

# Collegio Carlo Alberto

---

## Communication and Learning

Luca Anderlini  
Dino Gerardi  
Roger Lagunoff

Working Paper No. 82  
December 2008 (Revised, January 2010)  
[www.carloalberto.org](http://www.carloalberto.org)

# Communication and Learning\*

LUCA ANDERLINI  
*(Georgetown University)*

DINO GERARDI  
*(Collegio Carlo Alberto  
and Yale University)*

ROGER LAGUNOFF  
*(Georgetown University)*

January 2010

**Abstract.** We study strategic information transmission in an organization consisting of an infinite sequence of individual decision makers. Each decision maker chooses an action and receives an informative but imperfect signal of the once-and-for-all realization of an unobserved state. The state affects all individuals' preferences over present and future decisions. Decision makers do not directly observe the realized signals or actions of their predecessors. Instead, they must rely on cheap-talk messages in order to accumulate information about the state. Each decision maker is therefore both a receiver of information with respect to his decision, and a sender with respect to all future decisions.

We show that if preferences are not perfectly aligned “full learning” equilibria — ones in which the individuals' posterior beliefs eventually place full weight on the true state — do not exist. This is so both in the case of private communication, in which each individual only hears the message of his immediate predecessor, and in the case of public communication, in which a decision maker hears the message of all his predecessors. Surprisingly, in the latter case full learning may be impossible even in the limit as all members of the organization become perfectly patient. We also consider the case where all individuals have access to a mediator who can work across time periods arbitrarily far apart. In this case full learning equilibria exist.

JEL CLASSIFICATION: C70, C72, C73, D80, D83.

KEYWORDS: Communication, Learning, Dynamic Strategic Information Transmission.

ADDRESS FOR CORRESPONDENCE: Luca Anderlini, Georgetown University, 37<sup>th</sup> and O Streets NW, Washington DC 20057, USA. [1a2@georgetown.edu](mailto:1a2@georgetown.edu)

---

\*We are grateful to Bruno Biais, three anonymous referees, Axel Anderson, Leonardo Felli, Johannes Hörner, Navin Kartik, Qingmin Liu, George Mailath, Stephen Morris, Marco Ottaviani, Larry Samuelson, Leeat Yariv and numerous seminar audiences for helpful comments. Any remaining errors are our own responsibility. All three authors are grateful to the National Science Foundation for financial support (Grant SES-0617789).

## 1. Introduction

### 1.1. Motivation

In large, complex, ongoing organizations the flow of information is critical. Information is needed to take multiple decisions, across physical distance and/or time. Information does not reside with any single individual (or unit) within the organization. It generally evolves through time and is fragmented across a large number of participants.

In many instances, practical bounds on the capacity of communication channels preclude that all members directly report to a center within the organization. This necessarily creates a structure in which members act *both* as senders (of what they have gathered directly or have been given by others), and receivers of information.

Our interest in this paper is to study the strategic side of information transmission in a large group of agents with imperfectly aligned interests. Information improves through time and accrues to different agents, each of whom has some information that is useful to all others.

We concentrate on organizations consisting of a sequence of imperfectly informed decision makers, who receive information from their predecessors, take an action, and then choose what information to pass on to the organizational sequence.

The model we analyze fits well an ongoing organization (say a firm, a lobbying concern, or a non-profit) that is run by a sequence of leaders (say CEOs, lead-lobbyists, or managers).

Each of the successive persons in charge takes a hidden action and observes some piece of information about an underlying fact that affects everyone in the sequence, and communicates the current state of organizational knowledge to decision makers down the line. The (state-dependent) preferences of successive leaders are aligned in the “long run,” but each decision maker has a bias relative to his own decision.

The underlying state determines which of several activities/tasks (types of competitive strategy for a CEO, degrees of aggressiveness for a lead-lobbyist, types of fund-raising for a non-profit manager) is more rewarding. Each decision maker can then be seen as having a choice of how much time/effort to expend across these multiple activities/tasks. It is then natural to suppose that, even if the effects are completely internalized, it is the current decision maker who disproportionately (or entirely) bears the different costs of the activities/tasks. In turn, this biases his behavior away from the long run common interest of the organization.

We investigate whether the organization will learn as appropriate from the information of

its members, and our answer is largely negative. Even though the available information would easily lead to “full learning” of the underlying truth if it were aggregated, this will generally not happen in equilibrium because of the agents’ strategic behavior when they choose what information to pass on down the line.

It should be noted at the very outset that, while our interests place us in a close relationship with the classical contribution of Crawford and Sobel (1982) (henceforth CS), the world we analyze is emphatically not a “repeated sender-receiver model.” Each of our players is both a sender and a receiver in the CS sense, and this changes the framework dramatically. We return to CS and other related papers in Section 2 below.

### *1.2. Modeling Choices*

We postulate a single given state of nature. Each player in the sequence receives an imperfect informative signal of its realization. In the absence of strategic behavior at the information transmission stage, full learning is trivial because we take the signals to be (conditionally) i.i.d. across players. While other specifications are of course possible, this way of proceeding is compelling in its simplicity and transparency.

In our set up each player’s action is payoff relevant to others, but it is not directly observed by anyone else in the model. This hidden-action nature of the set up fits well the class of situations we have mentioned before (sequence of CEOs, lead-lobbyists, managers). Moreover, our focus is on strategic information transmission via cheap-talk messages as in CS, not on signaling via payoff relevant actions. A hidden-action framework keeps signaling (and hence possible herding-like phenomena) out of the picture and, again, ensures transparency. While a treatment of the general case, in which signaling and possible herding phenomena co-exist with the ones we study here is desirable, it is clearly beyond the scope of this paper.

Our players are forward looking; their payoffs are affected by the actions of future players, not by the past. Aside from the possibility that the effect of past actions on payoffs may reveal information and hence activate the signaling effects we have just referred to (this is clearly not necessarily the case, depending on the timing of payoff accrual and action choice in each period) this is in fact without loss of generality. If the past affects payoffs in a non-information revealing way, then it is in fact “sunk” in the familiar sense of the word and hence does not change the equilibrium set.

For tractability, we use a specific functional form for preferences. The state determines which (scalar) action is best, and preferences take the form of a quadratic loss around the (unknown) best action given the state. In the concluding section of the paper we discuss

possible generalizations.

As in CS, with preferences that are perfectly aligned the behavior of the model is not very interesting at all. The misalignment in our case takes an extremely simple form. Other things being equal, the current decision maker prefers an action that is “lower” than his predecessors would like.

Both our state and the players’ signals are binary. Besides obviously being the simplest possible specification of uncertainty, this way of proceeding highlights a key difference between the set-up of this paper and that of CS. In CS, the state space is continuous (and the sender knows the actual state, rather than a signal of it as is the case here). Because of the misalignment in preferences, the sender will not fully reveal his information to the receiver. She will instead partition the state space into cells with a non-empty interior and only reveal the cell to which the state belongs. Because our state is binary, it is straightforward to check that in our model, if any one player knew the state with certainty then she would want to reveal it to the subsequent players.

Finally, there are many possible cases to consider in terms of “communication technology;” which players “hear” directly the message of a given player. We examine two main cases, plus a third one. The first two are the obvious two polar cases to use as first steps.

In the first case, communication is *public*. Each player’s message is directly observed by all subsequent players in the sequence. Messages, once “published,” are difficult or impossible to destroy or manipulate.

In the second case, communication is *private*. Each player’s message is directly observed only by the player who immediately follows in the sequence. Messages are easy to hide, manipulate, or falsify. Consequently, each player can act as a “gate-keeper,” choosing how much of his knowledge is passed down the line.

Lastly, we consider the case of *mediated* communication. All players only send messages to a central mediator, who then recommends actions to each on the basis of all messages up that point in the sequence. This case seems contrived because the mediator can work across players who are arbitrarily far apart in the sequence. It also requires the mediator to have unlimited communication capacity, thus going at least in part against the grain of our motivation above. We consider this case in some detail because it is known to be an insightful benchmark. Among other things, it sheds light on the reasons for some of the differences in the behavior of the model under the first two types of communication technology.

### 1.3. Preview of Results

A natural benchmark to consider in our model is that of truth-telling strategy profiles. All players reveal truthfully their signal to subsequent players (and pass on truthfully the information they receive from previous players in the case of private communication).

In our model truth-telling would guarantee (ex-ante) a Pareto-efficient outcome. It is also obvious that, given the i.i.d. signal structure that we posit, in this case the organization would “fully learn” the state. As we move down the sequence, the players’ belief would be sure to eventually place probability one on the true state.

We show that, both under private and under public communication, not only truth-telling is ruled out, but in fact there are *no equilibria* in which *full learning* occurs. Indeed, in the public communication model we show that posterior beliefs about the state must eventually reach a threshold at which point learning ceases altogether. Only babbling can occur from that point on. Our no-full learning results hold for any discount factor, and any magnitude of the misalignment in preferences. Regardless of whether communication is public or private, all equilibrium outcomes are (ex-ante) bounded away from the Pareto-efficient one induced by the truth-telling strategy profile.

Our no-full learning results naturally beg the question of whether full learning can obtain in the limit as players become more and more patient. We refer to possible sequences of equilibria in which full learning occurs as discounting approaches zero as “limit learning.”

