur on the equilibrium path; and the Receiver responds to withholding with a maximally skeptical inference. These last two properties are standard in the equilibria of verifiable disclosure games (e.g., Milgrom and Roberts, 1986b, Proposition 2), with the caveat that of course the lowest type could just as well fully disclose or withhold.

Proposition 5 implies that the insights of the original non-withholding model can be applied to the current environment. Of particular interest are the limits as $k \to 0$ or $k \to \infty$.

Corollary 1. *When $k \approx 0$, any equilibrium in an augmented LSHP profile is uninformative, with all types pooling on some $m \in M_1$. As $k \to \infty$, any sequence of equilibria in augmented LSHP profiles has $\underline{t} \to 1$ and, for all t, $\rho(t) \to t$.*

Thus, at the cheap-talk limit of augmented LSHP equilibria, not only is there no information revealed,[27] but language is entirely inflated (all types claim to be the highest), whereas at the verifiable disclosure limit, there is full separation through truth-telling. In this sense, the augmented LSHP equilibria provide a sensible spanning of the extreme cases both in terms of how much information is conveyed and the messages that are used.

5.2. **Richer Forms of Withholding.** As previously noted, the withholding model discussed above is special because, at the limit of $k \to \infty$, the Sender has only two undominated choices: either completely withhold information about his type or completely reveal it. Verifiable disclosure models often endow the Sender with a much richer set of choices, such as reporting that his type lies in any closed subset of the type space that contains his true type. This is meant to capture the idea that the Sender can credibly reveal some information without revealing it all. To consider such possibilities when lying costs are not infinite requires a specification of the cost of each such message for each type. I will discuss a simple formulation which allows the previous analysis to extend.

Let $\mathcal{T}$ be any set of closed subsets of $T = [0, 1]$, such that as for each $t \in T$, $\{t\} \in \mathcal{T}$. The Sender is permitted to claim that his type lies in any element of $\mathcal{T}$, i.e. for any $X \in \mathcal{T}$, there is a (non-empty) set of messages M_X with the literal meaning "my type is in X." Let $\theta : \mathcal{T} \times T \to T$ be any function satisfying $\theta(X, t) \in X$ and $\theta(X, t) = t$ if $t \in X$. Starting with the usual cost function that is defined for all precise statements (i.e., any singleton $X \in \mathcal{T}$), I extend it to each vague statement (i.e., any non-singleton $X \in \mathcal{T}$) by assuming that for any $m \in M_X$, $C(m, t) = C(\theta(X, t), t)$. That is, the cost for the Sender of claiming that his type is in a set X is the same as the cost of claiming that his type is exactly $\theta(X, t)$. Note that by the assumed

[27] Note that since bias is assumed to be large, all cheap-talk equilibria are uninformative.

properties of θ, the cost for any type t of claiming to be in any set $X \in \mathcal{T}$ is the same as the cost of claiming to be t if and only if $t \in X$.[28]

The model without any withholding is the special case where $\mathcal{T}$ consists only of singleton sets. The withholding model discussed earlier is the case where $\mathcal{T}$ consists of just the singletons and the set T, with $\theta(T,t) = t$ for all t. In general, one natural possibility for θ is that $\theta(X,t) = \arg\min_{\tilde{t} \in X} |\tilde{t} - t|$. This captures the idea that a vague statement is a lie to the extent it is a lie when given its most favorable interpretation, i.e. the distance of the true type to the closest point in X. For any $\theta(\cdot,\cdot)$, the Sender's undominated choices at the limit of $k \to \infty$ are to claim that his type is in any $X \in \mathcal{T}$ such that $\theta(X,t) = t$, which is precisely any $X \in \mathcal{T}$ such that $t \in X$, just as in standard disclosure models.

The key to extending LSHP equilibria to this setting, for any lying cost intensity, is to have the Receiver respond to any message with the most skeptical inference among the set of types for whom that message is least costly. Formally, pick any LSHP equilibrium profile when withholding is not available that satisfies $\alpha(m) \geq \alpha(\tilde{m})$ if $m \in M_t$ and $\tilde{m} \in M_{\tilde{t}}$ for two types $t \geq \tilde{t}$. (Such an LSHP equilibrium profile exists by the proof of Theorem 2.) The profile specifies the Receiver's response, $\alpha(m)$, for any $m \in M_X$ with $|X| = 1$. Extend α to vague messages as follows: for any $m \in M_X$ with $|X| > 1$, $\alpha(m) = \alpha(\tilde{m})$ for some $\tilde{m} \in M_{\min X}$. In other words, if the Sender claims to be in a non-singleton set X, the Receiver responds by taking some action that she would in response to a claim that the Sender is the lowest type in X.[29] To see that this extended LSHP profile is an equilibrium for any k, suppose to contradiction that $m \in M_X$ is a profitable deviation for a type t. By construction, there exists $\tilde{m} \in M_{\min X}$ such that $\alpha(m) = \alpha(\tilde{m})$. If $t \leq \min X$, $C(m,t) = C(\theta(X,t),t) \geq C(\tilde{m},t)$, hence $\tilde{m}$ is a profitable deviation, a contradiction since we started with an LSHP equilibrium profile. If $t > \min X$, then any $m' \in M_t$ is a profitable deviation, because $C(m',t) \leq C(m,t)$ and $\alpha(m') \geq \alpha(m) = \alpha(\tilde{m})$, also a contradiction since we started with an LSHP equilibrium profile.

6. Conclusion

This paper has studied a model of communication between an informed Sender and an uninformed Receiver, where the Sender has a cost of lying or misrepresenting his private information in certain ways. These costs on the Sender may stem from various sources, such as technological, legal, or psychological constraints. The main results show that inflated language

[28] The assumptions on $\theta(\cdot,\cdot)$ can be weakened to relax this implication, but it would unnecessarily complicate the subsequent analysis.

[29] The reader may wonder why not use a simpler construction where the Receiver plays $\alpha(m) = a^R(0)$ if $m \in M_X$ with $|X| > 1$. The reason is that even though this simpler construction will also be an equilibrium, it will not satisfy standard refinements (such as the D1 criterion) for all k. In particular, as $k \to \infty$, it requires the Receiver to form the implausible belief that the Sender is type 0 upon observing any vague message, even if it is prohibitively costly for type 0 to send that message. The construction in the text is not vulnerable to this objection; rather, beliefs as $k \to \infty$ converge to equilibrium beliefs of the verifiable disclosure game.

naturally arises in this environment, where the Sender (almost) always claims to be a higher type than he would under complete information. Regardless of the intensity of lying cost, there is incomplete separation, with some pooling on the highest messages. How much information is revealed is a function of the intensity of lying cost. An extension of the basic model to allow costless withholding of information provides a unified framework to embed both cheap talk and verifiable disclosure as limiting cases.

The basic finding that lying costs enlarge the scope for direct communication to reduce the costs of asymmetric-information-based incentive problems is, on one hand, not surprising, but on the other hand, likely important in a wide variety of economic problems. A formal framework and analysis such as that developed may be useful in applications where it is important to understand exactly how incentive constraints are relaxed.

Section 4 provided one such example, focussing on the interaction between preference conflicts and lying costs. In an influential paper, Dessein (2002) argued that when a decision-maker or principal is faced with a biased but better-informed agent, unconstrained delegation of decision-making is better than cheap-talk communication whenever the conflict of interests are not too large. I have shown that in the simple but canonical setting of Section 4, if misrepresentation costs are even moderately large, this conclusion is reversed: the principal prefers to communicate than delegate regardless of how small the conflict of interest is, even though communication necessarily entails some loss of information. Besides matters of generalization, this raises a number of substantive questions. For example, Ivanov (2007) has shown that the principal is better off restricting the precision of the agent's information compared to either delegation or cheap-talk communication. Plainly, the magnitude of misrepresentation costs will have a significant effect on how beneficial such "information control" is. In the limiting case of verifiable information or infinite lying costs, the principal would strictly prefer to not restrict information at all. Whether information control is valuable whenever lying costs have finite intensity, and if so, how this control should be structured are open questions.

The fact that lying costs can mitigate inefficiencies resulting from asymmetric information suggests that applications to negotiation and bargaining will also be fruitful. For example, Valley et al. (2002) document striking experimental evidence that subjects in double auctions achieve greater efficiency given pre-play communication than the theoretical limits derived by Myerson and Satterthwaite (1983).[30] Importantly, the subjects do significantly better when communication is face-to-face than through written messages, which Valley et al. (2002) largely attribute to their evidence that subjects misrepresent valuations substantially less in face-to-face communication than in written messages. It seems plausible that the more direct the interface of communication, the more costly (difficult, uncomfortable, etc.) it is to lie. Accordingly, these empirical findings

[30] I thank Vince Crawford for this reference.

may be organized by a framework such as that developed here, although a model adapted to the bargaining environment would yield more finely-tuned insights.

Appendix A: Proofs

Proof of Lemma 1 on page 9. By hypothesis, types (t_l, t_h) are separating in a monotone PBE with mapping ρ. If ρ is constant on some interval $X \subseteq (t_l, t_h)$, then some $t \in X$ can profitably deviate and mimic a slightly higher type $t + \varepsilon \in X$, contradicting equilibrium separation; hence ρ must be strictly increasing on (t_l, t_h). This implies that for any $t \in (t_l, t_h)$, $\rho(t) \in (0, 1)$. Separation requires that for each $t \in (t_l, t_h)$, t must be a solution to (1). If ρ is differentiable on (t_l, t_h), then the first order condition obtained by differentiating (1) is (DE), as required. So it suffices to prove that ρ is differentiable on (t_l, t_h). This is done in series of steps.

Claim: $\rho(t) \neq r^S(t)$ for all $t \in (t_l, t_h)$.

Proof: Suppose there exists a $\hat{t} \in (t_l, t_h)$ such that $\rho(\hat{t}) = r^S(\hat{t})$. Without loss, we can assume that $r^S(\hat{t}) \in (0, 1)$, because otherwise ρ cannot be strictly increasing on (t_l, t_h). This implies that

$$C_1(\nu(r^S(\hat{t})), \hat{t}) = 0. \tag{A-1}$$

Define $g(\varepsilon)$ as the expected utility gain for a type $\hat{t} - \varepsilon$ by deviating from $\rho(\hat{t} - \varepsilon)$ to $\rho(\hat{t})$. Since we are on a separating portion of the type space when $|\varepsilon|$ is small,

$$g(\varepsilon) = \left[U^S\left(a^R\left(\hat{t}\right), \hat{t} - \varepsilon\right) - kC\left(\nu(r^S(\hat{t})), \hat{t} - \varepsilon\right)\right] - \left[U^S\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right) - kC\left(\nu(\rho\left(\hat{t} - \varepsilon\right)), \hat{t} - \varepsilon\right)\right].$$

Since $C\left(\nu(\rho\left(\hat{t} - \varepsilon\right)), \hat{t} - \varepsilon\right) \geq C\left(\nu(r^S(\hat{t} - \varepsilon)), \hat{t} - \varepsilon\right)$,

$$\begin{aligned} g(\varepsilon) \geq \phi(\varepsilon) \quad := \quad & \left[U^S\left(a^R\left(\hat{t}\right), \hat{t} - \varepsilon\right) - kC\left(\nu(r^S(\hat{t}), \hat{t} - \varepsilon)\right)\right] - \\ & \left[U^S\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right) - kC\left(\nu(r^S(\hat{t} - \varepsilon)), \hat{t} - \varepsilon\right)\right]. \end{aligned}$$

ϕ is differentiable in a neighborhood of 0, and has derivative

$$\begin{aligned} \phi'(\varepsilon) \quad = \quad & -U_2^S\left(a^R\left(\hat{t}\right), \hat{t} - \varepsilon\right) + kC_2\left(\nu(r^S(\hat{t})), \hat{t} - \varepsilon\right) - kC_1\left(\nu(r^S(\hat{t} - \varepsilon)), \hat{t} - \varepsilon\right)\nu'(r^S(\hat{t}) - \varepsilon)r_1^S(\hat{t} - \varepsilon) \\ & -kC_2\left(\nu(r^S(\hat{t} - \varepsilon)), \hat{t} - \varepsilon\right) + U_1^S\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right)a_1^R\left(\hat{t} - \varepsilon\right) + U_2^S\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right). \end{aligned}$$

Observe that $\phi(0) = 0$ and $\phi'(0) = U_1^S\left(a^R\left(\hat{t}\right), \hat{t}\right)a_1^R\left(\hat{t}\right) > 0$ (using (A-1)). Hence, for 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 of $\hat{t}$. $\diamond$

Claim: $\rho(t) > r^S(t)$ for all $t \in (t_l, t_h)$.

Proof: Suppose not. By the previous Claim, there exists a $t' \in (t_l, t_h)$ such that $\rho(t') < r^S(t')$. By monotonicity of ρ, $\rho(t' - \varepsilon) < \rho(t')$ for all $\varepsilon > 0$. It follows that for small enough $\varepsilon > 0$, $C\left(\nu(\rho(t' - \varepsilon)), t' - \varepsilon\right) > C\left(\nu(\rho(t')), t' - \varepsilon\right)$. On the other hand, for small enough $\varepsilon > 0$, $U^S\left(a^R(t'), t' - \varepsilon\right) > U^S\left(a^R(t' - \varepsilon), t' - \varepsilon\right)$. Therefore, for small enough $\varepsilon > 0$, a type $t' - \varepsilon$ strictly prefers to imitate t', contradicting equilibrium separation of t'. $\diamond$

Claim: ρ is continuous on (t_l, t_h).

Proof: Suppose there is a discontinuity at some $\hat{t} \in (t_l, t_h)$. First assume $\rho(\hat{t}) < \lim_{t \downarrow \hat{t}} \rho(t) =: \overline{\rho}$. By the continuity of C and ν, and the monotonicity of ρ, as $\varepsilon \downarrow 0$,

$$C\left(\nu(\rho\left(\hat{t} + \varepsilon\right)), \hat{t} + \varepsilon\right) - C\left(\nu(\rho\left(\hat{t}\right)), \hat{t} + \varepsilon\right) \quad \to \quad C\left(\nu(\overline{\rho}), \hat{t}\right) - C\left(\nu(\rho\left(\hat{t}\right)), \hat{t}\right) > 0,$$

where the inequality above follows from $\overline{\rho} > \rho\left(\hat{t}\right) \geq r^S(\hat{t})$. On the other hand, $U^S\left(a^R\left(\hat{t} + \varepsilon\right), \hat{t} + \varepsilon\right) - U^S\left(a^R\left(\hat{t}\right), \hat{t} + \varepsilon\right) \to 0$ as $\varepsilon \downarrow 0$; hence, for small enough $\varepsilon > 0$, $\hat{t} + \varepsilon$ prefers to imitate $\hat{t}$, contradicting equilibrium separation.

The argument for the other case where $\rho\left(\hat{t}\right) > \lim_{t\uparrow\hat{t}} \rho\left(t\right)$ is similar, establishing that $\hat{t}$ prefers to imitate $\hat{t} - \varepsilon$ for small enough $\varepsilon > 0$. $\diamond$

CLAIM: ρ is differentiable on (t_l, t_h).

PROOF: Given the previous claims that ρ is continuous and does not intersect r^S on (t_l, t_h), one can replicate the argument of Mailath (1987, Proposition 2 in Appendix). $\diamond$ *Q.E.D.*

The following Lemma is used in the proof of Theorem 1 and proves uniqueness of the separating function.

Lemma A.1. *Fix any $t_0 \in [r^S(0), 1]$. There is a unique solution to the problem of finding a $\bar{t} \in [0,1]$ and $\rho : [0, \bar{t}] \to [0,1]$ such that (i) ρ is strictly increasing and continuous on $[0, \bar{t}]$, (ii) $\rho(0) = t_0$, (iii) ρ solves (DE) on $(0, \bar{t})$, and (iv) $\rho(\bar{t}) = 1$ if $\bar{t} < 1$. Moreover, $\bar{t} = 0$ if $r^S(0) = 1$, otherwise $\bar{t} \in (0, \min\{t : r^S(t) = 1\})$.*

Proof. If $r^S(0) = 1$, the Lemma is trivial, so assume $r^S(0) < 1$. Since $C_1(\nu(r^S(t)), t) = 0$ for all t such that $r^S(t) \in (0,1)$, there is no Lipschitz condition on (DE) on the relevant domain, in particular even locally if $t_0 = r^S(0)$. Thus, standard results on differential equations do not apply. Instead, I proceed as follows, building on Mailath (1987).

STEP 1: (Local existence and uniqueness.) Consider the inverse initial value problem to find $\tau(\hat{t})$ such that:

$$\tau' = g(\hat{t}, \tau) := \frac{kC_1(\nu(\hat{t}), \tau)\nu'(\hat{t})}{U_1^S(a^R(\tau), \tau)a_1^R(\tau)}, \quad \tau(\hat{t}_0) = 0. \tag{A-2}$$

By the assumptions on C, ν, and U^S, g is continuous and Lipschitz on $[0,1] \times [0,1]$. Hence, standard theorems (e.g. Coddington and Levinson, 1955, Theorem 2.3, p. 10) imply that there is a unique location solution, $\tilde{\tau}$, to (A-2) on $[\hat{t}_0, \hat{t}_0 + \delta)$, for some $\delta > 0$; moreover, $\tilde{\tau} \in \mathcal{C}^1([\hat{t}_0, \hat{t}_0 + \delta))$.[31] Inverting $\tilde{\tau}$ gives a strictly increasing $\tilde{\rho}$ on $[0, \tilde{t})$ such that $\tilde{\rho}(0) = t_0$, $\tilde{\rho}$ solves (DE) on $(0, \tilde{t})$, and $\tilde{\rho} \in \mathcal{C}^1([0, \tilde{t}))$. Note that $\tilde{\rho}(t) > r^S(t)$ for all $t \in (0, \tilde{t})$. Since the inverse of any increasing local solution to (DE) is a local solution to (A-2) on $[r^S(0), r^S(0) + \eta)$ for some $\eta > 0$, local uniqueness of an increasing solution to (DE) follows from the fact that $\tilde{\tau}$ is the unique local solution to (A-2) above $r^S(0)$.

STEP 2: ($\tilde{\rho}$ is bounded away from r^S above 0.) Suppose $\tilde{\rho}$ is a solution to (DE) on $[0, \delta)$, with $\tilde{\rho} \in \mathcal{C}^1([0, \delta))$ and $\tilde{\rho}' > 0$. Let $t_\delta := \lim_{t\uparrow\delta} \tilde{\rho}\left(t\right)$. I claim that $t_\delta > r^S(\delta)$. This is obviously true if $r^S(\delta) = 0$, so assume $r^S(\delta) > 0$, and $t_\delta \leq r^S(\delta)$, towards contradiction. We must have $t_\delta = r^S(\delta)$ (because $\tilde{\rho}(t) > r^S(t)$ for all $t \in (0, \delta)$) and $\lim_{t\uparrow\delta} \tilde{\rho}'(t) = \infty$. Let $z := \max r_1^S(t)$. By the assumptions on $C(\cdot,\cdot)$, $z < \infty$. Since $\tilde{\rho} \in \mathcal{C}^1([0, \delta))$, there exists $t_1(0, \delta)$ such that $\tilde{\rho}'(t) > z$

[31] $\mathcal{C}^1([a,b))$ is the set of all Real-valued functions on $[a,b)$ that have a continuous derivative at all $\hat{t} \in (a,b)$ and in addition have a right-hand derivative at a that is continuous from the right at a.

for all $t \in [t_1, \delta)$. Pick $\varepsilon > 0$ such that $\tilde{\rho}(t_1) > r^S(t_1) + \varepsilon$. We have

$$\begin{aligned} t_\delta &= \tilde{\rho}(t_1) + \lim_{t \uparrow \delta} \int_{t_1}^{t} \tilde{\rho}'(y) dy \\ &> r^S(t_1) + \varepsilon + \int_{t_1}^{\delta} \tilde{\rho}'(y) dy \\ &> r^S(t_1) + \varepsilon + \int_{t_1}^{\delta} r_1^S(y) dy \\ &= r^S(\delta) + \varepsilon, \end{aligned}$$

which contradicts $t_\delta = r^S(\delta)$.

STEP 3: The proof is completed as follows. If $\tilde{\rho}$ is defined on $[0, \delta)$ with $\lim_{t \uparrow \delta} \tilde{\rho}(t) < 1$, then by the previous step, $\frac{U_1^S(a^R(t),t) a_1^R(t)}{kC_1(r,t)}$ is continuous, Lipschitz, and bounded in a neighborhood of $(\delta, \lim_{t \uparrow \delta} \tilde{\rho}(t))$, hence standard theorems (e.g. Coddington and Levinson, 1955, Theorem 4.1 and preceeding dicussion, p. 15) imply that there is a unique extension of $\tilde{\rho}$ to $[0, \delta + \eta)$ for some $\eta > 0$; this extension is strictly increasing, in $\mathcal{C}^1([0, \delta + \eta))$, and by the previous step, $\lim_{t \uparrow \delta + \eta} \tilde{\rho}(t) > r^S(\delta + \eta)$. Now let $\bar{t} := \sup\{x : \tilde{\rho} \text{ can be extended to } [0, x)\})$. Clearly, $\bar{t} \in (0, \min\{t : r^S(t) = 1\})$. Moreover, we must have $\lim_{t \uparrow \bar{t}} \tilde{\rho}(t) = 1$, for otherwise $\tilde{\rho}$ can be extended beyond $\bar{t}$. We are done by setting $\tilde{\rho}(\bar{t}) = 1$. *Q.E.D.*

Proof of Theorem 1 on page 9. Suppose there is a separating monotone equilibrium, (α, μ). By Lemma 1, the induced map $\rho = \Psi \circ \alpha$ solves (DE) on $(0, 1)$. Let $t_0 := \lim_{t \downarrow 0} \rho(t)$. Since $\rho(t) > r^S(t)$ for all $t \in (0, 1)$, $t_0 \geq r^S(0)$. However, by Lemma A.1, there is no solution on $[0, 1)$ to the initial value problem given by $\rho(0) = t_0$ and (DE) on $(0, 1)$, a contradiction. *Q.E.D.*

The following Lemma is used in the proof of Theorem 2.

Lemma A.2. *There exists $\underline{t} \in [0, \bar{t}]$ and a strictly increasing sequence, $\langle t_0 = \underline{t}, t_1, \ldots, t_J = 1 \rangle$, that solve (4), (5), and (6). Moreover, if $r^S(0) < r^S(1)$ and k is large enough, this is true with some $\underline{t} > 0$.*

Proof. The proof is constructive.

STEP 0: PRELIMINARIES.

Start by defining the function

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

$\phi(t)$ is the utility gain for type t from sending a message in M_1 and receiving its ideal action over separating itself (thus inducing $a^R(t)$) with a message in $M_{\rho^*(t)}$. Note that in an LSHP equilibrium, the gain from pooling over separating can be no more than $\phi(t)$, and will generally be strictly less. 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

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

ϕ is continuous and $\phi\left(\bar{t}\right) > 0$; hence $\hat{t} < \bar{t}$ and for all $t \in \left(\hat{t}, \bar{t}\right]$, $\phi(t) > 0$. In everything that follows, we are only concerned with $t \in \left[\hat{t}, \bar{t}\right]$. So statements such as "for all t" are to be read as "for all $t \in [\hat{t}, \bar{t}]$" and so forth unless explicitly specified otherwise.

STEP 1: CONSTRUCTING THE NECESSARY SEQUENCES.

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

$$\Delta(a,t) := U^S(a,t) - kC(\nu(1),t) - \left[U^S\left(a^R(t),t\right) - kC\left(\nu(\rho^*(t)),t\right)\right].$$

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 $t \in \left[\hat{t}, \bar{t}\right]$, it follows that for any such 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 $a_1^r(t)$. By continuity of Δ, a_1^l and a_1^r are continuous, $a_1^l\left(\bar{t}\right) = a_0^r\left(\bar{t}\right)$, and $a_1^r\left(\hat{t}\right) = a_1^l\left(\hat{t}\right) = a^S\left(\hat{t}\right)$ if $\hat{t} > 0$. By the monotonocity of $a^R(\cdot,\cdot)$, for $q \in \{l,r\}$ and t, there is either no solution or a unique t' that solves $a^R(t,t') = a_1^q(t)$. If there is a solution, call it $p_1^q(t)$, otherwise set $p_1^q(t) = 1$. It is straightforward 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 $t = \bar{t}$, and $p_1^r(t) > p_0^r(t)$. Note that $p_1^r(t) \geq p_1^l(t)$, and $p_1^l\left(\hat{t}\right) = p_1^r\left(\hat{t}\right)$ if $\hat{t} > 0$.

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

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

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 of a^R and $U_{11}^S < 0$, $p_j^q(t)$ is well-defined and unique. Define $a_j^q(t) := a^R\left(p_{j-1}^q(t), p_j^q(t)\right)$. 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, $p_j^l\left(\hat{t}\right) = p_j^r\left(\hat{t}\right)$ if $\hat{t} > 0$, and $p_{j+1}^l\left(\bar{t}\right) = p_j^r\left(\bar{t}\right)$ (these follow easily by induction, given that we noted all these properties for $j = 1$).