As it turns out, the possibility of limit learning depends on which mode of communication (public or private) is considered. We show that limit learning is possible in the private communication case when the preference misalignment is not too large relative to the precision of the signal. By contrast, we show that limit learning is not possible in the public communication model if the signal precision is not too large. The pattern that seems to emerge from these results — parameter configurations in which limit learning is possible with private communication and impossible with public communication — is that “more” communication does not necessarily facilitate learning.

In the case of mediated communication, the possibility of full learning is restored. Full learning equilibria do exist in the mediated communication model when the bias is not large relative to the signal’s precision, regardless of how heavily the individuals discount the future. The fact that full learning can obtain in this case, sheds light on the seemingly odd comparison of the private and public communication cases above. In fact, private communication allows for limited mediation in the information flow. Communication between any player and any

subsequent one, can effectively be mediated by the intervening players. As it turns out, the possibility of “filtered” information being passed down the line facilitates learning.

Going back to our leading examples above, the role of information filtering has implications for the design of available information channels in a firm, a lobbying concern or a non-profit organization. To facilitate learning, the information channels of the organization should allow the CEO, lead-lobbyist or manager to filter the information that is passed down the line, and should not compel him to announce publicly his “report” about the information he has. Rather, in the absence of an unlikely technology that mimics the mediator described above, each agent in charge should be allowed to send a private report directly to his successor, who, in turn, can choose to filter what is sent on.

#### 1.4. *Outline*

The rest of the paper is organized as follows. Section 2 which follows reviews some related literature. In Section 3 we specify the model and equilibrium concept, and we define formally what it means that full learning occurs. Section 4 gives a very basic intuition for our no-full learning results and outlines some of the difficulties that arise in proving them. Sections 5 and 6 contain our main results, asserting that full learning is impossible both under private and under public communication. Section 7 is concerned with the limit learning results we mentioned above. Section 8 considers the case of mediated communication and Section 9 concludes.

For ease of exposition, all proofs have been placed in an Appendix. In the numbering of equations and all other items, a prefix of “A” points to the Appendix. For the sake of brevity, some proofs have been omitted from the Appendix as well. These can be found in a supplement to the paper available at <http://www.anderlini.net/learning-omitted-proofs.pdf>

## 2. **Related Literature**

This paper sits between the literature on social learning and the one on strategic cheap talk. Both these literatures examine why full learning might or might not occur. The social learning literature tends to focus on costly information acquisition and/or herding effects as the primary sources of incomplete learning, while the cheap talk literature is more concerned with agents’ incentives for withholding or manipulating information through the strategic use of messages. Each of these literatures is too vast to attempt a reasonable summary. Instead, we refer the interested reader to Gale (1996), Sobel (2000), and Moscarini, Ottaviani, and Smith (1998) for surveys and references on social learning, and to Farrell and Rabin (1996),

Sobel (2007), Krishna and Morgan (2001, 2008), and Ottaviani and Sørensen (2006) for surveys and references on strategic communication. Here, we limit ourselves to highlighting some similarities and differences with a few contributions that we feel are more closely related to our work.

The present paper has elements in common with both literatures we just mentioned; our costless messages are as in the cheap talk literature, and the possible updating from accumulated informative signals is akin to the social learning literature. The social learning side of our model, however, is deliberately kept trivial in order to concentrate on the strategic role of cheap talk. Crucially, past actions are not observed in the present model.

More importantly, most of the social learning literature differs from the present paper by assuming the linkages between agents are purely informational. The standard assumption is that the decisions of one individual do not directly affect the payoffs of another (Banerjee, 1992, among many others). A notable exception is a recent paper by Smith and Sørensen (2008) who examine herding with altruistically linked agents in a setup somewhat similar to ours. Each agent cares about how his action affects the beliefs of future decision makers, thereby adding a signaling component to the model. The result is a tendency for each decision maker to skew one's action against one's belief ("contrarianism"), and this incentive becomes stronger the more certain are the agents about the true state.

The present model shares some features with Smith and Sørensen (2008), notably the presence of dynastic or altruistic linkages between agents, a common value of the unobserved state, and, significantly, the potential to learn from others. The key differences are that, here, a payoff bias is assumed between different generations of decision makers, and the focus is on learning from costless messages rather than costly actions. Smith and Sørensen (2008) study an equivalent team problem (all individuals maximize the same average discounted sum) in which the externalities that generate herding are internalized. Further, they show that under some conditions this is equivalent to a planner's experimentation problem with incomplete information.

Payoff bias and cheap talk go hand in hand in the present set-up. While these two features are not necessarily inseparable — recent work by Kartik (2009) incorporates lying costs into the CS model thereby bridging the gap between costly signaling and cheap talk — our motivation is in isolating incentive effects of intergenerational learning via pure cheap talk.<sup>1</sup>

---

<sup>1</sup>Kartik's model shows, among other things, that lying costs turn cheap talk into costly signaling. He shows that when the sender's bias makes him prefer higher actions relative to the receiver, then full separation is

Within the cheap talk literature, our model is closer to those with a dynamic or dynastic element. Spector (2000), for instance, examines a repeated game of communication in a large population with a multi-dimensional policy space. Everyone in the population shares the same underlying policy preferences. The population is partitioned into two groups characterized by different priors about an unobserved state. Each individual receives an informative signal of the state. The two groups alternate in the role of senders and receivers with a new sender and receiver randomly drawn from the appropriate group each period.

Unlike ours, Spector’s players are myopic. Hence he does not look at the dynamic component of strategic information transmission. Rather, his focus is on the effect of conflict on a multi-dimensional policy space which, surprisingly, in a steady-state reduces to at most a single-dimensional disagreement.

A dynamic version of the basic sender-receiver game is found in Taub (1997). Both sender and receiver have ideal policy positions that vary across time according to serially correlated processes that are uncorrelated with one another. There are obvious similarities between our paper and his — both involve learning via accumulated messages. In Taub’s model, however, the learning (though not full learning) is possible due to the serial correlation in the sender’s bias. In our case whatever learning occurs takes place automatically, provided information is properly transmitted through time.

Li (2007) considers sequential arrival of information in a sender-receiver set up. Her focus is on what the sequence of reports by the sender reveals about his *ability* (the quality of his signals). Intriguingly, she finds that a sender who “changes his mind” may be a signal of high ability. In our set up all signals have the same precision since they are i.i.d. and hence there is nothing to infer about the sender’s ability in the sense of Li (2007).

Golosov, Skreta, Tsyvinski, and Wilson (2008) also consider a dynamic sender receiver game. Their set up is close to the original model in CS. A state is realized once and for all and observed by the sender. Then, in each of a finite number of periods, the sender sends a message to the same receiver who takes an action each time. They find that there are equilibria that do not have the simple partitional structure found in CS. Learning is incomplete, but more information is transmitted than in the static set up.

Carrillo and Mariotti (2000) examine voluntary ignorance in a single agent problem with present-biased preferences. Given the bias, the problem may be re-interpreted as a dynastic game similar to ours. The agent can costlessly access an informative signal with precision

---

possible in the lower regions of the type space. This is not possible in the standard CS model with a continuum of types.

of his choosing. The signal is public and is observed by all future agents. The authors consider Markov equilibria, showing that one with full learning always exists. They also exhibit Markov equilibria with incomplete learning in which experimentation stops at some finite date with positive probability.

Strictly speaking, the costless information acquisition in Carrillo and Mariotti (2000) is not cheap-talk. However, there is a sense in which their model is similar to the case of public communication that we analyze here because of the public availability of information in their case, and of messages in ours. The key difference between the two models is that in Carrillo and Mariotti (2000) any agent can unilaterally decide to become arbitrarily well-informed. This is not possible in our model since each agent receives one signal of a given precision, which remains private even when messages are public. Moreover, in their model perfect recall precludes an agent from withholding his information from the view of future agents. Hence, each agent has an incentive to become informed if all others do. As a consequence, in Carrillo and Mariotti (2000) there always is a full-learning equilibrium, whereas full-learning is ruled out in our case.

Finally, this paper builds on our previous work on dynastic communication in Anderlini, Gerardi, and Lagunoff (2008, 2010), Anderlini and Lagunoff (2005), and Lagunoff (2006).<sup>2</sup> These papers consider dynastic games in which messages are about the past history of play. Because there are no “objective” types (i.e., no exogenous states or payoff types), there is no learning in the traditional sense of the word. These contributions focus on whether social memory, embodied in the beliefs of the new entrants, is accurate, and what pitfalls may befall a society if it is not.

### 3. The Framework

#### 3.1. Model

Time is discrete, and indexed by  $t = 0, 1, 2, \dots$ . In period 0 Nature selects the state  $\omega$  which can be either 1 or 0. State  $\omega = 1$  is chosen with probability  $r \in (0, 1)$ . Nothing else takes place in period 0, and no-one in the model observes Nature’s choice of  $\omega$ , which is determined once and for all.

There is a period-by-period flow of imperfect information about the state  $\omega$ . We take this to have the simplest form that will allow for an interesting model: a sequence of imperfect but informative conditionally i.i.d. signals. The realized signal in period  $t$  is denoted  $s_t$ , and

---

<sup>2</sup>Related models are also found in Kobayashi (2007) and Lagunoff and Matsui (2004).

its symmetric structure is parameterized by a single number  $p \in (1/2, 1)$  as

$$\Pr \left\{ s_t = 1 \mid \omega = 1 \right\} = \Pr \left\{ s_t = 0 \mid \omega = 0 \right\} = p \quad (1)$$

With obvious terminology, we refer to  $p$  as the signal's quality.<sup>3</sup>

There is a countable infinity of decision makers, each indexed by  $t = 1, 2, \dots$ . Each decision maker's tenure lasts for one period; the decision maker in period  $t$  chooses an action  $a_t \in \mathbb{R}$  at time  $t$ , but he also derives state-dependent utility not only from his own choice but also from all the choices  $a_{t+\tau}$ , with  $\tau = 1, 2, \dots$ , of all his successors. We refer to this decision maker simply as “player  $t$ .”

We work with a specific payoff function that captures our desiderata and at the same time ensures tractability.<sup>4</sup> Fix a sequence of actions  $a_1, \dots, a_t, \dots$  and a state  $\omega$ . The payoff of player  $t$  is written as

$$-(1 - \delta) \left[ \sum_{\tau=0}^{\infty} \delta^{\tau} (\omega - a_{t+\tau})^2 + 2\alpha a_t \right] \quad (2)$$

with  $\delta$  a common discount factor in  $(0, 1)$ . The squared term in (2) is a quadratic loss function that reaches a maximum at  $a_t = \omega$  for  $\omega = 0, 1$ . The term  $2\alpha a_t$  in (2) embodies the bias in the decision makers' preferences. Each player  $t$  bears an extra “effort” cost  $2\alpha a_t$ , increasing in  $a_t$ .<sup>5</sup> Aside from this, for given beliefs, the decision makers' preferences are aligned. Of course, what matters to player  $t$  is the expected value of (2), given his beliefs about  $\omega$  and the sequence of all future actions  $\{a_{t+\tau}\}_{\tau=1}^{\infty}$ .

Because of the way the bias term is parameterized by  $\alpha$ , it is immediate to check that if player  $t$  believes that  $\omega = 1$  with probability  $x_t$ , then his choice of  $a_t$  to maximize (2) is simply

$$a_t = x_t - \alpha \quad (3)$$

---

<sup>3</sup>While the assumption of conditionally i.i.d. signals is important for our arguments, the symmetry of the signal structure could easily be dispensed with.

<sup>4</sup>As we anticipated, Subsection 9.1 is devoted to a discussion of possible generalizations of payoffs.

<sup>5</sup>We can also interpret each  $a_t$  (appropriately normalizing so that  $a_t \in [0, 1]$ ) as the fraction of time that  $t$  spends on one of two tasks available, with each task yielding different losses in the two states. The bias term is then interpreted as embodying the fact that the task on which the fraction of time  $a_t$  is spent is more costly for  $t$  than the other. See Subsection 1.1 above.

Given (3), we refer to  $\alpha$  simply as *the* bias. Throughout, we assume that  $\alpha \in (0, 1/2)$ .<sup>6</sup>

Player  $t$ 's information comes from his predecessor(s) in the form of payoff irrelevant messages. He does not observe directly any part of the past history of play or realized signals.

In the case of *private communication* player  $t$  observes a message  $m_{t-1}$  sent by player  $t-1$ , which is observed by no-one else. In the case of *public communication* player  $t$  observes all messages  $m_{t-\ell}$  (with  $\ell = 1, \dots, t-1$ ) sent by all preceding players. It is useful to establish a piece of notation for what player  $t$  observes by way of messages that encompasses both the private and public communication cases. We let  $m^{t-1} = m_{t-1}$  in the private communication case, while  $m^{t-1} = (m_1, \dots, m_{t-1})$  in the public communication case. (Set  $m_0 = m^0 = \emptyset$ .)

The choice of message spaces is irrelevant: our main results hold for *any* choice of message spaces. This will be reflected in the formal statement of each of our propositions below. For the time being we simply establish that the message space available to each player  $t$  from which to pick  $m_t$  is denoted by  $M_t \neq \emptyset$ . (Set  $M_0 = \emptyset$ .) We also let  $M$  represent the entire collection  $\{M_t\}_{t=0}^{\infty}$  of message spaces.

After observing  $m^{t-1}$ , player  $t$  picks an action  $a_t$ . After choosing  $a_t$  he observes the informative but imperfect signal  $s_t$ , and after observing  $s_t$  he decides which message  $m_t$  to send.<sup>7</sup> Note that, crucially, player  $t$  does not learn the state before sending  $m_t$ . This can be interpreted as his period  $t$  payoff simply accruing after he sends  $m_t$ , or as his experiencing a payoff that depends on his action  $a_t$  and the signal  $s_t$  in an appropriate way.<sup>8</sup>

Given this time line, player  $t$ 's strategy has two components. One that returns an action  $a_t$  as a function of  $m^{t-1}$ , and one that returns a message  $m_t$  as a function of  $m^{t-1}$  and  $s_t$ . We denote the first by  $\lambda_t$  and the second by  $\sigma_t$ . Formally

$$\lambda_t : m^{t-1} \mapsto a_t \quad \text{and} \quad \sigma_t : (m^{t-1}, s_t) \mapsto m_t \quad (4)$$

Notice that there is a sense in which the  $\lambda_t$  part of  $t$ 's strategy is trivial. Given his beginning-of-period belief  $x_t \in [0, 1]$  that  $\omega = 1$ , using (2) and (3) we know that in any equilibrium it will be the case that  $a_t = x_t - \alpha$ . But since the only mechanism for transmitting information are the messages described above, it is clear that  $x_t$  is entirely determined by the message-sending behavior of  $t$ 's predecessors  $(\sigma_1, \dots, \sigma_{t-1})$  together with the observed  $m^{t-1}$ . Thus,

<sup>6</sup>The restriction  $\alpha < 1/2$  simplifies some of our computations, and only disposes of uninteresting cases.

<sup>7</sup>The fact that  $s_t$  is observed after the choice of  $a_t$  simplifies the analysis but is inessential to our results.

<sup>8</sup>The latter would be the analogue of what is routinely done in the large literature on repeated games with imperfect private monitoring. See Mailath and Samuelson (2006) for an overview and many references.

all strategic behavior (and hence any equilibrium) is entirely pinned down by a profile  $\sigma = (\sigma_1, \dots, \sigma_t, \dots)$ .

To avoid unnecessary use of notation, and since this does not cause ambiguity, from now on we will refer to  $\sigma = (\sigma_1, \dots, \sigma_t, \dots)$  as a strategy profile, leaving it as understood that the associated  $\lambda = (\lambda_1, \dots, \lambda_t, \dots)$ , determining actions via (2) and (3), is to be considered as well. Of course, actions and messages sent in equilibrium will be determined by (2), (3) and the players' on- and off-path beliefs. Luckily, off-path beliefs need not be specified with any piece of notation. The reason is simply that off-path beliefs will not play any role in our results. Since the on-path beliefs are entirely pinned down by  $\sigma$  via Bayes' rule, referring to this profile will always suffice.

Throughout the paper, an ‘‘Equilibrium’’ will refer to a Weak Perfect-Bayesian Equilibrium (henceforth WPBE) of the game at hand (see for instance Fudenberg and Tirole, 1991). This is the weakest equilibrium concept that embodies sequential rationality that applies to our set up. Our main focus is on ruling out certain outcomes as possible in any equilibrium. Therefore, a weaker equilibrium concept strengthens our main results. It is in fact also the case that whenever we construct an equilibrium in any of our arguments below (Propositions 3 and 5 below), then the equilibrium will also satisfy the more stringent requirements of a Sequential Equilibrium in the sense of Kreps and Wilson (1982).

We note that only pure strategies are considered. We strongly conjecture that our main results hold when mixed strategies are allowed. Note that, since (2) is strictly concave in  $a_t$ , mixing in actions is clearly impossible in any equilibrium. Moreover, recall that our main results concern the absence of equilibria in which the state is learned through time. Mixed message strategies seem very unlikely to help in this respect, even if one could construct strategy profiles that guarantee the indifferences needed in any mixed strategy equilibrium. We return to this point at some length in Subsection 9.2 in the conclusions.

### 3.2. Full Learning

Our main question (mostly answered in the negative) is whether the accumulated information in this organization will suffice to eventually uncover, or fully learn, the true state. Hence, we now need to be precise about what we mean by full learning in our model.

Fix a strategy profile  $\sigma$ . As we noted above, the belief of player  $t$  at the beginning of period  $t$  that  $\omega = 1$  is entirely determined by  $\sigma$  and the  $m^{t-1}$  he observes. Denote it by  $x_t(\sigma, m^{t-1})$  with the dependence on  $\sigma$  or  $m^{t-1}$ , or both, sometimes suppressed as convenient.