STEP 2: THE CRITICAL SEGMENT B.

I claim there exists $B \geq 1$ such that $p_{B-1}^r\left(\bar{t}\right) < 1 = p_B^r\left(\bar{t}\right)$. (Obviously, if it exists, it is unique.) To see this, first note that by definition, $p_0^r\left(\bar{t}\right) = \bar{t} < 1$. Let $\overline{K} = \inf\{K : p_K^r(\bar{t}) = 1\}$.[32] It is sufficient to show that $\exists \varepsilon > 0$ such that for any $j < \overline{K}$, $\left|a_{j+1}^r\left(\bar{t}\right) - a_j^r\left(\bar{t}\right)\right| \geq \varepsilon$. By construction, $a_j^r\left(\bar{t}\right) < a^S\left(p_j^r\left(\bar{t}\right)\right) < a_{j+1}^r\left(\bar{t}\right)$ and $a_j^r\left(\bar{t}\right) \leq a^R\left(p_j^r\left(\bar{t}\right)\right) \leq a_{j+1}^r\left(\bar{t}\right)$. Since $\min_{t\in[0,1]}\left|a^S(t) - a^R(t)\right| > 0$, we are done.

STEP 3: EXISTENCE WHEN $\hat{t} > 0$.

Consider the functions p_B^l and p_B^r. These are continuous, and $p_B^l\left(\bar{t}\right) = p_{B-1}^r\left(\bar{t}\right) < 1 = p_B^r\left(\bar{t}\right)$. Moreover, $p_B^l\left(\hat{t}\right) = p_B^r\left(\hat{t}\right)$; hence either $p_B^l(\hat{t}) = 1$ or $p_B^r(\hat{t}) < 1$. It follows that there is some type $\tilde{t} \in [\hat{t}, \bar{t}]$ such that either (i) $p_B^l(\tilde{t}) = 1$ and $p_B^l(t) < 1$ for all $t > \tilde{t}$; or (ii) $p_B^r(\tilde{t}) = 1$ and $p_B^r(t) < 1$ for all $t < \tilde{t}$. Let $q = l$ if (i) is the case; $q = r$ if (ii) is the case. By construction, setting $\underline{t} = \tilde{t}$ and $t_j = p_B^q(\tilde{t})$ for $j = 1, \ldots, B$ satisfies the needed properties.

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

STEP 4: EXISTENCE WHEN $\hat{t} = 0$.

By the continuity of p_B^l and p_B^r, the logic in Step 3 can fail when $\hat{t} = 0$ only if $p_B^l(0) < 1 = p_B^r(0)$. So suppose this is the case. Note that this requires $p_1^l(0) < p_1^r(0)$. For any $t \in \left[p_1^l(0), p_1^r(0)\right]$, $U^S\left(a^R(0,t),0\right) - kC(1,0) - \left[U^S\left(a^r(0),0\right) - kC(0,0)\right] \geq 0$, 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 $a^R(0,t)$ with a message in M_1 over inducing $a^R(0)$ with message in $M_{r^S(0)}$. This follows from the construction of p_1^l and p_1^r, and $U_{11}^S < 0$. 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

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

if a solution exists that is strictly greater than $\tau_{j-1}(t)$, and otherwise set $\tau_j(t) = 1$. It is straightforward that for all $j \geq 0$, $\tau_j(t)$ is continuous in t. Since $\tau_B\left(p_1^l(0)\right) = p_B^l(0) < 1 = p_B^r(0) = \tau_B\left(p_1^r(0)\right)$, it follows that $\tilde{t} = \min_{t \in [p_1^l(0), p_1^r(0)]}\{t : \tau_B(t) = 1\}$ is well-defined and lies in $\left(p_1^l(0), p_1^r(0)\right]$. By construction, setting $\underline{t} = 0$ and $t_j = \tau_j(\tilde{t})$ for $j = 1, \ldots, B$ satisfies the needed properties.

Finally, the proof of the Lemma is completed by noting that $\hat{t} > 0$ for all k large enough if $r^S(0) < r^S(1)$, because $\underline{t} > 0$ if $r^S(0) < r^S(1)$ and $\phi(0) < 0$ for all k large enough in this case. *Q.E.D.*

Proof of Theorem 2 on page 11.

The necessity of (4) and (5) follows from the discussion in the text preceding the Theorem. Fix any $\underline{t} \geq 0$ and a strictly increasing sequence $\langle t_0 = \underline{t}, t_1, \ldots, t_J = 1\rangle$ that satisfy conditions (4), (5), and (6). I will show that there is a corresponding LSHP equilibrium. This suffices to prove the Theorem because of Lemma A.2.

I first define the strategy profile. Based on the discussion in the main text preceding the Theorem, the Sender's strategy, μ, is clear, with the only possible ambiguity being for types $t_0, \ldots, t_{J-1}$: assume for concreteness that they "pool up" rather than "down" (so type $\underline{t}$ does not separate). Let $M^\mu := \bigcup_t \mu(t)$ be the set of messages used in this strategy. It is clear what the Receiver's response, $\alpha(m)$, must be for any $m \in M^\mu$. For any $m \notin M^\mu$, the Receiver plays as follows:

(1) For any $m \in M_t \setminus M^\mu$ with $t < r^S(0)$, $\alpha(m) = a^R(0)$.
(2) For any $m \in M_t \setminus M^\mu$ with $t \in [\rho^*(0), \rho^*(\underline{t}))$, $\alpha(m) = a^R((\rho^*)^{-1}(t))$.
(3) For any $m \in M_t \setminus M^\mu$ with $t \in [\rho^*(\underline{t}), 1)$, $\alpha(m) = a^R(\underline{t})$.
(4) For any $m \in M_1 \setminus M^\mu$, $\alpha(m) = a^R(\underline{t}, t_1)$.

I now argue that (μ, α) as constructed above is an LSHP equilibrium. Obviously, μ is an LSHP strategy and α is optimal given beliefs that are derived by Bayes rule on the equilibrium path. So it only needs to be shown that μ is optimal for the Sender.

No type $t \geq \underline{t}$ has a profitable deviation to some $m \in M_1$, because $\alpha(m) = a^R(t_0, t_1)$ for any $m \in M_1 \setminus M^\mu$ and $U^S(\alpha(\mu(t)), t) \geq U^S(a^R(t_{j-1}, t_j), t)$ for all $j = 1, \ldots, J$ since (3) holds for each t_j $(j = 1, \ldots, J)$ and $U_{12}^S > 0$. This logic also implies that if any $t < \underline{t}$ has a profitable deviation to some $m \in M_1$, it has a profitable deviation to some $m \in M_1 \cap M^\mu$. Note also that

if any type has a profitable deviation to some $m \in M_t$ where $t < r^S(0)$, then it has a profitable deviation to some $\tilde{m} \in M_{r^S(0)}$, because $\tilde{m}$ is cheaper than m for all types, and they both lead to the same action, $\alpha(m) = \alpha(\tilde{m}) = a^R(0)$.

CLAIM: Assume $\underline{t} = 0$. Then no type has a profitable deviation to any $m \in \bigcup_{t\in[r^S(0),1)} M_t$.

PROOF: By Lemma A.2, type 0 does not have a profitable deviation to any $m \in M \setminus M_1$. So I prove the result by showing that if some type has a profitable deviation to some $m \in \bigcup_{t\in[r^S(0),1)} M_t$, then type 0 has a profitable deviation to m. For this, it suffices to show that if $t > 0$ and $\hat{t} \in [r^S(0), 1)$, then

$$\begin{array}{rcl} U^S\left(a^R(0), t\right) - kC\left(\nu(\hat{t}), t\right) & > & U^S\left(\alpha(\mu(t)), t\right) - kC\left(\nu(1)\left(t\right), t\right) \\ & \Downarrow & \\ U^S\left(a^R(0), 0\right) - kC\left(\nu(\hat{t}), 0\right) & > & U^S\left(\alpha(\mu(0)), 0\right) - kC\left(\nu(1), 0\right). \end{array}$$

Fix a $t > 0$ and $\hat{t} \in [r^S(0), 1)$. Since $\underline{t} = 0$, we have

$$U^S\left(\alpha(\mu(t)), t\right) - kC\left(\nu(1), t\right) \geq U^S\left(\alpha(\mu(0), t\right) - kC\left(\nu(1), t\right),$$

and hence it suffices to show that

$$\begin{aligned} &U^S\left(a^R(0), 0\right) - kC\left(\nu(\hat{t}), 0\right) - \left[U^S\left(\alpha(\mu(0)), 0\right) - kC\left(\nu(1), 0\right)\right] \\ &> U^S\left(a^R(0), t\right) - kC\left(\nu(\hat{t}), t\right) - \left[U^S\left(\alpha(\mu(0)), t\right) - kC\left(\nu(1), t\right)\right]. \end{aligned}$$

This inequality can be rewritten as

$$\int_{a^R(0)}^{\alpha(\mu(0))} \int_0^t U_{12}^S\left(y, z\right) dzdy > k \int_{\nu(\hat{t})}^{\nu(1)} \int_0^t C_{12}\left(y, z\right) dzdy,$$

which holds because $C_{12} < 0 < U_{12}^S$. $\diamond$

CLAIM: Assume $\underline{t} > 0$. Type $\underline{t}$ is indifferent between playing $\mu(t)$ and playing any $m \in M_{\rho^*(t)} \setminus M^\mu$, but strictly prefers $\mu(t)$ to any $m \in M^\mu \setminus \{M_{\rho^*(t)} \cup M_1\}$. Any $t \neq \underline{t}$ strictly prefers $\mu(t)$ to any $m \in M^\mu \setminus M_{\rho(t)}$.

PROOF: The indifference property for type $\underline{t}$ is proved by two observations: if $\rho^*(\underline{t}) < 1$, then it is immediate from Lemma A.2 and that $\alpha(m) = a^R(\underline{t})$ for all $m \in M_{\rho^*(t)}$; on the other hand, if $\rho^*(\underline{t}) = 1$, then for any $m \in M_{\rho^*(t)} \setminus M^\mu$, $\alpha(\mu(t)) = \alpha(m) = a^R(\underline{t}, 1)$ and $\mu(t)$ and m have the same cost. Note that if $\rho^*(\underline{t}) = 1$, Lemma A.2 implies that $U^S(a^R(\underline{t}, t_1), \underline{t}) = U^S(a^R(\underline{t}), \underline{t})$.

Equation (DE) implies that for all $t \in (0, \underline{t})$,

$$U_1^S(a^R(t), t)\frac{da^R}{dt}(t) - kC_1(\nu(\rho^*(t)), t)\nu'(\rho^*(t))\frac{d\rho^*}{dt}(t) = 0.$$

Since $U_{12}^S > 0 > C_{12}$, this implies that for any $\tilde{t} < t \in (0, \underline{t})$,

$$U_1^S(a^R(t), \tilde{t})\frac{da^R}{dt}(t) - kC_1(\nu(\rho^*(t)), \tilde{t})\nu'(\rho^*(t))\frac{d\rho^*}{dt}(t) < 0.$$

Therefore, by continuity, any $t \leq \underline{t}$ strictly prefers $\mu(t)$ to any $m \in \bigcup_{\hat{t}\leq\underline{t}} M_{\rho^*(\hat{t})} \setminus M_1$.

To show that any $t < \underline{t}$ strictly prefers $\mu(t)$ to any $m \in M_1$, pick any such t. The highest utility for t when constrained to sending $m \in M_1$ is attained by sending $\mu(\underline{t})$. From the previous

arguments,

$$U^S(\alpha(\mu(t)),t) - kC(\nu(\rho(t)),t) > U^S(a^R(\underline{t}),t) - kC(\nu(\rho(\underline{t})),t),$$

and hence it suffices to show that

$$U^S(a^R(\underline{t}),t) - kC(\nu(\rho(\underline{t})),t) \geq U^S(\alpha(\mu(\underline{t})),t) - kC(\nu(1),t). \tag{A-3}$$

Since $U^S(\alpha(\mu(\underline{t})),\underline{t}) - kC(\nu(1),\underline{t}) = U^S(a^R(\underline{t}),t) - kC(\nu(\rho^*(\underline{t})),\underline{t})$, (A-3) is true if

$$\begin{aligned} &U^S(\alpha(\mu(\underline{t})),\underline{t}) - kC(\nu(1),\underline{t}) - [U^S(a^R(\underline{t}),t) - kC(\nu(\rho^*(\underline{t})),\underline{t})] \\ &> U^S(a^R(\underline{t}),t) - kC(\nu(\rho(\underline{t})),t) - [U^S(\alpha(\mu(\underline{t})),t) - kC(\nu(1),t)], \end{aligned}$$

which can be rewritten as

$$\int_{a^R(\underline{t})}^{a^R(\underline{t},t_1)} \int_t^{\underline{t}} U_{12}^S(y,z)\,dzdy > k \int_{\nu(\rho^*(\underline{t}))}^{\nu(1)} \int_t^{\underline{t}} C_{12}(y,z)\,dzdy,$$

which holds since $U_{12}^S > 0 > C_{12}$.