Now fix a realization  $\mathbf{s}$  of the entire stochastic process  $\{s_t\}_{t=1}^\infty$ . Via the strategy profile

$\sigma$ , the realization  $\mathbf{s}$  clearly determines  $m^{t-1}$  for every  $t = 1, 2, \dots$ . This makes it clear that a given profile  $\sigma$  (together with a realization of the state  $\omega$  and the signal structure (1) parameterized by  $p$ ) defines a stochastic process  $\{x_t\}_{t=1}^{\infty}$  governing the (on-path) evolution of the players' posterior. The probabilities (conditional when needed) of events in this process will be indicated by  $\Pr_{\sigma}$ .<sup>9</sup>

We are now able to define formally what it means that full learning obtains given  $\sigma$ .

**Definition 1.** *Full Learning:* We say that a profile  $\sigma$  induces full learning of the state if and only if the associated stochastic process  $\{x_t\}_{t=1}^{\infty}$  converges in probability to the true state  $\omega$ .

In symbols, we say that full learning of the state obtains given  $\sigma$  if and only if for every  $\varepsilon > 0$  there exists a  $\bar{t} \geq 1$  such that  $t > \bar{t}$  implies

$$\Pr_{\sigma} \left\{ x_t > 1 - \varepsilon \mid \omega = 1 \right\} > 1 - \varepsilon \quad \text{and} \quad \Pr_{\sigma} \left\{ x_t < \varepsilon \mid \omega = 0 \right\} > 1 - \varepsilon \quad (5)$$

#### 4. Some Bare-Bones Intuition

Our purpose here is to set out some basic intuition behind our no-full learning results (Propositions 1 and 2 below), highlight some of the difficulties that the arguments run into, and outline how these are resolved in very broad terms. More detailed sketches of the actual arguments follow each proposition below.

##### 4.1. Looking One Period Ahead

To see that the bias in preferences may make the truthful revelation of information problematic just from one period to the next one can reason in much the same way as in the CS world. Consider just a single individual (sender) and his successor (receiver). Let  $x$  be the initial belief (that the state is equal to one) of the sender. Suppose first that the sender's strategy is to reveal his signal  $s \in \{0, 1\}$ . Since the signal is informative, the receiver will now have beliefs either  $x^0$  (if the report is  $s = 0$ ) or  $x^1$  (if the report is  $s = 1$ ) that satisfy  $x^0 < x < x^1$ . Because of the bias, the receiver will take action  $x^0 - \alpha$  in one case and  $x^1 - \alpha$  in the other. Suppose on the other hand that the sender babbles and reveals no information. For instance, he reports  $s = 1$  regardless of what he really observes. Then the receiver's belief

---

<sup>9</sup>It is clear that  $\sigma$  also induces a stochastic process  $\{m_t\}_{t=0}^{\infty}$  governing the (on-path) evolution of messages sent by each player. With a slight abuse of notation the probabilities (conditional when needed) of events in this process will be indicated in the same way by  $\Pr_{\sigma}(\cdot)$ . The argument of the operator will ensure that this does not cause any ambiguity. Finally, because  $\sigma$  pins down the on-path evolution of both  $\{x_t\}_{t=1}^{\infty}$  and  $\{m_t\}_{t=0}^{\infty}$ , it also induces *updated* probabilities that  $\omega$  is one or zero, conditional on any (positive probability) event concerning  $\{x_t\}_{t=1}^{\infty}$  or  $\{m_t\}_{t=0}^{\infty}$ , or both. These will also be identified using the notation  $\Pr_{\sigma}(\cdot)$ , with the arguments used to avoid any ambiguity.

is the same as the sender’s before he observes  $s$ , namely  $x$ . Hence the receiver takes action  $x - \alpha$ .

For the information to be truthfully transmitted between sender and receiver we need to satisfy the usual two incentive compatibility constraints. The sender must *both* (weakly) prefer action  $x^1 - \alpha$  to action  $x^0 - \alpha$  when his belief is  $x^1$  *and* he must (weakly) prefer action  $x^0 - \alpha$  to action  $x^1 - \alpha$  when his belief is  $x^0$ . The first constraint does not bind since actions of the receiver are always skewed downward.

The second constraint on the other hand might be impossible to satisfy. This constraint is satisfied if  $x^0 - \alpha$  is no further from  $x^0$  than is  $x^1 - \alpha$ , as is for instance the case in Figure 1. Since the constraint can be written as  $2\alpha < x^1 - x^0$ , it is clear that if  $x^1 - x^0$  is small relative to  $\alpha$  then the second incentive compatibility constraint cannot be satisfied. If this is the case, after observing  $s = 0$ , the sender prefers the action that the receiver would take if he believed the sender’s message and was told that  $s = 1$ . Hence there is no equilibrium in which the sender’s messages fully reveal the signal  $s$ .

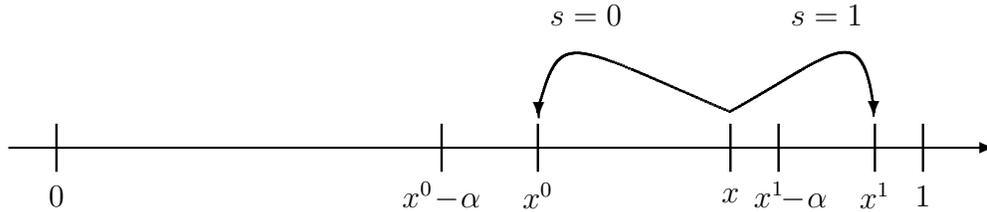


Figure 1: One Period Incentive Compatibility

It is critical to notice that the possibility of truthful revelation just one period ahead depends on the the *relative* size of  $x^1 - x^0$  and  $\alpha$ . Given this, it is immediate that one period ahead truth-telling will be possible when the belief  $x$  takes “intermediate” values (provided that  $\alpha$  is not too large), while it will become impossible as  $x$  approaches either zero or one. This is simply because Bayes’ rule tells us that the updating effect will be larger for intermediate values of  $x$  and will shrink all the way to zero as  $x$  goes to the extremes of  $[0, 1]$ .

#### 4.2. Difficulties Of The Full-Blown Dynamic Case

Armed with the observations of Subsection 4.1, now consider the full-blown dynamic model we described in Section 3. To fix ideas, suppose communication is public (similar issues arise when communication is private). Suppose first that we had an equilibrium with perpetual truth-telling. This is a particularly simple way to achieve full learning of course. Suppose for the sake of argument that the true state is one. Then, in equilibrium, the beginning-of-period

belief of player  $t$ ,  $x_t$ , converges to one (with probability one). This of course means that, the difference between  $x_t$  updated with a signal of one (denoted  $x_t^1$ ) and  $x_t$  updated with a signal of zero (denoted  $x_t^0$ ) must eventually become arbitrarily small. So eventually in this putative truthful equilibrium we must be in a situation in which  $2\alpha > x_t^1 - x_t^0$  and hence looking ahead only one period truthful revelation is no longer possible. This, however is not sufficient to dismiss perpetual truth-telling as a possible equilibrium. The reason is that the decision makers look indefinitely ahead at discounted payoffs.

A player  $t$  who finds that  $2\alpha > x_t^1 - x_t^0$  before deciding to deviate or not from the putative truth-telling equilibrium, will have to reason through the following. Suppose that he observes a signal  $s_t$  equal to zero. Since  $2\alpha > x_t^1 - x_t^0$ , as far as his period  $t + 1$  payoff, player  $t$  clearly prefers to deviate and report  $s_t = 1$ . However, think now of his payoff in period  $t + \tau$ , with  $\tau$  possibly large. Taking the truthful strategies of the others as given, the beliefs between  $t + 1$  and  $t + \tau$  will go up and down through time, following the realizations of the signals each period. Deviating in period  $t$  and reporting  $s_t = 1$  when in fact  $s_t = 0$  will simply take the entire sequence of beliefs two “notches” up relative to where it would be if  $t$  did not deviate (in the same sense that  $x_t^0$  is two notches down from  $x_t^1$  — it takes two realizations of the signal equal to one for Bayes’ rule to take beliefs from  $x_t^0$  to  $x_t^1$ ). But then consider a sequence of signals that takes the beliefs gradually towards the center of the interval  $[0, 1]$ . As we noted before, (provided that  $\alpha$  is not too large) the one-period-ahead calculation in this case tells the player not to deviate from truthful revelation. So, with some probability, deviation at  $t$  has the same effect (in period  $t + \tau$ ) than deviating when the beliefs are in an intermediate position and hence there is no gain from deviation at  $t + \tau$ . In fact deviations could entail a sizable loss when the beliefs are in the intermediate range of  $[0, 1]$ , especially if the signal is reasonably precise.

Of course, the outcome of calculations for the entire path following a deviation will depend crucially on the signal precision (how fast the beliefs move through time), and on how much the individuals discount the future. As it turns out, the computation is not trivial. One of the main reasons for this is that while the loss from deviating when beliefs are intermediate is multiplied by a probability that approaches zero, the gain from deviating shrinks to zero as well when  $x_t$  goes to either one or zero, because in these cases the difference  $x_t^1 - x_t^0$  must shrink to zero too.