The argument for types $t > \underline{t}$ strictly preferring $\mu(t)$ to any $m \in M^\mu \setminus M_1$ is analogous to above. $\diamond$

Claim: Assume $\underline{t} > 0$. No type has a profitable deviation to any $m \in \bigcup_{t\in[\rho^*(\underline{t}),1)} M_t$.

Proof: The result is vacuously true if $\rho^*(\underline{t}) = 1$, so assume that $\rho^*(\underline{t}) < 1$. Any $m \in \bigcup_{t\in[\rho^*(\underline{t}),1)} M_t$ is responded to with $\alpha(m) = a^R(\underline{t})$, hence when constrained to such messages, any type $t \leq \underline{t}$ maximizes utility by sending an $m \in M_{\rho^*(t)}$, because $\rho^*(\underline{t}) > r^S(t)$. However, by the previous Claim, all $t \leq \underline{t}$ weakly prefer playing $\mu(t)$ to any $m \in M_{\rho^*(t)}$. This proves the claim for any $t \leq \underline{t}$.

To prove the claim for all $t > \underline{t}$, it suffices to show that if $t > \underline{t}$ has a profitable deviation to an $m \in \bigcup_{t\in[\rho^*(\underline{t}),1)} M_t$, then type $\underline{t}$ has a profitable deviation to m, since the latter is not possible. To prove this, it suffices to show that for any $t > \underline{t}$ and $\hat{t} \in [\rho^*(\underline{t}),1)$,

$$\begin{array}{rcl} U^S\left(a^R(\underline{t}),t\right) - kC\left(\nu(\hat{t}),t\right) & > & U^S\left(\alpha(\mu(t)),t\right) - kC\left(\nu(1),t\right) \\ & \Downarrow & \\ U^S\left(a^R(\underline{t}),\underline{t}\right) - kC\left(\nu(\hat{t}),\underline{t}\right) & > & U^S\left(\alpha(\mu(\underline{t})),\underline{t}\right) - kC\left(\nu(1),\underline{t}\right). \end{array}$$

Fix a $t > \underline{t}$ and $\hat{t} \in [\rho^*(\underline{t}),1)$. Since $U^S\left(\alpha(\mu(t)),t\right) \geq U^S(\alpha(\mu(\underline{t})),t)$, it suffices to show that

$$\begin{array}{rcl} U^S\left(a^R(\underline{t}),t\right) - kC\left(\nu(\hat{t}),t\right) & > & U^S\left(\alpha(\mu(\underline{t})),t\right) - kC\left(\nu(1),t\right) \\ & \Downarrow & \\ U^S\left(a^R(\underline{t}),\underline{t}\right) - kC\left(\nu(\hat{t}),\underline{t}\right) & > & U^S\left(\alpha(\mu(\underline{t})),\underline{t}\right) - kC\left(\nu(1),\underline{t}\right). \end{array}$$

This is true if

$$\begin{aligned} &U^S\left(a^R(\underline{t}),\underline{t}\right) - kC\left(\nu(\hat{t}),\underline{t}\right) - [U^S\left(\alpha(\mu(\underline{t})),\underline{t}\right) - kC\left(\nu(1),\underline{t}\right)] \\ > \;& U^S\left(a^R(\underline{t}),t\right) - kC\left(\nu(\hat{t}),t\right) - [U^S\left(\alpha(\mu(\underline{t})),t\right) - kC\left(\nu(1),t\right)], \end{aligned}$$

which can be be rewritten as

$$\int_{a^R(\underline{t})}^{\alpha(\mu(\underline{t}))} \int_{\underline{t}}^{t} U_{12}^S(y,z)\,dzdy > k \int_{\nu(\hat{t})}^{\nu(1)} \int_{\underline{t}}^{t} C_{12}(y,z)\,dzdy,$$

which is true because $U_{12}^S > 0 > C_{12}$. $\diamond$

The above claims establish that μ is optimal for the Sender, completing the proof. *Q.E.D.*

Proof of Proposition 1 on page 14. Large bias implies that for any t, $U^S(\cdot, t)$ is strictly increasing on $[a^R(0), a^R(1)]$. Hence there cannot be multiple pools in an LSHP equilibrium because all pools use messages in M_1, imposing the same direct message cost on any type, whereas higher actions are strictly preferred by all types.

Necessity of (7) for $\underline{t} > 0$ is straightforward, since if it does not hold, continuity implies that for small enough $\varepsilon > 0$, either some type $\underline{t} - \varepsilon$ will deviate from $\mu(\underline{t} - \varepsilon)$ to $\mu(\underline{t})$, or $\underline{t}$ will deviate from $\mu(\underline{t})$ to $\mu(\underline{t} - \varepsilon)$. Sufficiency of (7) with $\underline{t} \in (0, \bar{t})$ for an LSHP equilibrium follows from the observation that such a $\underline{t}$ and (trivial) sequence $\langle t_0 = \underline{t}, t_1 = 1\rangle$ satisfies and (4) and (5), and therefore the equilibrium construction of Theorem 2 applies.

Similarly, (8) must hold for an LSHP equilibrium with $\underline{t} = 0$, since otherwise type 0 will deviate to some $m \in r^S(0)$. It is sufficient because then $\underline{t} = 0$ and the (trivial) sequence $\langle t_0 = \underline{t} = 0, t_1 = 1\rangle$ satisfies (4) and (6). *Q.E.D.*

Proof of Proposition 2 on page 15. Chen et al. (2008, Proposition 1) show that there is always a strictly increasing sequence $\langle t_0 = 0, t_1, \ldots, t_J = 1\rangle$ $(J \geq 1)$ such that for $j = 1, \ldots, J$, t_j satisfies (3), and moreover, $U^S(a^R(t_0, t_1), 0) \geq U^S(a^R(0), 0)$. Under the Proposition's assumption, we cannot have $U^S(a^R(t_0, t_1), 0) = U^S(a^R(0), 0)$. This implies that for all k sufficiently small enough, the sequence $\langle t_0 = 0, t_1, \ldots, t_J = 1\rangle$ $(J \geq 1)$ combined with $\underline{t} = 0$ satisfies (4) and (6), and therefore the equilibrium construction of Theorem 2 applies. *Q.E.D.*

The following Lemma is used in the proof of Proposition 3; it shows that when k increases, the separating function decreases pointwise.

Lemma A.3. *Assume $r^S(0) < r^S(1)$. Fix $k_1 < k_2$, and let ρ^1 and ρ^2 denote the separating function when $k = k_1$ and $k = k_2$ respectively, with respective domains $[0, t^1]$ and $[0, t^2]$. Then $t^2 > t^1$ and for all $t \in (0, t^1]$, $\rho^2(t) < \rho^1(t)$.*

Proof. Since $r^S(0) < r^S(1)$, Lemma A.1 implies that $t^1 > 0$ and $t^2 > 0$. It suffices to show that $\rho^2(t) < \rho^1(t)$ for all $t \in (0, t^1]$, because then $\rho^2(t^1) < 1$ and Lemma A.1 implies that $t^2 > t^1$. Suppose, to contradiction, that $\rho^2(\hat{t}) \geq \rho^1(\hat{t})$ for some $\hat{t} \in (0, t^1]$. For any $t > 0$, (DE) implies that if $\rho^1(t) \leq \rho^2(t)$, then $\frac{d\rho^1}{dt}(t) > \frac{d\rho^2}{dt}(t) > 0$. Consequently, we must have $\rho^2(t) > \rho^1(t)$ for all $t \in (0, \hat{t})$. This implies that for all $t \in (0, \hat{t})$, $\frac{d\rho^1}{dt}(t) > \frac{d\rho^2}{dt}(t)$. Reformulating in terms of inverse functions, we have that for all $y \in (r^S(0), \rho(\hat{t}))$, $\frac{d[(\rho^1)^{-1}]}{dy} < \frac{d[(\rho^2)^{-1}]}{dy}(y)$, yet $(\rho^1)^{-1}(r^S(0)) = (\rho^2)^{-1}(r^S(0)) = 0$ and $(\rho^1)^{-1}(\hat{y}) > (\rho^2)^{-1}(\hat{y})$ for $\hat{y} = \rho^1(\hat{t})$. This contradicts the Fundamental Theorem of Calculus applied to $(\rho^1)^{-1}$ and $(\rho^2)^{-1}$. *Q.E.D.*

Proof of Proposition 3 on page 17. (a) Equation (12) and simple algebra shows that $\bar{t} \geq 1 - 4b$ is equivalent to (16). It is obvious that (16) is satisfied if $k \geq \frac{1}{4}$; to see that it is also satisfied if $b \geq \frac{1}{4}$, let $s(k, b) := e^{\frac{k}{b}}(4k - 1)$ and compute that when $k < \frac{1}{4}$, $\frac{\partial s}{\partial b} = e^{\frac{k}{b}}(1 - 4k)\frac{k}{b^2} > 0$, $\frac{\partial s}{\partial k} = \frac{e^{\frac{k}{b}}}{b}(4k + 4b - 1) > 0$, while $s(0, \frac{1}{4}) = -1$.

Now suppose $\bar{t} \geq 1 - 4b$. If $-(\frac{1}{2} - b)^2 + b^2 \geq k$, then $\underline{t} = 0$ and $t_1 = 1$ satisfy (6), and we are done. So assume $-(\frac{1}{2} - b)^2 + b^2 < k$. A little algebra shows that $-(\frac{1-\bar{t}}{2} - b)^2 + b^2 \geq 0$. By continuity, there exists $\underline{t} \in (0, \bar{t}]$ that satisfies (14). This $\underline{t}$ and $t_1 = 1$ satisfy (5), and we are done.

Now suppose $\bar{t} < 1 - 4b$. Simple algebra shows that for any $t \in [0, \bar{t}]$, $-(\frac{1-t}{2} - b)^2 + b^2 < 0$, hence (14) cannot be satisfied with any $\underline{t} \in (0, \bar{t}]$ and $t_1 = 1$. So all single-pool LSHP equilibria must have $\underline{t} = 0$, inducing the Receiver to play $\frac{1}{2}$ in equilibrium. Since $\bar{t} < 1 - 4b$ requires $b < \frac{1}{4}$, type 0 strictly prefers any action in $[0, \frac{1}{2})$ to $\frac{1}{2}$. Consider any $m \in M_{1-\varepsilon}$ for $\varepsilon < \frac{1}{2}$. If $\alpha(m) \leq \frac{1}{2}$, type 0 has a profitable deviation to m since it weakly prefers $\alpha(m)$ to action $\frac{1}{2}$ and strictly saves on lying cost. So, $\alpha(m) > \frac{1}{2}$. But then, type $1 - \varepsilon > \frac{1}{2}$ has a profitable deviation to m because it strictly prefers $\alpha(m)$ to $\frac{1}{2}$ and saves on lying cost. Consequently, there are no LSHP equilibria.

(b) A single-pool LSHP equilibrium with $\underline{t} = 0$ exists if $b \geq k + \frac{1}{4}$ because in this case, $-(\frac{1}{2} - b)^2 + b^2 \geq k$, hence $\underline{t} = 0$ and t_1 satisfy (6). If $b < \frac{1}{4}$, then the argument used in part (a) to show that there is no uninformative LSHP equilibrium applies.

So it remains to show that there is no uninformative LSHP equilibrium when $b \in [\frac{1}{4}, k + \frac{1}{4})$. Suppose, to contradiction, that there is. Consider any $m \in M_0$. Let $a := \alpha(m)$ be the Receiver's response to m. Type 0 not deviating to m requires $-(\frac{1}{2} - b)^2 - k \geq -(a - b)^2$. Since $a \in [0, 1]$, we must have

$$1 \geq a \geq b + \sqrt{k + \left(\frac{1}{2} - b\right)^2}, \tag{A-4}$$

which implies that $k \leq \frac{1}{2}$, since $b \geq \frac{1}{4}$. On the other hand, type 1 not deviating to m requires $-(\frac{1}{2} - 1 - b)^2 \geq -(a - 1 - b)^2 - k$, which implies that

$$a \leq 1 + b - \sqrt{\left(b + \frac{1}{2}\right)^2 - k}. \tag{A-5}$$

We have a contradiction between (A-4) and (A-5) if

$$b + \sqrt{\left(k + (\frac{1}{2} - b\right)^2} > 1 + b - \sqrt{\left(b + \frac{1}{2}\right)^2 - k},$$

which simplifies after some algebra to $k < 2b$, which holds since $k \leq \frac{1}{2}$ and $b \geq \frac{1}{4}$.