In order to rule out *any* equilibrium with full learning, a further layer of difficulties arises. These stem from the fact that full learning can in principle obtain in equilibria that are very different from the simple truth-telling equilibrium hypothesized above. In particular,

full learning could obtain in an equilibrium that fails to be Markov in the sense that its continuation equilibria could be different after different histories leading to the same belief. This structure of equilibrium can in principle be used to “punish” deviations from truth-telling behavior. The evaluation of continuation payoffs we have sketched above becomes more subtle as a result.

To see the extra difficulties involved in ruling out full learning in general, consider again a player  $t$  who finds that  $2\alpha > x_t^1 - x_t^0$ , and observes a signal  $s_t = 0$ . Just as before, in a full learning equilibrium this must happen at some point with probability one. Looking one period ahead, player  $t$  prefers to report  $s_t = 1$  as before. However, we can no longer be sure that the following individuals use truth-telling strategies. The continuation path of play could be very different after a deviation at  $t$ . As before, it is possible that a sequence of signals eventually takes the beliefs back towards the middle of  $[0, 1]$ , where the deviation at  $t$  entails a loss. Because the equilibrium need not be Markov in the first place, the continuation equilibrium could for instance be such that the beliefs get “trapped” in this range for long periods (because of prolonged spells of babbling) *only* following a report of  $s_t = 1$  at  $t$ . So, by deviating player  $t$  may be “punished” in that the future path of beliefs may yield the losses associated with beliefs in the intermediate range of  $[0, 1]$  with much higher probability than in the case of a truth-telling strategy profile. Our results establishing that full learning cannot take place overcome the difficulties we have described mostly by bounding continuation payoffs or their changes in case of deviations in a variety of ways.

### 4.3. Broad Outline Of No Full Learning Proofs

The proofs of Propositions 1 and 2 below share some structure that arises naturally from our intuitive discussion so far. In both cases, the payoff to player  $t$  other than the current-period component is divided into the payoff that accrues to  $t$  in period  $t + 1$ , and the rest of  $t$ 's continuation payoff, stemming from the actions taken in  $t + 2$  and beyond.<sup>10</sup> For the purposes of this discussion we will refer to the former as the payoff one period ahead, and to the latter as the continuation payoff.

In both the private and public communication cases, the argument begins in the following way. If we had any equilibrium with full learning, then for some large  $t$  the belief of player  $t$  must be eventually arbitrarily close to 0 (if  $\omega = 0$ ) or to 1 (if  $\omega = 1$ ), *and* player  $t$  must reveal

---

<sup>10</sup>Dealing with the part of  $t$ 's payoff that accrues at  $t$  — determined by the  $\lambda_t(\cdot)$  part of his strategy as in (4) — is trivial as we noticed in Subsection 3.1. It will simply be given by (3) and not affect any of the play that follows.

his information truthfully in order for learning to continue. Using Bayes' rule this means that the beliefs of player  $t + 1$  if player  $t$  "tells the truth" or "lies" must be arbitrarily close to each other. However, for a given  $\alpha$ , using the observations we made in Subsection 4.1 above, player  $t$  wants to deviate from truth-telling looking at his payoff one period ahead. In other words, if we had an equilibrium with full learning, it must be the case that, with probability one, some player will want to deviate looking at his payoff one period ahead.

As we noticed above in Subsection 4.2, the fact that a deviation is profitable looking one period ahead is far from implying that a deviation pays for  $t$  overall because the continuation payoff needs to be taken into account. In this respect the two arguments for Propositions 1 and 2 differ. In the case of Proposition 1 dealing with private communication, we are able to bound the possible loss in terms of continuation payoff that player  $t$  may incur from deviation from truth-telling at  $t$ . In the case of Proposition 2, dealing with public communication, we are able to show that the gain from deviation one period ahead still dominates even if we assume that the continuation payoff is the worse possible one (equilibrium or not) after deviation, and it is the best possible (equilibrium or not) after truth-telling.

In the case of private communication, the pivot of the argument is the maximizing behavior of player  $t + 1$ . Communication is *private*, therefore player  $t + 1$  can "wipe the past" entirely if he so decides. Since player  $t + 1$ 's beliefs following a deviation and truth-telling are arbitrarily close to each other we can then use explicitly his incentives to show that the possible loss from deviation in the continuation payoff of player  $t$  shrinks to zero faster than the gain one period ahead. This is sufficient to destroy the putative full learning equilibrium.

In the case of public communication, the incentives of player  $t + 1$  do not provide such a tight handle on the possible loss from deviation as in the private communication case. Intuitively, the argument we have just sketched out no longer works because player  $t + 1$  cannot lie about past messages (they are observed directly by his successors). Hence he cannot "undo"  $t$ 's messages as above. In fact, in the case of public communication, as we discussed in Subsection 4.2  $t$ 's message may have large unavoidable effects on play at  $t + 2$  and beyond, regardless of the behavior of player  $t + 1$ . The argument for Proposition 2 below proceeds by comparing the overall payoff to  $t$  in the following two cases. In the first, player  $t$  does not deviate and hence reports his signal truthfully,  $t + 1$  acts according to his belief (arising from  $t$ 's truthful behavior), and the continuation payoff is the *best* possible, regardless of what equilibrium strategies prescribe. In the second case, player  $t$  deviates and mis-reports his signal, as his gain looking one period ahead tells him to do, just as we outlined in Subsection 4.1 above, and the continuation payoff is the *worst* possible, regardless

of what equilibrium strategies prescribe. Using the features of the payoffs we have postulated in (2) we can then show that  $t$ 's overall payoff in the second case is larger than in the first. This obviously implies that  $t$  wants to deviate from truthful revelation, and hence destroys the putative full learning equilibrium. It is worth noting here that identifying the appropriate best and worst continuation payoffs are relatively more tractable problems than, for instance, finding extremal values of equilibrium continuation payoffs. Intuitively, this is the consequence of the fact that solutions to problems of maximization or minimization of payoffs are guaranteed to have a Markov structure that, as we have noted already, equilibria are not guaranteed to have.

### 5. No Full Learning with Private Communication

With private communication full learning of the state cannot occur in equilibrium. This is so regardless of the message spaces, discount factor, signal precision, bias or prior.

**Proposition 1.** *No Full Learning with Private Communication: Fix any  $M, \delta, p, \alpha$  and  $r$ . Then there is no equilibrium  $\sigma^*$  of the model with private communication that induces full learning of the state.*

By the Proposition, posterior beliefs do not converge in probability to the true state. Concretely, this means that for a set of signal realizations arising with positive probability, beliefs are not close to the truth at infinitely many dates.

What follows is an informal sketch of the argument we use in the Appendix to prove Proposition 1. It is convenient to divide it into three parts.

The first part of the argument shows that off-path behavior can be safely ignored in the second and third parts. To see that this is the case consider an equilibrium  $\sigma^*$ , and begin by noticing that, regardless of the cardinality of the message spaces, since we are only considering pure strategies, only a finite set of messages can be sent with positive probability in any period for any realization of  $\omega \in \{0, 1\}$ .<sup>11</sup> We can then proceed to delete from the players' strategy sets all messages that are sent with probability zero for both realizations of  $\omega \in \{0, 1\}$ . The strategy profile  $\hat{\sigma}^*$  that we obtain in this way clearly induces the same stochastic process  $\{x_t\}_{t=1}^\infty$  on the players' posterior given any realization of the state  $\omega$ . Moreover, the profile  $\hat{\sigma}^*$  must also be an equilibrium in the game with the reduced message spaces. This is simply because by deleting messages from the players' strategy sets we only reduce the possible

---

<sup>11</sup>This is because the possible realizations of the string of binary signals up to and including any time  $t$  is finite.

deviations to be considered to verify that a given strategy profile is in fact an equilibrium. Hence, there is no loss of generality in considering only message spaces and equilibria like  $\hat{\sigma}^*$ , in which all messages are sent with positive probability.

The second part of the argument concerns continuation payoffs as a function of a player's beliefs, when the beliefs are sufficiently close to one. In particular, consider a player's continuation payoff computed after he has chosen an action, but before he observes the signal, as a function of the message he receives, conditional on the true state  $\omega$ . We show that if  $m$  and  $m'$  are two messages that are associated with beliefs  $x$  and  $x'$  that are both arbitrarily close to one, then the continuation payoffs that they induce are arbitrarily close to each other, both conditional on the true state being one and on the true state being zero. This is not hard to show when we are conditioning on  $\omega = 1$ . Using the incentive constraints of the player, it must be that the continuation payoffs are close to each other since the conditioning event ( $\omega = 1$ ) has probability close to one in the player's beliefs. If the payoffs were not close to each other one of the two "types" (either the one receiving  $m$  or the one receiving  $m'$ ) would necessarily have an incentive to behave like the other. To show that the continuation payoffs are close to each other when conditioning on  $\omega = 0$  involves an extra difficulty since the conditioning event has probability close to zero in the player's beliefs and hence in this case the conditional continuation payoffs have a vanishingly small effect on the player's incentives. We circumvent this difficulty by ruling out some possible equilibrium continuation paths, which in turn can be done using the incentives of subsequent players.