(c) If $\bar{t} < 1 - 4b$, the statement is vacuously true because there are no single-pool LSHP equilibria by part (a). If $\bar{t} = 1 - 4b$, simple algebra shows that $-(\frac{1-t}{2} - b)^2 + b^2 < 0$ for all $t \in [0, \bar{t})$. By part (b), there is no single-pool LSHP equilibrium with $\underline{t} = 0$, and by the necessity of (14) for any LSHP equilibrium, there is no single-pool LSHP equilibrium with $\underline{t} \in (0, \bar{t})$ either. Therefore, by part (a), all LSHP equilibria have cutoff $\underline{t} = \bar{t}$.

It remains to consider $\bar{t} > 1 - 4b$. Define

$$v(t) := -\left(\frac{1-t}{2} - b\right)^2 + b^2, \tag{A-6}$$

$$w(t) := k\left((1-t)^2 - (\rho^*(t) - t)^2\right). \tag{A-7}$$

By (14), there is a single-pool LSHP equilibrium with strictly positive cutoff if and only if $v(\underline{t}) = w(\underline{t})$ for some $\underline{t} > 0$; by part (b), there is one with cutoff $\underline{t} = 0$ if and only if $v(0) \geq w(0)$.

Taking derivatives yields $v'(t) = -\left(\frac{1-t}{2} - b\right)$, $v''(t) = -1$, $w'(t) = -2k\left(1 - \rho^*(t) + \frac{b}{k}\right) < 0$, and $w''(t) = \frac{2b}{\rho(t)-t} > 0$, where the derivatives of $w(\cdot)$ have been simplified using the fact that ρ^* solves (9). Thus, $v''(t) - w''(t) < 0$. Since $\bar{t} > 1 - 4b$ implies $v(\bar{t}) - w(\bar{t}) > 0$, we conclude that there is at most one solution to $v(t) - w(t) = 0$ on the domain $t \in [0, \bar{t}]$. Moreover, $v(0) - w(0) \geq 0$ if and only if there is no solution to $v(t) - w(t) = 0$ with $t \in (0, \bar{t}]$, yielding the the desired conclusion.

(d) When $b \geq \frac{1}{4}$, there cannot be multiple pools in an LSHP equilibrium because (13) will be violated. Furthermore, all single-pool LSHP equilibria are essentially equivalent because $b \geq \frac{1}{4}$ implies $\bar{t} > 1 - 4b$, and the rest of the argument of part (c) applies.

(e) Fix any $b > 0$ any $\hat{k}$ such that (16) is satisfied. In what follows, I add k as an argument to various objects to emphasize the dependence on the cost intensity. First consider $b > \hat{k} + \frac{1}{4}$. By part (b), $\underline{t}(k) = 0$ for all k in a neighborhood of $\hat{k}$, hence $\underline{t}'(\hat{k}) = 0$.

Now consider $b < \hat{k} + \frac{1}{4}$. Then $v(\underline{t}(k)) - w(\underline{t}(k); k) = 0$ for all k in a neighborhood of $\hat{k}$, where v and w are defined in equations (A-6) and (A-7) (note that $v(\cdot)$ does not depend on k). It is routine to verify that for any t, $\rho^*(t; \cdot)$ is continuous on the domain of k for which $\rho^*(t; k)$ is well-defined. This implies that for any t, $w(t; \cdot)$ is continuous on the domain of k for which it is defined. Hence $\underline{t}(\cdot)$ is continuous at $\hat{k}$. Next I argue that it is strictly increasing and differentiable at $\hat{k}$. By the Implicit Function Theorem,

$$\underline{t}'(\hat{k}) \quad = \quad \frac{\frac{\partial w(\underline{t}(\hat{k}); \hat{k})}{\partial k}}{\frac{\partial [v(\underline{t}(\hat{k})) - w(\underline{t}(\hat{k}); \hat{k})]}{\partial t}}, \tag{A-8}$$

so long as the denominator is non-zero, which will be shown below. The numerator of the right hand side of (A-8) computes as

$$-2k(\rho^*(\underline{t}(\hat{k}); \hat{k}) - \underline{t}(\hat{k}))\frac{\partial \rho^*(\underline{t}(\hat{k}); \hat{k})}{\partial k} + \left((1 - \underline{t}(\hat{k}))^2 - (\rho^*(\underline{t}(\hat{k}); k) - \underline{t}(\hat{k}))^2\right),$$

which is strictly positive because $1 \geq \rho^*(t) > t$ and $\frac{\partial \rho^*(\underline{t}(\hat{k}); \hat{k})}{\partial k} < 0$ by Lemma A.3 (that the derivative exists is routinely verified). To sign the denominator, note that $\frac{\partial^2 [v(t) - w(t; \hat{k})]}{\partial t^2} < 0$ for all t by the calculation performed in part (c). Hence, if $\frac{\partial [v(\underline{t}(\hat{k})) - w(\underline{t}(\hat{k}); \hat{k})]}{\partial t} \leq 0$, we would have $v(\bar{t}(\hat{k})) - w(\bar{t}(\hat{k}); \hat{k}) < 0$ because $\bar{t}(\hat{k}) > \underline{t}(\hat{k}) = 0$, which implies that $v(\bar{t}(\hat{k})) < 0$ since $w(\bar{t}(\hat{k})) = 0$. But this is impossible because by (A-6), $v(t) \geq 0$ for all $t \in [1 - 4b, 1)$, and $\bar{t}(\hat{k}) \geq 1 - 4b$ by part (a). Consequently, the denominator of the right hand side of (A-8) is strictly positive. Therefore, $\underline{t}'(\hat{k}) > 0$.

Finally, consider $b = \hat{k} + \frac{1}{4}$. Then by part (b), $\underline{t}(\hat{k}) = 0$, hence $\underline{t}(\cdot)$ is left-continuous at $\hat{k}$. Since $v(\underline{t}(k)) - w(\underline{t}(k); k) = 0$ holds for all $k \geq \hat{k}$, the continuity argument used above implies that $\underline{t}(\cdot)$ is also right-continuous at $\hat{k}$. *Q.E.D.*

Proof of Proposition 5 on page 21. Since on-path behavior in an augmented LSHP profile is identical to an LSHP equilibrium, I only need to verify that deviating to some $m \in M_\phi$ is not profitable for any type. Fix any type t and some $m \in M_\phi$. Since the (non-augmented) LSHP profile is an equilibrium, it is not profitable for type t to deviate some unused $\tilde{m} \in r^S(t)$, which induces an action $\alpha(r^S(t)) \geq a^R(0)$. But then, since all $\tilde{m} \in r^S(t)$ cost the same as $m \in M_\phi$, and

by large bias, $U^S(\alpha(r^S(t)),t) \geq U^S(a^R(0),t)$, deviating to m is not profitable for type t. *Q.E.D.*

Proof of Corollary 1 on page 22. Since bias is large, Proposition 1 implies that if $\underline{t} > 0$, (7) must hold. But the left-hand side of (7) is bounded above zero on the domain $\underline{t} \in [0, \overline{t}]$ for any set of k that is bounded away from ∞; whereas the right-hand side is converging to 0 as $k \to 0$. Hence, for small enough k, there is no solution to (7) in the domain $\underline{t} \in [0, \overline{t}]$, which proves the first part of the Proposition.

The second part follows from the discussion in Section 3.6, noting that $\inf\{t : r^S(1) = 1\} = 1$ since $r^S(t) = t$ in the withholding model. *Q.E.D.*

Proof of Proposition 4 on page 19. I will show that for $k \geq \frac{1}{4}$ and $b \in (0, \frac{3}{16})$, there is an equilibrium under communication in which the Receiver's ex-ante utility is strictly higher than $-b^2$, which is her utility from delegation. This proves the Proposition because by the discussion in Section 3.6, for all $b \in \left[\frac{3}{16}, \sqrt{\frac{1}{12}}\right]$, communication is arbitrarily close to fully-revealing and hence dominates delegation for all large enough k. For all $b > \sqrt{\frac{1}{12}}$, communication dominates delegation for any k, because even an uninformative equilibrium under communication yields ex-ante utility to the Receiver strictly greater than $-b^2$.

Accordingly, assume $k \geq \frac{1}{4}$ and $b \in (0, \frac{3}{16})$. By Proposition 3, there are single-pool LSHP equilibria for these parameters. R's utility in a single-pool LSHP equilibrium with cutoff $\underline{t}$ is $-\int_{\underline{t}}^1 \left(\frac{1+\underline{t}}{2} - t\right)^2 dt$, since all types below $\underline{t}$ separate and all types above $\underline{t}$ form a single pool. Hence this equilibrium yields higher utility than delegation if

$$\frac{1}{4}\underline{t} - \frac{1}{4}\underline{t}^2 + \frac{1}{12}\underline{t}^3 - \frac{1}{12} > -b^2,$$

or equivalently,

$$\underline{t} > 1 - \left(\sqrt[3]{12}\right) b^{\frac{2}{3}}. \tag{A-9}$$

Proposition 3 implies that $\underline{t} \geq 1 - 4b$ (because $-\left(\frac{1-t}{2} - b\right)^2 + b^2 < 0$ for all $t < 1 - 4b$), hence (A-9) holds if

$$1 - 4b > 1 - \left(\sqrt[3]{12}\right) b^{\frac{2}{3}},$$

which is true because $b \in (0, \frac{3}{16})$. *Q.E.D.*

Appendix B: Equilibrium Refinement

This appendix develops a rigorous justification for focussing on LSHP equilibria, by applying an equilibrium refinement proposed by Bernheim and Severinov (2003) called the *monotonic D1* (mD1) criterion. The first part below explains and formalizes the criterion; the second part applies it to the current model, deriving the sense in which it justifies LSHP equilibria.

The Monotonic D1 Criterion. The underlying idea of the mD1 criterion is the same as Cho and Kreps's (1987) D1 criterion, which requires that the Receiver not attribute a deviation to a particular type if there is some other type that is willing to make the deviation for a strictly larger set of possible Receiver responses. The mD1 criterion strengthens this by applying the test to only those responses from the Receiver which satisfy an *action monotonicity* restriction. Analogous to message monotonicity for the Sender, action monotonicity requires the Receiver's strategy to satisfy $\alpha(m) \geq \alpha(m')$ if $\Psi(m) > \Psi(m')$. In words, if the Sender claims to be a strictly higher type, the Receiver should respond with a weakly higher action. Intuitively, given message monotonicity, action monotonicity is natural (and indeed, is an implication of message monotonicity on the equilibrium path).

Some notation is needed for the formal statement. With respect to a given profile (μ, α), which induces some $\rho = \Psi \circ \alpha$, define

$$\begin{aligned} \xi_l\left(\tilde{t}\right) &:= \max\{a^R(0), \sup_{t:\rho(t)\leq\tilde{t}} \alpha\left(\mu(t)\right)\}, \\ \xi_h\left(\tilde{t}\right) &:= \min\{a^R(1), \inf_{t:\rho(t)\geq\tilde{t}} \alpha\left(\mu(t)\right)\}, \end{aligned}$$

where I follow the convention that $\sup(\emptyset) = +\infty$ and $\inf(\emptyset) = -\infty$. To understand these functions, suppose (μ, α) is an equilibrium with message and action monotonicity. If some type claims to be $\tilde{t}$ on the equilibrium path, then $\xi_l(\tilde{t})$ (resp. $\xi_h(\tilde{t})$) is just the "lowest" (resp. "highest") equilibrium action taken in response to a claim of $\tilde{t}$. If no type claims to be $\tilde{t}$, yet some type does claim to be a $t < \tilde{t}$ (resp. $t > \tilde{t}$), then $\xi_l\left(\tilde{t}\right)$ (resp. $\xi_h\left(\tilde{t}\right)$) is the "highest" (resp. "lowest") action taken by the Receiver in equilibrium to a claim of being lower (resp. higher) than $\tilde{t}$. If no type claims to be any $t \leq \tilde{t}$ (resp. $t \geq \tilde{t}$), then $\xi_l\left(\tilde{t}\right)$ (resp. $\xi_h\left(\tilde{t}\right)$) just specifies the lowest (resp. highest) rationalizable action for the Receiver.

With respect to some profile (μ, α), let

$$A(m,t) := [\xi_l(\Psi(m)), \xi_h(\Psi(m))] \cap \{a : U^S(a,t) - kC(\nu(\Psi(m)), t) \geq U^S(\alpha(\mu(t)), t) - kC(\nu(\rho(t)), t)\},$$
$$\overline{A}(m,t) := [\xi_l(\Psi(m)), \xi_h(\Psi(m))] \cap \{a : U^S(a,t) - kC(\nu(\Psi(m)), t) > U^S(\alpha(\mu(t)), t) - kC(\nu(\rho(t)), t)\}.$$

To interpret, consider an unused message m in some equilibrium. $A(m,t)$ (resp. $\overline{A}(m,t)$) is the set of responses *within the set* $[\xi_l(\Psi(m)), \xi_h(\Psi(m))]$ that give type t a weak (resp. strict) incentive to deviate to m.

The following is Bernheim and Severinov's (2003) refinement criterion adapted to the current model.

Definition B.1. An equilibrium, (μ, α), satisfies the mD1 criterion if it satisfies (i) message monotonicity, (ii) action monotonicity, and (iii) $\alpha(m) = a^R(t')$ for any t' and any out-of-equilibrium message m such that $A(m,t) \subseteq \overline{A}(m,t')$ for all $t \neq t'$ and $A(m,t') \neq \emptyset$.

The first two parts of the definition are clear. In the third part, the requirement that $\alpha(m) = a^R(t')$ could alternatively be posed as $support[G(\cdot|m)] = \{t'\}$. If we replace $[\xi_l(\Psi(m)), \xi_h(\Psi(m))]$ in the definitions of A and $\overline{A}$ with $\left[a^R(0), a^R(1)\right]$, then the above test is basically the D1 criterion (cf. Cho and Kreps, 1987, p. 205). However, given action monotonicity, any out-of-equilibrium message m satisfies $\alpha(m) \in [\xi_l(\Psi(m)), \xi_h(\Psi(m))]$. Accordingly, the definition above applies the idea behind the D1 criterion on the restricted action space $[\xi_l(\Psi(m)), \xi_h(\Psi(m))]$. That is, it requires that for some out-of-equilibrium message m, if there is some type t' that would *strictly* prefer to deviate to m for any response $a \in [\xi_l(\Psi(m)), \xi_h(\Psi(m))]$ that a type $t \neq t'$ would *weakly* prefer to deviate for, then upon observing the deviation m, the Receiver should believe it is type t', so long as there is some response in $a \in [\xi_l(\Psi(m)), \xi_h(\Psi(m))]$ such that t' would at least weakly prefer sending m to its equilibrium payoff.[33]

The Connection to LSHP Equilibria. The mD1 criterion provides a justification for focussing on LSHP equilibria in the following sense: any mD1 equilibrium is an LSHP equilibrium, modulo the behavior of the cutoff type; conversely, there is always an LSHP equilibrium that is identical to some mD1 equilibrium. Formally,

Theorem B.1. *(1) If (μ, α) is an mD1 equilibrium, then there is an LSHP equilibrium, $(\tilde{\mu}, \tilde{\alpha})$ such that for all but at most a single t, $\tilde{\mu}(t) = \mu(t)$ and $\tilde{\alpha}(\tilde{\mu}(t)) = \alpha(\mu(t))$.*

(2) There is an LSHP equilibrium that is also an mD1 equilibrium.

Remark B.1. *It is possible to further show that for any LSHP equilibrium, (μ, α), that satisfies action monotonicity, there is some mD1 equilibrium, (μ, α'), such that $\alpha' \circ \mu = \alpha \circ \mu$, i.e. there are only off-the-equilibrium-path differences if any. A proof is available on request. Note that the existence proof for LSHP equilibria (Theorem 2) constructs an LSHP equilibrium that satisfies action monotonicity.*

The proof of Theorem B.1 requires a series of lemmata that derive necessary conditions for an mD1 equilibrium. Some notation is helpful, where all of the following are defined with implicit respect to some strategy, μ. Let $\mu^{-1}(m) := \{t : \mu(t) = m\}$ and $\Gamma(t) := \bigcup_{m \in M_t} \mu^{-1}(m)$. Hence, $\Gamma(t)$ is the set of types who claim to be type t, i.e. who use messages in M_t. To say that there is a pool on message m is to say that $|\mu^{-1}(m)| > 1$, type t is separating if $|\mu(t)| = 1$, and a message m is used if $|\mu^{-1}(m)| > 0$. Next, define $t_l(m) := \inf\{t : \rho(t) = \Psi(m)\}$ and $t_h(m) := \sup\{t : \rho(t) = \Psi(m)\}$. In words, $t_l(m)$ identifies the "lowest" type who claims to be type $\Psi(m)$, and similarly for $t_h(m)$. Note that under message monotonicity, $t_l(m) < t_h(m)$ implies that any type $t \in (t_l(m), t_h(m))$ satisfies $\mu(t) \in M_{\Psi(m)}$, although not necessarily $\mu(t) = m$.

Lemma B.1. *In any monotone equilibrium, for any message m,*

1. *$\mu^{-1}(m)$ and $\Gamma(\Psi(m))$ are convex sets, and $|\Gamma(\Psi(m)| \in \{0, 1, \infty\}$.*
2. *If $|\Gamma(\Psi(m))| > 0$, there is a strictly increasing sequence, $\langle t_l(m) = t_0, t_1, \ldots, t_J = t_h(m)\rangle$ and a set of distinct messages $\{m_j\}_{j=1}^J \subseteq M_{\Psi(m)}$ such that (i) for all $j = 1, \ldots, J-1$, t_j solves (3), (ii) for any $j = 1, \ldots, J$ and $t \in (t_{j-1}, t_j)$, $\mu(t) = m_j$, and $\alpha(\mu(t)) = a^R(t_{j-1}, t_j)$.*

Proof. Just apply the logic of CS. *Q.E.D.*

[33] The caveat that $A(m, t') \neq \emptyset$ prevents internal inconsistencies in the criterion when there exists m, t, and t' such that $A(m, t) = A(m, t') = \emptyset$.

Lemma B.2. *In any monotone equilibrium, if there is pooling on some message $m_p \notin M_1$, then there exists $\eta > 0$ such that every $m \in \bigcup_{t\in(\Psi(m_p),\Psi(m_p)+\eta)} M_t$ is unused.*

Hence, if there is pooling on some message, then there are some unused messages "just above."

Proof. Suppose there is pooling on $m_p \notin M_1$. To minimize notation in this proof, write t_h instead of $t_h(m_p)$. If $\rho(t_h) > \Psi(m_p)$, then by message monotonicity, we are done, since all messages in $\bigcup_{(\Psi(m_p),\Psi(\alpha(t_h)))} M_t$ are unused. So assume that $\rho(t_h) = \Psi(m_p)$. If $t_h = 1$, we are done since all messages in $\bigcup_{(\Psi(m_p),1)} M_t$ are unused, by message monotonicity. So assume $t_h < 1$. Let $m_h := \mu(t_h)$. We must have $|\mu^{-1}(m_h)| > 1$, for if not, t_h is separating, and a slightly lower type would strictly prefer to mimic type t_h, contradicting equilibrium. It follows that $\alpha(m_h) < a^R(t_h)$. Let $\tilde{t} := \lim_{t\downarrow t_h} \rho(t)$. ($\rho'$ is well-defined by message monotonicity, though it may not be used.)

Claim: $\tilde{t} > \Psi(m_p)$.

Proof: Suppose not. By message monotonicity, $\tilde{t} = \Psi(m_p) = \rho(t_h)$. Note that $\rho(\cdot)$ is then continuous at t_h. Since $\rho(t) > \Psi(m_p)$ for all $t > t_h$, it follows that ρ is strictly increasing on $(t_h, t_h + \delta)$ for some $\delta > 0$, i.e. all types in this interval are separating. Hence, $a_\varepsilon := \alpha(\mu(t_h + \varepsilon)) = a^R(t_h + \varepsilon)$ for small enough $\varepsilon > 0$. By picking $\varepsilon > 0$ small enough, we can make $C(\nu(\rho(t_h + \varepsilon)), t_h) - C(\nu(\rho(t_h)), t_h)$ arbitrarily close to 0, whereas $U^S(a_\varepsilon, t_h) - U^S(\alpha(m_h), t_h)$ is positive and bounded away from 0, because $\alpha(m_h) < a^R(t_h) < a_\varepsilon < a^S(t_h)$. Therefore, for small enough $\varepsilon > 0$, t_h prefers to imitate $t_h + \varepsilon$, contradicting equilibrium. $\diamond$

This completes the argument because messages in $\bigcup_{t\in(\Psi(m_p),\tilde{t})} M_t$ are unused. *Q.E.D.*

Lemma B.3. *In any mD1 equilibrium, if $|\Gamma(\tilde{t})| = 0$ for some $\tilde{t} > \rho(0)$, then for all $m \in M_{\tilde{t}}$, $\alpha(m) = a^R(\sup\{t : \rho(t) < \tilde{t}\})$.*

Hence, if no type claims to be $\tilde{t} > \rho(0)$, then the Receiver must respond to any claim of being $\tilde{t}$ by inferring that the Sender is the "highest type who claims to be below t." The proof shows that it is this type who has the biggest incentive to claim to be $\tilde{t}$ (given action monotonicity for the Receiver).

Proof. Fix a $\tilde{t} > \rho(0)$ such that $|\Gamma(\tilde{t})| = 0$. Let $\hat{t} := \sup\{t : \rho(t) < \tilde{t}\}$ and, for any t, $a_t := \alpha(\mu(t))$. There are two distinct cases: either $\hat{t} < 1$, or $\hat{t} = 1$.

Case 1: $\hat{t} < 1$

Let $t^+ := \inf_{t>\hat{t}} \rho(t)$, $t^- := \sup_{t<\hat{t}} \rho(t)$, $a^+ := \xi_h(\tilde{t})$, and $a^- := \xi_l(\tilde{t})$. Message monotonicity implies that $\rho(\hat{t}), \tilde{t} \in [t^-, t^+]$.

Claim: The inequalities below, (B-1) and (B-2), hold for all t, with equality for $t = \hat{t}$:

$$U^S(a_t, t) - kC(\nu(\rho(t)), t) \geq U^S(a^-, t) - kC(\nu(t^-), t), \tag{B-1}$$

$$U^S(a_t, t) - kC(\nu(\rho(t)), t) \geq U^S(a^+, t) - kC(\nu(t^+), t). \tag{B-2}$$

Proof: I prove it for (B-1); it is analogous for (B-2). Suppose first that $\rho(\hat{t}) > \tilde{t}$. Then, by message and action monotonicity, $a^- = \lim_{t\uparrow\hat{t}} a_t$ and $(a_t, \rho(t)) \uparrow (a^-, t^-)$ as $t \uparrow \hat{t}$. Continuity of U^S, C, and ν imply that any t for which (B-1) does not hold has a profitable

deviation to $\mu(\hat{t}-\varepsilon)$ for small enough $\varepsilon>0$. For type $\hat{t}$, suppose towards contradiction that (B-1) holds strictly. Continuity of U^S, C, and ν then imply that for all sufficiently small $\varepsilon>0$, $U^S\left(a_{\hat{t}},\hat{t}-\varepsilon\right)-kC\left(\nu(\rho\left(\hat{t}\right)),\hat{t}-\varepsilon\right)>U^S\left(a_{\hat{t}-\varepsilon},\hat{t}-\varepsilon\right)-kC\left(\nu(\rho(\hat{t}-\varepsilon)),\hat{t}-\varepsilon\right)$, which contradicts optimality of $\mu(\hat{t}-\varepsilon)$.

Now suppose that $\rho(\hat{t})<\tilde{t}$. Then $a^-=a_{\hat{t}}$ and $t^-=\rho(\hat{t})$, and it is obvious that (B-1) holds with equality for $\hat{t}$. Moreover, any t for which (B-1) does not hold has a profitable deviation to $\mu(\hat{t})$. $\diamond$

The following two claims show that the mD1 criterion yields the desired conclusion.

Claim: For all $m\in M_{\tilde{t}}$, $A(m,\hat{t})\neq\emptyset$.

Proof: Pick any $m\in M_{\tilde{t}}$. If $\tilde{t}\leq r^S(0)$, then since (B-1) holds with equality for $\hat{t}$, $a^-\in A(m,\hat{t})$. If $\tilde{t}>r^S(0)$, then since (B-2) hold with equality for $\hat{t}$, $a^+\in A(m,\hat{t})$. $\diamond$

Claim: For all $m\in M_{\tilde{t}}$ and $t\neq\hat{t}$, $A(m,t)\subseteq\overline{A}(m,\hat{t})$.