The third part of the argument proceeds by contradiction. Suppose then that we have an equilibrium with full learning. Conditional on  $\omega = 1$ , the players' beliefs must converge to one; learning cannot stop. So some player  $t - 1$  must have a beginning-of-period belief arbitrarily close to one.<sup>12</sup> For the purposes of this illustration we focus on the case in which this player fully transmits his end-of-period belief to player  $t$ . In other words, after observing  $s_{t-1}$  equal to zero or to one he passes on to  $t$  his end-of-period belief,  $x'$  or  $x$ , using messages  $m'$  or  $m$  respectively.<sup>13</sup> Since  $t - 1$ 's beginning-of-period belief is arbitrarily close to one,  $x$  and  $x'$ , and  $m$  and  $m'$  are as in the second part of the argument above.

---

<sup>12</sup>While in Section 4 above we reasoned intuitively from the point of view of "player  $t$ ," here we refer to the putative deviator as player  $t - 1$ . Using player  $t - 1$  is in line with what we do in the actual proof of Proposition 1 in the Appendix, and it simplifies notation considerably.

<sup>13</sup>Notice that clearly it must be that  $x > x'$ . Notice also that since we are in the private communication case, the picture may be different from the case we are concentrating on here. In particular,  $t - 1$  may send different messages after observing  $s_{t-1}$  equal to zero and to one, but may at the same time fail to transmit his end-of-period belief to player  $t$ . This is because other "types" of player  $t - 1$  may be using the same messages on the equilibrium path. Of course, in the Appendix we treat an exhaustive set of possible cases.

Now consider the continuation payoff of player  $t - 1$ , at the end of the period, as a function of the message he sends. This can be divided into the payoff that accrues to him at  $t$ , and the stream of payoffs that accrue to him from  $t + 1$  onward — the continuation payoffs of player  $t$  about which we know from the second part of the argument. Assume that player  $t - 1$  has observed  $s_{t-1} = 0$  and hence that he is supposed to send message  $m'$  to player  $t$ , thus inducing him to have belief  $x'$ .

As far as period  $t$  is concerned, player  $t - 1$  now has an incentive to deviate and send message  $m$ . This is for exactly the same reason that we pointed out in Subsection 4.1 above, with the help of Figure 1. It follows directly from the fact that  $x - x'$  is small. Notice that the gain from deviation that accrues to player  $t - 1$  at time  $t$  shrinks to zero as  $x$  and  $x'$  approach one and hence  $x - x'$  approaches zero. Nevertheless, this gain is shown to be bounded below by  $(x - x')\alpha$ .

Next, consider the change in continuation payoffs from  $t + 1$  onwards that is generated by a deviation from  $t - 1$ 's part to send  $m$  instead of  $m'$ . This term is hard to sign. The difficulty is similar to the one we outlined in Subsection 4.2 above. A single deviation at  $t - 1$  could make a large difference a long way down the line. However, we do know that this payoff difference has an upper bound of the form  $(x - x')A$ , with  $A$  a term that depends on  $t$ 's continuation payoffs. This is because  $t - 1$ 's continuation payoffs from  $t + 1$  onwards are averages of player  $t$ 's continuation payoffs from the second part of the argument. Using the fact that  $t$ 's continuation payoffs are all close together as in the second part of the argument, using his incentives we can then show that as  $x - x'$  shrinks to zero, the term  $A$  in the bound also shrinks to zero. Hence, as  $x$  and  $x'$  approach one and hence  $x - x'$  approaches zero, the gain to  $t - 1$  from deviating that accrues to him at time  $t$  must be larger than all possible losses from the deviation in periods after  $t$ . So, player  $t - 1$  will want to deviate and send  $m$  after observing  $s_{t-1} = 0$ . This destroys the putative equilibrium in which full learning occurs.

Before moving on, we remark that a relatively minor modification of the argument we use to prove Proposition 1, yields a further result concerning the *set* of equilibria of the model with private communication. In particular, it can be ruled out that there exists a sequence of equilibria that yield outcomes that are closer and closer to full learning.<sup>14</sup>

---

<sup>14</sup>More formally, once we fix any  $\delta, p, \alpha$  and  $r$ , the set of  $\varepsilon$  for which any equilibrium satisfies (5) is bounded away from zero.

## 6. No Full Learning with Public Communication

Just as in the case of private communication, with public communication full learning of the state cannot occur in equilibrium. In this case too, this is so regardless of the message spaces, discount factor, signal precision, bias or prior.

**Proposition 2.** *No Full Learning with Public Communication: Fix any  $M$ ,  $\delta$ ,  $p$ ,  $\alpha$  and  $r$ . Then there is no equilibrium  $\sigma^*$  of the model with public communication that induces full learning of the state.*

The conclusion of Proposition 2 is the same as for Proposition 1: posterior beliefs do not converge in probability to the true state. In this case, this is true despite the fact that decision makers can access the entire history of messages that precedes them.

As with Proposition 1, a formal proof of the claim is in the Appendix. Here we give an informal sketch of the argument, which for convenience is divided into five parts.

The first part of the argument consists of the observation that in the public communication case, since we are only considering pure strategies, each player's behavior at the message stage can only be of two varieties. He can either babble and reveal nothing, or he can fully reveal his signal to all subsequent players. Hence, there is no loss of generality in setting the message space of all players equal to  $\{0, 1\}$ . There is also no loss of generality in considering (as we do both here and in the Appendix) only the case in which all players who babble send message one, and all players who reveal their signal send message one when the signal is one, and message zero when the signal is zero.

The second part of the argument considers the possible continuation payoffs of a player, say  $t$ , who truthfully reveals his signal to his successor  $t + 1$ . These are the possible continuation payoffs at the message stage of a player  $t$  whose end-of-period belief is the same as the beginning-of-period belief of  $t + 1$ . As the strategy profile of all players from  $t + 1$  on (the "continuation profile") varies, so will the continuation payoff of player  $t$ . It can be shown that as we consider all possible continuation profiles (regardless of whether they are part of an equilibrium or not), the continuation payoff of  $t$  achieves a maximum when the continuation profile is truthful revelation from all players from  $t + 1$  on, and achieves a minimum when the continuation profile is babbling for all players from  $t + 1$  on.

The third part of the argument concerns a different thought experiment concerning  $t$ 's continuation payoffs at the end of period  $t$ . In particular, we consider the hypothetical situation in which  $t$ 's end-of-period belief is zero (so that he is certain that  $\omega = 0$ ), but  $t + 1$

and all subsequent players think that  $t$ 's end-of-period belief is  $x_{t+1} > 0$ .<sup>15</sup> Somewhat abusing terminology we will refer to  $x_{t+1}$  as a “fresh prior” for  $t + 1$  and all subsequent players. When  $x_{t+1}$  is close to one we are able to show that, as we consider all possible continuation profiles (regardless of whether they are part of an equilibrium or not),  $t$ 's continuation payoff achieves a minimum when the continuation profile is babbling for all players from  $t + 1$  on.

In this case, unlike in the second part of the argument above, considering directly all possible continuation profiles in the comparison with babbling turns out not to be a practical line of argument. We proceed by simplifying the problem with an observation. It obviously suffices to show that the babbling continuation profile is a solution to the problem of *minimizing*  $t$ 's continuation payoff, given  $t$ 's belief of zero and an  $x_{t+1}$  close to one. However, once we consider this minimization problem, it is not hard to see that only continuation profiles that are Markov — so that behavior only depends on beliefs about the state and not on which history leads to those beliefs — need to be considered. This simplifies the problem considerably, and allows us to reach the stated conclusion. In fact, restricting attention to Markov strategies for the purpose of this characterization is a way to circumvent precisely the difficulty of “trapped” beliefs that we outlined in Subsection 4.2 above.

The fourth part of the argument simply puts together what we know from the second and third parts. Notice that the continuation payoff of player  $t$  for a given continuation strategy profile and a given fresh prior  $x_{t+1}$  is actually linear in  $t$ 's end-of-period belief. This, using what we know about how  $t$ 's continuation payoff changes as we change the continuation profile, has the following implication. Whenever the fresh prior  $x_{t+1}$  is near one and  $t$ 's end-of-period belief is anywhere below  $x_{t+1}$ , the continuation payoff to  $t$  is minimized when the continuation profile is babbling for all players from  $t + 1$  on.

The fifth part closes the argument by contradiction, assuming that there is a putative equilibrium in which full learning occurs. Clearly, this implies that we must be able to find a player  $t$  with a beginning-of-period belief arbitrarily close to one who reveals his signal to player  $t + 1$ . Let  $x_{t+1}^0$  and  $x_{t+1}^1$  be the beginning-of-period belief of  $t + 1$  after he receives message zero and one respectively. In this putative equilibrium these are also the end-of-period beliefs of player  $t$  after observing a signal of zero and one respectively. Moreover, both  $x_{t+1}^0$  and  $x_{t+1}^1$  are arbitrarily close to one and they obviously satisfy  $x_{t+1}^0 < x_{t+1}^1$ .

---

<sup>15</sup>This is obviously not something that can take place given any strategy profile when communication is public. Nevertheless, it is a useful case to consider as an “anchor” to look at  $t$ 's continuation payoffs after possible deviations from a putative equilibrium with full learning. This is exactly how this part of the argument will feed into the next one.