Proof: Fix any $m\in M_{\tilde{t}}$ and $t\neq\hat{t}$. I must show that $\forall a\in[a^-,a^+]$,

$$\begin{array}{rcl} U^S\left(a,t\right)-kC\left(\nu(\tilde{t}),t\right) & \geq & U^S\left(a_t,t\right)-kC\left(\nu(\rho\left(t\right)),t\right) \\ & \Downarrow & \\ U^S\left(a,\hat{t}\right)-kC\left(\nu(\tilde{t}),\hat{t}\right) & > & U^S\left(a_{\hat{t}},\hat{t}\right)-kC\left(\nu(\rho\left(\hat{t}\right)),\hat{t}\right). \end{array}$$

I provide the argument assuming $t<\hat{t}$; it is analogous for $t>\hat{t}$ (using (B-2) instead of (B-1)). Since (B-1) holds for all t, with equality for $\hat{t}$, it suffices to show that

$$\begin{array}{rcl} U^S\left(a,t\right)-kC\left(\nu(\tilde{t}),t\right) & \geq & U^S\left(a^-,t\right)-kC\left(\nu(t^-),t\right) \\ & \Downarrow & \\ U^S\left(a,\hat{t}\right)-kC\left(\nu(\tilde{t}),\hat{t}\right) & > & U^S\left(a^-,\hat{t}\right)-kC\left(\nu(t^-),\hat{t}\right). \end{array}$$

This is true if

$$\begin{aligned} &U^S\left(a,\hat{t}\right)-kC\left(\nu(\tilde{t}),\hat{t}\right)-\left[U^S\left(a^-,\hat{t}\right)-kC\left(\nu(t^-),\hat{t}\right)\right] \\ &>U^S\left(a,t\right)-kC\left(\nu(\tilde{t}),t\right)-\left[U^S\left(a^-,t\right)-kC\left(\nu(t^-),t\right)\right], \end{aligned}$$

which can be be rewritten as

$$\int_{a^-}^{a}\int_{t}^{\hat{t}}U_{12}^S\left(y,z\right)dzdy>k\int_{\nu(t^-)}^{\nu(\tilde{t})}\int_{t}^{\hat{t}}C_{12}\left(y,z\right)dzdy,$$

which is true because $U_{12}^S>0>C_{12}$. $\diamond$

Case 2: $\hat{t}=1$

It needs to be shown that for any $m\in M_{\tilde{t}}$, $\alpha(m)=a^R(1)$. If $a_1=a^R(1)$, this is an immediate consequence of action monotonicity, so assume that $a_1<a^R(1)$. If $\rho\left(1\right)>\tilde{t}$, then the same proof as in Case 1 works, except that one now defines $t^+:=\rho\left(1\right)$. So consider $\rho\left(1\right)<\tilde{t}$. Then $a^-=a_1$ and $a^+=a^R(1)$. Since $a^+\in A(m,1)$, the following claim shows that the mD1 criterion yields the desired conclusion.

Claim: For all $t<1$ and $m\in M_{\tilde{t}}$, $A(m,t)\subseteq\overline{A}(m,1)$.

PROOF: Fix any $t<1$ and $m\in M_{\tilde{t}}$. I must show that $\forall a\in\left[a_1,a^R(1)\right]$,

$$\begin{array}{rcl} U^S(a,t)-kC\left(\nu(\tilde{t}),t\right) & \geq & U^S(a_t,t)-kC\left(\nu(\rho(t)),t\right) \\ & \Downarrow & \\ U^S(a,1)-kC\left(\nu(\tilde{t}),1\right) & > & U^S(a_1,1)-kC\left(\nu(\rho(1)),1\right). \end{array}$$

Since optimality of $\mu(t)$ implies $U^S(a_t,t)-kC(\rho(t),t)\geq U^S(a_1,t)-kC(\rho(1),t)$, it suffices to show that

$$\begin{aligned} &U^S(a,1)-kC\left(\nu(\tilde{t}),1\right)-\left[U^S(a_1,1)-kC\left(\nu(\rho(1)),1\right)\right] \\ &>U^S(a,t)-kC\left(\nu(\tilde{t}),t\right)-\left[U^S(a_1,t)-kC\left(\nu(\rho(1)),t\right)\right], \end{aligned}$$

which can be rewritten as

$$\int_{a_1}^{a}\int_{t}^{1}U_{12}^S(y,z)\,dzdy>k\int_{\nu(\rho(1))}^{\nu(\tilde{t})}\int_{t}^{1}C_{12}(y,z)\,dzdy.$$

This inequality holds because $U_{12}^S>0>C_{12}$. $\diamond$ *Q.E.D.*

Lemma B.4. *In any mD1 equilibrium, if $\rho(0)>r^S(0)$, then $\alpha(m)=a^R(0)$ for any m such that $\Psi(m)\in[0,\rho(0))$.*

Hence, if the lowest type is using inflated language, then all unused messages at the bottom must be attributed by the Receiver to type 0.

Proof. Assume $r^S(0)<\rho(0)$. It suffices to show that $\alpha(m)=a^R(0)$ for any m such that $\Psi(m)\in[r^S(0),\rho(0))$, because action monotonicity then implies that $\alpha(m)=a^R(0)$ for any m such that $\Psi(m)\in[0,r^S(0))$. Accordingly, pick any m such that $\Psi(m)\in[r^S(0),\rho(0))$. For any t, let $a_t:=\alpha(\mu(t))$. It follows from message and action monotonicity that $\xi_l(\Psi(m))=a^R(0)$ and $\xi_h(\Psi(m))=a_0$. Plainly, $a_0\in A(m,0)$, so the mD1 criterion yields the result if for all $t>0$, $A(m,t)\subseteq\overline{A}(m,0)$. Equivalently, I must show that $\forall a\in\left[a^R(0),a_0\right]$ and $\forall t>0$,

$$\begin{array}{rcl} U^S(a,t)-kC\left(\nu(\Psi(m)),t\right) & \geq & U^S(a_t,t)-kC\left(\nu(\rho(t)),t\right) \\ & \Downarrow & \\ U^S(a,0)-kC\left(\nu(\Psi(m)),0\right) & > & U^S(a_0,0)-kC\left(\nu(\rho(0)),0\right). \end{array}$$

That (μ,α) is an equilibrium implies

$$U^S(a_t,t)-kC\left(\nu(\rho(t)),t\right)\geq U^S(a_0,t)-kC\left(\nu(\rho(0)),t\right),$$

and hence it suffices to show that

$$\begin{aligned} &U^S(a,0)-kC\left(\nu(\Psi(m)),0\right)-\left[U^S(a_0,0)-kC\left(\nu(\rho(0)),0\right)\right] \\ &>U^S(a,t)-kC\left(\nu(\Psi(m)),t\right)-\left[U^S(a_0,t)-kC\left(\nu(\rho(0)),t\right)\right]. \end{aligned}$$

This inequality can be rewritten as

$$\int_{a}^{a_0}\int_{0}^{t}U_{12}^S(y,z)\,dzdy>k\int_{\nu(\Psi(m))}^{\nu(\rho(0))}\int_{0}^{t}C_{12}(y,z)\,dzdy,$$

which holds because $C_{12}<0<U_{12}^S$. *Q.E.D.*

Lemma B.5. *In any mD1 equilibrium, there is a cutoff type $\underline{t}\in[0,\bar{t}]$ such that all types $t<\underline{t}$ are separating with $\rho(t)=\rho^*(t)$, while all $\mu(t)\in M_1$ for any $t>\underline{t}$.*

Hence, the Sender's behavior in an mD1 equilibrium is an LSHP strategy, modulo perhaps the behavior of the cutoff type.

Proof. I prove the result in two steps.

CLAIM: There is a cutoff type $\underline{t} \in [0,1)$ such that all types below are separating and all types above are sending messages in M_1.

PROOF: We know there must be some pooling, by Theorem 1. So it suffices to prove that $|\Gamma(m)| > 1$ implies $m \in M_1$, because then $\underline{t} = \inf\{t : \rho(t) = 1\}$ satisfies the needed properties for a cutoff type. Suppose there is some m_p such that $\Psi(m_p) < 1$ and $|\Gamma(m_p)| > 1$. Let $\hat{a} := \lim_{\varepsilon \to 0} \alpha(\mu(t_h(m_p) - \varepsilon))$. By Lemma B.1, for all small enough $\varepsilon > 0$, $\mu(t_h(r_p) - \varepsilon) \in M_{\Psi(m_p)}$, $\alpha(\mu(t_h(m_p) - \varepsilon) = \hat{a}$, and $\hat{a} < a^R(t_h(m_p))$. By Lemma B.2, $\exists \eta > 0$ such that for any m with $\Psi(m) \in (\Psi(m_p), \Psi(m_p) + \eta)$, $|\Gamma(m)| = 0$. By Lemma B.3, for any m with $\Psi(m) \in (\Psi(m_p), \Psi(m_p) + \eta)$, $\alpha(m) = a^R(t_h(r_p))$. It follows that for small enough $\varepsilon > 0$ and $\delta > 0$, a type $t_h(r_p) - \varepsilon$ strictly prefers to send message any $m \in M_{\Psi(m_p)+\delta}$ rather than playing $\mu(t_h(r_p) - \varepsilon)$, a contradiction. $\diamond$

CLAIM: $\underline{t} \leq \bar{t}$ and $\rho(t) = \rho^*(t)$ for all $t < \underline{t}$.

PROOF: Assume $\underline{t} > 0$, for there is nothing to prove otherwise. Since all types in $[0, \underline{t})$ are separating, Lemma 1 implies that $\rho(t) > r^S(t)$ for all $t \in (0, \underline{t})$. The same continuity argument as in the proof of Lemma 1 shows that ρ must be continuous at 0, hence $\rho(0) \geq r^S(0)$. But then, we must have $\rho(0) = r^S(0)$, for if not, action monotonicity implies that action monotonicity implies that $\alpha(m) = a^R(0)$ for all $m \in M_{r^S(0)}$, and type 0 can profitably deviate to some $m \in M_{r^S(0)}$. Both parts of the desired conclusion now follows from Lemma A.1. $\diamond$ *Q.E.D.*

Lemma B.6. *Assume $r^S(0) < 1$. In any mD1 equilibrium with cutoff $\underline{t} = 0$, $\rho(0) \in \{r^S(0), 1\}$.*

This says that so long as not all types have the same cost-minimizing messages, the lowest type must either minimize its costs or maximally inflate its message.

Proof. Assume $\underline{t} = 0$ and $\rho(0) \neq 1$. By Lemma B.5, type 0 is separating, hence $\alpha(\mu(0)) = a^R(0)$, while $\rho(t) = 1$ for all $t > 0$. If $\rho(0) > r^S(0)$, then by action monotonicity, $\alpha(m) = a^R(0)$ for any $m \in M_{r^S(0)}$, and type 0 can profitably deviate to such a message. Therefore, $\rho(0) \leq r^S(0)$. Suppose $\rho(0) < r^S(0)$. Let $\hat{a} := \lim_{\varepsilon \downarrow 0} \alpha(\mu(\varepsilon))$. We must have $U^S(\hat{a}, 0) - kC(\nu(1), 0) = U^S(a^R(0), 0) - kC(\rho(0), 0)$, because otherwise, a small enough type $\varepsilon > 0$ would have a profitable deviation to $\mu(0)$. Pick an arbitrary $m \in M_{r^S(0)}$. By action monotonicity, $\alpha(m) \in [a^R(0), \hat{a}]$, and hence $U^S(\alpha(m), 0) \geq \min\{U^S(a^R(0), 0), U^S(\hat{a}, 0)\}$. Since $r^S(0) \in (\rho(0), 1)$, we have $C(\nu(r^S(0)), 0) < \min\{C(\nu(\rho(0)), 0), C(\nu(1), 0)\}$, and it follows that $U^S(\alpha(m), 0) - kC(\nu(r^S(0)), 0)) > U^S(\hat{a}, 0) - kC(\nu(1), 0)$. Therefore, type 0 has a profitable deviation to m, a contradiction. *Q.E.D.*

Lemma B.7. *In any mD1 equilibrium with cutoff $\underline{t}$, letting $a_1 := \lim_{t \downarrow \underline{t}} \alpha(\mu(t))$,*

(i) *if $\underline{t} > 0$, $U^S(a^R(\underline{t}), \underline{t}) - kC(\nu(\rho^*(\underline{t})), \underline{t}) = U^S(a_1, \underline{t}) - kC(\nu(1), \underline{t})$;*
(ii) *if $\underline{t} = 0$, $U^S(a^R(0), 0) - kC(\nu(r^S(0)), 0) \leq U^S(a_1, 0) - kC(\nu(1), 0)$.*

Proof. (i) Assume $\underline{t} \in (0, 1)$. First suppose $\rho^*(\underline{t}) = 1$. Then $\mu(\underline{t}) \in M_1$, and we must have $U^S(a^R(\underline{t}), \underline{t}) = U^S(a_1, \underline{t})$. Since all types below $\underline{t}$ are separating, equilibrium requires that for all

$\varepsilon > 0$,

$$U^S(a^R(\underline{t}-\varepsilon),\underline{t}) - kC(\nu(\rho^*(\underline{t}-\varepsilon)),\underline{t}) \leq U^S(a_1,\underline{t}) - kC(\nu(1),\underline{t}).$$

The desired conclusion follows from the above inequality and continuity of all the relevant functions.