Now consider player  $t$ , after observing  $s_t = 0$ . His end-of-period belief is  $x_{t+1}^0$ . His putative equilibrium strategy tells him to send message zero. Upon receiving this message  $t + 1$ 's belief will be  $x_{t+1}^0$ . Hence, using what we know from the second part of the argument, an upper bound on  $t$ 's continuation payoff at this point can be obtained by computing his payoff when the continuation profile is truthful. Suppose instead that  $t$  deviates and sends message one. Since the putative equilibrium prescribes that he is truthful, his message is believed by  $t + 1$  and all subsequent players. After the deviation the end-of-period belief of  $t$  is  $x_{t+1}^0$ , but  $t + 1$  and all subsequent players believe it to be  $x_{t+1}^1$ . So,  $x_{t+1}^1$  is now just like the fresh prior above for  $t + 1$  and all subsequent players. Therefore from part four of the argument we know that in this case a lower bound on  $t$ 's continuation payoff can be obtained assuming that the continuation profile is babbling from  $t + 1$  on.

Since the two continuation profiles involved in the bounds are perpetual truth-telling in one case and perpetual babbling in the other, the comparison is relatively simple to carry out. This computation tells us that a player  $t$  with end-of-period belief  $x_{t+1}^0$  prefers the continuation payoff he gets under a babbling continuation profile with  $t + 1$  and all subsequent players having a fresh prior of  $x_{t+1}^1$  to the continuation payoff he gets under a truth-telling continuation profile with player  $t + 1$  having a belief of  $x_{t+1}^0$ . Hence, after observing  $s_t = 0$  player  $t$  wants to deviate from the putative equilibrium and send  $t + 1$  a message of one, instead of a message of zero as the equilibrium prescribes.

Finally, as we noted in Subsection 1.3 above, it should be pointed out that our argument yields a stronger result than the formal statement of Proposition 2. In particular, the bounds on continuation payoffs that hold for any continuation profile (equilibrium or not) and their comparison as in the fifth part of the argument above show that there are thresholds (upper and lower in fact) to the possible equilibrium beliefs across all equilibria. Therefore in any equilibrium there are points at which learning ceases altogether. Only babbling can occur after they are reached. Of course, this also implies that, as in the case of private communication, the set of equilibria is "bounded away" from full learning in this case too.<sup>16</sup>

## 7. Perfect Patience

### 7.1. Limit Learning

Intuitively, as the discount factor  $\delta$  increases the players' preferences become closer to complete alignment. The bias term only affects the current period and the weight of the current

---

<sup>16</sup>See footnote 14 above.

period in each player's preferences obviously decreases as  $\delta$  grows.

Since the main source of the difficulties in learning the state is the mis-alignment in the players' preferences, intuitively we would expect learning to become easier as the players' patience grows. We say that *limit learning* is possible if there are equilibria with full learning in the limit as  $\delta$  approaches 1.

We investigate whether limit learning is possible in our model. The answer depends on, among other things, the type of communication allowed. In the model with public communication we are able to show that when  $p$  is not too large limit learning is impossible. On the other hand in the model with private communication we can show that, if  $\alpha$  is not too large relative to  $p$ , then limit learning can occur.

Before proceeding, we need to make the notion of limit learning precise.

**Definition 2.** *Limit Learning:* We say that limit learning of the state is possible if and only if as  $\delta$  approaches 1 we can find a sequence of equilibria along which the true state is learned with a precision and a probability both approaching 1.

Fix  $p$ ,  $\alpha$  and  $r$ . In symbols, we say that limit learning is possible if and only if for every  $\varepsilon > 0$  we can find a  $\bar{t} \geq 1$  and a  $\bar{\delta} < 1$  such that for every  $\delta > \bar{\delta}$  there is a collection of message spaces  $M$  and an equilibrium  $\sigma^*(\delta)$  such that for every  $t > \bar{t}$

$$\Pr_{\sigma^*(\delta)} \left\{ x_t > 1 - \varepsilon \mid \omega = 1 \right\} > 1 - \varepsilon \quad \text{and} \quad \Pr_{\sigma^*(\delta)} \left\{ x_t < \varepsilon \mid \omega = 0 \right\} > 1 - \varepsilon$$

We are now ready for our first result on limit learning.

## 7.2. Private Communication

Provided the bias is not too large relative to the signal's quality, limit learning is possible when communication is private.

**Proposition 3.** *Limit Learning with Private Communication:* Fix any  $r$  and suppose that  $\alpha < p - 1/2$ . Then limit learning is possible in the model with private communication.<sup>17</sup>

At this point it is worth mentioning again that, as we alluded to in Subsection 3.1, the limit learning of the statement of Proposition 3 can be sustained with Sequential Equilibria rather than just with WPBE.

---

<sup>17</sup>In the Appendix we prove a slightly stronger statement. The one given here is phrased to allow direct comparability with the statement of Proposition 4 below. The proof of Proposition 3 in the Appendix yields a collection  $M$  of message spaces and an equilibrium  $\sigma^*(\delta)$  satisfying the requirements of Definition 2 which do not in fact depend on  $\delta$  but only on the value of  $\varepsilon$  (or equivalently  $\bar{\delta}$ ) in Definition 2.

In the Appendix we construct an appropriate sequence of Sequential Equilibria. As with our previous two results, we only give a sketch here. For the sake of clarity, this is divided into four parts. Throughout this sketch, for simplicity we imagine that the prior  $r$  is equal to one half. If this were not the case the review phase we sketch below would not be symmetric making the outline a lot more cumbersome to follow. The role of the condition  $\alpha < p - 1/2$  would also be lengthier to explain.

The argument begins with two observations. The condition  $\alpha < p - 1/2$  in the statement of the proposition has an immediate interpretation. This condition tells us that any player with a belief of  $1 - p$  or less will strictly prefer action  $-\alpha$  to action  $1 - \alpha$  to be taken in the future. Moreover, these preferences will be reversed for any player with a belief of  $p$  or more. Since  $r = 1/2$  it is clear that the belief of any player who knows (or believes) the history of signals to contain more zeroes than ones will have a belief of  $1 - p$  or less, while any player who knows (or believes) the history of signals to contain more ones than zeroes will have a belief of  $p$  or more. So, the majority of signals being zero or one will determine a player's strict preferences between future actions  $-\alpha$  and  $1 - \alpha$  as above.

The second part of the argument outlines the mechanics of the review phase that characterizes the equilibria we construct to support limit learning. In the first  $T$  periods of play, with  $T$  a large odd number, equilibrium consists of a review phase. The purpose of the review phase is to establish whether the majority of signals in the first  $T$  periods is zero or one. The review phase is said to have outcome zero or one according to which one of these events occur. Of course if the proportion of zeroes or of ones is very large in the initial part of the review phase, its outcome may be determined early.

During the first  $T$  periods, unless the outcome has already been determined, the players collectively count the number of signals equal to one that have been realized so far. Each player adds one to his predecessor's message if he observes a signal of one, and leaves the count unchanged if he observes a signal of zero. He then sends the new tally as a message to the next player. If at any point the outcome of the review phase is determined, the current player sends a message that indicates that the outcome of the review phase has been determined and whether it is zero or one. From that point on, up to  $T$  and beyond, only that message is sent forward through time each period and thus reaches all subsequent players. No further information is ever added.

The third part of the argument consists of noticing that, using the Weak Law of Large Numbers, making the review phase longer and longer, we can ensure that the beliefs of all players from  $T + 1$  onwards are arbitrarily close to the true state (within  $\varepsilon$  of it), with

arbitrarily large probability (at least  $1 - \varepsilon$ ) as required for limit learning to occur. Intuitively, this is clear from the fact that the signal is informative and, in equilibrium, all players from  $T + 1$  onwards only know whether the majority of a large number  $T$  of realizations of the signal is zero or one.

The fourth part of the argument consists of the verification that no player wants to deviate from the prescribed behavior. It is useful to consider two cases. Begin with any player  $t \geq T + 1$ , after the review phase is necessarily over. As we noted in the third part of the argument, this player's belief is either arbitrarily close to zero or arbitrarily close to one, depending on the message he receives about the outcome of the review phase. If he deviates and reports an outcome of the review phase different from the one he received, all subsequent actions will be arbitrarily close to  $-\alpha$  and his belief arbitrarily close to one in one case, and will be arbitrarily close to  $1 - \alpha$  with his belief arbitrarily close to zero in the other case. Therefore, because of what we know from the first part of the argument, no player  $t \geq T + 1$  stands to gain from deviating from the proposed equilibrium.

Next, consider any player  $t \leq T$  when the outcome of the review phase may or may not have been determined. Consider his continuation payoff accruing from time  $T + 1$  onward. Note that, whatever  $t$  tells his successor, the actions from  $T + 1$  onward will all be arbitrarily close to either  $-\alpha$  or  $1 - \alpha$ , according to whether the outcome of the review phase will be declared to be zero or one.