So suppose $\rho^*(\underline{t}) < 1$. I claim that $\rho(\underline{t}) \in \{\rho^*(\underline{t}), 1\}$. If this is not the case, $\rho(\underline{t}) \in (\rho^*(\underline{t}), 1)$ by message monotonicity, hence $\underline{t}$ is separating. But then, action monotonicity implies that $\alpha(m) = a^R(t)$ for all $m \in M_{\rho^*(t)}$, and since $\rho^*(\underline{t}) > r^S(\underline{t})$, $\underline{t}$ can profitably deviate to any $m \in M_{\rho^*(t)}$.

Consider first that $\rho(\underline{t}) = \hat{t}$, in which case $\underline{t}$ is separating. Define for $\varepsilon \geq 0$,

$$W(\varepsilon) := U^S\left(a^R(\underline{t}), \underline{t}+\varepsilon\right) - kC\left(\nu(\rho^*(\underline{t})), \underline{t}+\varepsilon\right) - \left[U^S\left(a_1, \underline{t}+\varepsilon\right) - kC\left(\nu(1), \underline{t}+\varepsilon\right)\right].$$

If the Lemma does not hold, $W(0) > 0$ (the reverse inequality is inconsistent with optimality of $\mu(\underline{t})$). But then, by continuity of W, for small enough $\varepsilon > 0$, a type $\underline{t}+\varepsilon$ would prefer to deviate to playing $\mu(\underline{t})$, contradicting equilibrium.

It remains to consider $\rho(\underline{t}) = 1$. Lemma B.3 implies that for all $m \in M_{\rho^*(\underline{t})}$, $\alpha(m) = a^R(\underline{t})$. Thus, if the Lemma does not hold, optimality of $\mu(\underline{t})$ implies that $U^S(a_1, \underline{t}) - kC(\nu(1), \underline{t}) > U^S\left(a^R(\underline{t}), \underline{t}\right) - kC(\nu(\rho^*(\underline{t})), \underline{t})$. But then, by continuity of the relevant functions, we have that that for small enough $\varepsilon > 0$,

$$U^S(a_1, \underline{t}-\varepsilon) - kC(\nu(1), \underline{t}-\varepsilon) > U^S\left(a^R(\underline{t}-\varepsilon), \underline{t}-\varepsilon\right) - kC\left(\nu(\rho^*(\underline{t}-\varepsilon)), \underline{t}-\varepsilon\right),$$

implying that a type $\underline{t}-\varepsilon$ has a profitable deviation, contradicting equilibrium.

(ii) Assume $\underline{t} = 0$. The result is obvious if $r^S(0) = 1$, so assume $r^S(0) < 1$. Suppose first $\rho(0) = 1$. Then Lemma B.4 implies that for any $m \in M_{r^S(0)}$, $\alpha(m) = a^R(0)$. The desired result now follows from from optimality of $\mu(0)$. So consider $\rho(0) < 1$. By Lemma B.6, type 0 is separating with $\rho(0) = r^S(0)$. If the Lemma is not true, then optimality of $\mu(0)$ implies $U^S(a^R(0), 0) - kC(\nu(r^S(0), 0) > U^S(a_1, 0) - kC(\nu(1), 0)$. But then, by continuity of the relevant functions, for small enough $\varepsilon > 0$, a type ε has a profitable deviation to $\mu(0)$, contradicting equilibrium. *Q.E.D.*

Proof of Theorem B.1 on page 38. (1) Fix an mD1 equilibrium (μ, α). Lemmas B.1, B.5, and B.7 imply that there is an associated cutoff type $\underline{t}$ and strictly increasing sequence $\langle t_0 = \underline{t}, t_1, \ldots, t_J = 1\rangle$ that satisfy (4), (5), and (6). Moreover, by Lemma B.5, all types $t < \underline{t}$ are playing $\mu(t) \in M_{\rho^*(t)}$, while all types $t > \underline{t}$ are playing $\mu(t) \in M_1$. Hence, the proof of Theorem 2 shows that there is an LSHP equilibrium, $(\tilde{\mu}, \tilde{\alpha})$, such that for all $t \neq \underline{t}$, $\tilde{\mu}(t) = \mu(t)$ and $\tilde{\alpha}(\tilde{\mu}(t)) = \alpha(\mu(t))$.

(2) Pick an LSHP equilibrium, (μ, α), that is constructed in the proof of Theorem 2. I will argue that this is an mD1 equilibrium. Message and action monotonicity are obviously satisfied, so it only remains to check that every out-of-equilibrium response $\alpha(m)$ passes the third part of the mD1 criterion. Let M^μ denote the range of μ, i.e. the set of messages used in equilibrium.

Step 1: First consider any $m \notin M^\mu$ such that there is an $\tilde{m} \in M^\mu$ with $\Psi(\tilde{m}) = \Psi(m)$. Then we must have $\overline{A}(m, t) = \emptyset$ for all t, for otherwise some type can profitably deviate to some message in $M_{\Psi(m)}$. Hence, the mD1 criterion permits any response to m, in particular $\alpha(m)$.

Step 2: Suppose $\underline{t} > 0$. Consider any m such that $\Psi(m) \in [\rho^*(\underline{t}), 1)$. We have $\xi_l(\Psi(m)) = a^R(\underline{t}) < \xi_h(\Psi(m)) = a^R(\underline{t}, t_1)$, and $\alpha(m) = a^R(\underline{t})$. By construction, (5) holds. So

the argument used in the proof of Lemma B.3 (in particular, the Claim in Case 1) can be applied to show that $A(m,t) \subset \overline{A}(m,\underline{t})$ for any $t \neq \underline{t}$. Moreover, it can be verified that there exists some $a \in [\xi_l(m), \xi_h(m)]$ such that

$$U^S(a,\underline{t}) - C(\nu(\Psi(m)),\underline{t}) = U^S(\alpha(\mu(\underline{t})) - kC(\nu(\rho(\underline{t})),\underline{t}),$$

and because of $U^S_{12} > 0 > C_{12}$,

$$U^S(a,t) - C(\nu(\Psi(m)),t) < U^S(\alpha(\mu(t)) - kC(\nu(\rho(t)),t)$$

for all $t \neq \underline{t}$. Hence, $A(m,\underline{t}) \not\subseteq \overline{A}(m,t)$ for any $t \neq \underline{t}$. Therefore, $\alpha(m) = a^R(\underline{t})$ passes the mD1 test.

STEP 3: Suppose $\underline{t} > 0$. Consider any m such that $\Psi(m) < r^S(0)$. We have $\xi_l(\Psi(m)) = \xi_h(\Psi(m)) = \alpha(m) = a^R(0)$. It follows that $A(m,t) = \emptyset$ for all t, hence $\alpha(m)$ passes the mD1 test.

STEP 4: Suppose $\underline{t} = 0$, hence $M^\mu \subseteq M_1$. Pick any m such that $\Psi(m) < 1$. We have $\xi_l(\Psi(m)) = \alpha(m) = a^R(0) < \xi_h(\Psi(m)) = a^R(\underline{t}, t_1)$. The argument of Lemma B.4 can be applied to show that $A(m,t) \subseteq \overline{A}(m,0)$ for all t. There are now two cases to consider. First, suppose $A(m,0) = \emptyset$. Then it follows that $A(m,t) = \emptyset$ for all t, hence any response to m passes the mD1 test. On the other hand, suppose $A(m,0) \neq \emptyset$. Then because (6) holds by construction, continuity implies that exists some $a \in [\xi_l(\Psi(m)), \xi_h(\Psi(m))]$ such that

$$U^S(a,0) - C(\nu(\Psi(m)),0) = U^S(\alpha(\mu(0)) - kC(\nu(\rho(0)),0),$$

and because of $U^S_{12} > 0 > C_{12}$,

$$U^S(a,t) - C(\nu(\Psi(m)),t) < U^S(\alpha(\mu(t)) - kC(\nu(\rho(t)),t)$$

for all $t > 0$. Hence, $A(m,0) \not\subseteq \overline{A}(m,t)$ for any $t > 0$. Therefore, $\alpha(m) = a^R(\underline{t})$ passes the mD1 test. *Q.E.D.*

REFERENCES

ALLINGHAM, M. AND A. SANDMO (1972): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1, 323–338.

AUSTEN-SMITH, D. AND J. S. BANKS (2000): "Cheap Talk and Burned Money," *Journal of Economic Theory*, 91, 1–16.

BANKS, J. S. AND J. SOBEL (1987): "Equilibrium Selection in Signaling Games," *Econometrica*, 55, 647–661.

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

CAI, H. AND J. T.-Y. WANG (2006): "Overcommunication in Strategic Information Transmission Games," *Games and Economic Behavior*, 56, 7–36.

CHEN, Y. (2007): "Perturbed Communication Games with Honest Senders and Naive Receivers," Mimeo, Arizona State University.

CHEN, Y., N. KARTIK, AND J. SOBEL (2008): "Selecting Cheap-Talk Equilibria," *Econometrica*, 76, 117–136.

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

CHO, I.-K. AND J. SOBEL (1990): "Strategic Stability and Uniqueness in Signaling Games," *Journal of Economic Theory*, 50, 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, 1431–1451.

DENECKERE, R. AND S. SEVERINOV (2007): "Optimal Screening with Costly Misrepresentation," Mimeo.

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

ESÖ, P. AND A. GALAMBOS (2008): "Disagreement and Evidence Production in Pure Communication Games," Mimeo, Northwestern University.

ESÖ, P. AND J. SCHUMMER (2004): "Bribing and Signaling in Second-Price Auctions," *Games and Economic Behavior*, 47, 299–324.

GNEEZY, U. (2005): "Deception: The Role of Consequences," *American Economic Review*, 95, 384–394.

GREEN, J. R. AND J.-J. LAFFONT (1986): "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 53, 447–56.

GREEN, J. R. AND N. L. STOKEY (2007): "A Two-person Game of Information Transmission," *Journal of Economic Theory*, 127, 90–104.

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

HURKENS, S. AND N. KARTIK (2008): "Would I Lie to You? On Social Preferences and Lying Aversion," Forthcoming in *Experimental Economics*.

IVANOV, M. (2007): "Informational Control and Organizational Design," Mimeo, Pennsylvania State University.

KARTIK, N. (2005): "On Cheap Talk and Burned Money," Mimeo, University of California, San Diego.

——— (2007): "A Note on Cheap Talk and Burned Money," *Journal of Economic Theory*, 136, 749–758.

KARTIK, N., M. OTTAVIANI, AND F. SQUINTANI (2007): "Credulity, Lies, and Costly Talk," *Journal of Economic Theory*, 134, 93–116.

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

LACKER, J. M. AND J. A. WEINBERG (1989): "Optimal Contracts under Costly State Falsification," *Journal of Political Economy*, 97, 1345–1363.

LIN, H. AND M. F. MCNICHOLS (1998): "Underwriting Relationships, Analysts' Earnings Forecasts and Investment Recommendations," *Journal of Accounting and Economics*, 25, 101–127.

MAGGI, G. AND A. RODRIGUEZ-CLARE (1995): "Costly Distortion of Information in Agency Problems," *RAND Journal of Economics*, 26, 675–689.

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

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

MICHAELY, R. AND K. L. WOMACK (1999): "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations," *Review of Financial Studies*, 12, 653–86.

MILGROM, P. AND J. ROBERTS (1986a): "Pricing and Advertising Signals of Product Quality," *Journal of Political Economy*, 94, 796–821.

——— (1986b): "Relying on the Information of Interested Parties," *RAND Journal of Economics*, 17, 18–32.

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

Myerson, R. and M. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 28, 265–281.

Ottaviani, M. and F. Squintani (2006): "Naive Audience and Communication Bias," *International Journal of Game Theory*, 35, 129–150.

Sánchez-Pagés, S. and M. Vorsatz (2006): "Enjoy the Silence: An Experiment on Truth-Telling," Mimeo, University Maastricht.

Seidmann, D. J. and E. Winter (1997): "Strategic Information Transmission with Verifiable Messages," *Econometrica*, 65, 163–170.

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

Stein, J. C. (1989): "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior," *Quarterly Journal of Economics*, 104, 655–669.

Valley, K., L. Thompson, R. Gibbons, and M. H. Bazerman (2002): "How Communication Improves Efficiency in Bargaining Games," *Games and Economic Behavior*, 38, 127–155.

Verrecchia, R. E. (1983): "Discretionary disclosure," *Journal of Accounting and Economics*, 5, 179–194.