Now consider what the belief of player  $t$  would be at the beginning of period  $T + 1$ , if he could observe the signal realizations in all periods from  $t$  to  $T$ . Assume that all players, including  $t$ , obey the equilibrium strategies we have described. Using Bayes' rule, this belief would clearly be  $p$  or more for those realizations of signals that entail an outcome of the review phase equal to one, and would be  $1 - p$  or less for those realizations of signals that entail an outcome of the review phase equal to zero. If he plays according to his equilibrium strategy, from  $T + 1$  onwards player  $t$ 's payoff will be the one corresponding to an action close to  $1 - \alpha$  in the former case and the one corresponding to an action close to  $-\alpha$  in the latter case. It then follows from the first part of the argument that, if  $t$  plays according to his equilibrium strategy, the continuation actions from  $T + 1$  onwards will be matched to what he prefers, contingent on every possible realizations of signals.

If on the other hand  $t$  deviates from his equilibrium strategy, it is not hard to see that the deviation will make his beliefs and preferred continuation actions from  $T + 1$  onwards mismatched with positive probability. Since  $\delta$  is appropriately high, gains accruing before  $T + 1$  can be ignored. Hence, no player  $t \leq T$  wants to deviate from the proposed equilibrium.

Our next concern is limit learning in the public communication case.

### 7.3. Public Communication

Provided that the signal quality is not too high, limit learning cannot occur when communication is public. As is the case in Proposition 3, our next statement identifies a simple set of sufficient conditions.

**Proposition 4.** *No Limit Learning with Public Communication: Fix any  $\alpha$  and  $r$  and suppose that  $p < \sqrt{2}/(1 + \sqrt{2})$ . Then limit learning is impossible in the model with public communication.*

Proposition 4 is proved in the Appendix. A brief sketch of the argument follows.

Fix a discount factor  $\delta$  and an equilibrium. From the proof of Proposition 2 we know that there is a greatest lower bound to the values that the belief of any player can take in this equilibrium. Denote this bound by  $\underline{x}$ , and denote  $\underline{x}^0$  and  $\underline{x}^1$  the beliefs obtained by updating  $\underline{x}$  on the basis of a signal of zero and one respectively. In the terminology of Section 4, the beliefs  $\underline{x}^0$  and  $\underline{x}^1$  are one notch down and one notch up from  $\underline{x}$  respectively.

The argument proceeds by contradiction. Suppose that the proposition is false, then as  $\delta$  approaches one we must be able to find a sequence of equilibria with  $\underline{x}$  approaching zero, so that limit learning can occur. (Limit learning of course requires more, but this is clearly a necessary condition for it to be possible at all.)

Since  $\underline{x}$  is a greatest lower bound on beliefs, we must be able to find a player, say  $t$ , who has a beginning-of-period belief  $\underline{x}^1$  and reveals his signal truthfully to subsequent players. So, after observing  $s_t = 0$ , if  $t$  follows his equilibrium strategy, the beginning-of-period belief of all players from  $t + 1$  onwards is  $\underline{x}$  and the action  $\underline{x} - \alpha$  is chosen in every period.

Now focus on  $t$  entertaining a deviation to reporting a signal realization of one after observing  $s_t = 0$ . It is convenient to look at  $t$ 's continuation payoff component by component. Fix a given period ahead, say  $t + \tau$ , and a sequence of realized signals between  $t$  and  $t + \tau$ . Two cases are possible. Given the sequence of signals, either at  $t + \tau$  the lower bound has been achieved and hence action  $\underline{x} - \alpha$  is chosen, or the bound has not been achieved and an action  $\hat{x} - \alpha > \underline{x} - \alpha$  is chosen instead. In the former case, clearly player  $t$  is indifferent between following his equilibrium strategy and the hypothesized deviation.

Consider then the case in which action  $\hat{x} - \alpha > \underline{x} - \alpha$  is chosen at  $t + \tau$ . Notice that since player  $t + \tau$  does not know that  $t$  deviated, his beginning-of-period belief is two notches up from what it would have been if he had known about  $t$ 's deviation. Clearly, the latter is

how  $t$  evaluates this particular component of his continuation payoff after the hypothesized deviation. Moreover,  $t$ 's updated belief must be at least  $\underline{x}^0$  since the lower bound has not been reached and hence  $t + \tau$ 's beginning-of-period belief must be at least  $\underline{x}^1$ .

The condition that  $p < \sqrt{2}/(1 + \sqrt{2})$  of the statement of the proposition now allows us to close the argument. In fact when  $p$  is low in this sense, so that each notch of updating makes an appropriately small difference to beliefs, straightforward algebra is sufficient to show the following. Assume that  $\underline{x}$  is sufficiently low, and consider any player with a belief  $x$  at least as large as  $\underline{x}^0$ . Denote by  $x^{11}$  the belief two notches up from  $x$ . Then the player with belief  $x$  strictly prefers action  $x^{11} - \alpha$  to action  $\underline{x} - \alpha$ . However, this implies that for  $\underline{x}$  sufficiently low, player  $t$  finds the deviation hypothesized above strictly profitable. This destroys the putative equilibrium with  $\underline{x}$  arbitrarily low corresponding to  $\delta$  arbitrarily close to one, and hence closes the argument for Proposition 4.

It is instructive to piece together what we know from Propositions 3 and 4 about limit learning in the private and public communication cases. While the characterization is by no means complete because of the parametric restrictions imposed in Propositions 3 and 4, on a subset of the parameter space these two results allow for a comparison.

In particular, directly from the sufficient conditions stated in Propositions 3 and 4 we get that if  $\alpha < p - 1/2 < \sqrt{2}/(1 + \sqrt{2}) - 1/2$ , then limit learning is possible in the private communication case but impossible if communication is instead public.<sup>18</sup> This observation substantiates the remark we made in the introduction that the investigation of limit learning yields the insight that “more” (public) communication does not necessarily enable the organization to “learn more.”

Our next and last concern is the case of communication that takes place via a mediator. As we anticipated above, this helps our understanding of the comparative behavior of the public and private communication cases that we have just pointed out.

## 8. Mediated Communication

We now turn to the well-known theoretical benchmark case of mediated communication model (Forges, 1986, Myerson, 1982). In this case our main results are overturned: full learning can occur. Because mediated communication seems so unnatural in this context, in our view this strengthens our original no-full learning results. But, of course, this is a matter of interpretation.

---

<sup>18</sup> $\sqrt{2}/(1 + \sqrt{2}) - 1/2 \approx 0.086$ .

The structure of the equilibria that yield this possibility also seems of independent interest. They are a natural variation on those that yielded limit learning in Proposition 3 above, exploiting the ability to filter information to make appropriate use of the Weak Law of Large Numbers. In this sense, the mediated communication case sheds light on the comparison between Propositions 3 and 4 above. The ability to pass on filtered information down the line is (in principle) unlimited in the mediated communication case. In the case of private communication filtering is restricted by the fact that communication between two decision makers can only be mediated by the intervening players. In the case of public communication case, no filtering is possible at all.

In the mediated communication model, we take it to be the case that each player  $t$  reports to a mediator the signal  $s_t$  that he observes, so that we can simply choose  $M_t = \{0, 1\}$  for every  $t$ . The mediator then recommends an action to each player  $t$  on the basis of the reports of all players from 1 to  $t - 1$ .

This structure can be shown to be without loss of generality (Forges, 1986, Myerson, 1982), but this is besides the point in our case. Since we are establishing a *possibility* result, doing so for one particular class of mediated communication protocols is all that is required.

A mediated communication protocol  $\mathcal{M}$  is a sequence of maps  $\{\mu_t\}_{t=0}^{\infty}$ . Each  $\mu_t$  is from  $\{0, 1\}^{t-1}$  into  $\mathbb{R}$ . The interpretation is that  $\mu_{t-1}(m_1, \dots, m_{t-1})$  is the action that the mediator recommends to player  $t$  if the reports of the previous players are  $(m_1, \dots, m_{t-1})$ . Consistently with the notation we established in Subsection 3.1, we set  $m^{t-1} \equiv \mu_{t-1}(m_1, \dots, m_{t-1})$ .

We call  $\mathcal{M}^*$  a mediated communication equilibrium if it satisfies the standard conditions that, given  $\mathcal{M}^*$ , all players are willing to truthfully reveal the signal they observe, and all players are willing to take whatever action  $a_t = m^{t-1}$  prescribed by  $\mathcal{M}^*$ . Notice that any given  $\mathcal{M}^*$ , mapping sequences of signals into actions, induces a stochastic process  $\{a_t\}_{t=1}^{\infty}$ .

Recall that, using (2) and (3), player  $t$  is willing to take action  $a_t$  if and only if his belief is equal to  $x_t = a_t + \alpha$ . Therefore, in line with Definition 1, we say that a mediated communication equilibrium  $\mathcal{M}^*$  induces full learning of the state if and only if the induced stochastic process  $\{a_t + \alpha\}_{t=1}^{\infty}$  converges in probability to the true state.

Our next result asserts that full learning can obtain in the mediated communication case.

**Proposition 5.** *Full Learning With a Mediator: Fix  $\delta, p$  and  $r$ . Suppose that  $\alpha < p - 1/2$ . Then there is a mediated communication equilibrium  $\mathcal{M}^*$  that induces full learning of the state.*

The proof of Proposition 5 is omitted for reasons of space. Once one has the details of

the proof of Proposition 3, it is largely a notational and reinterpretational exercise. As with our previous results, we provide a sketch of the argument here.

The communication equilibrium  $\mathcal{M}^*$  that we propose induces full learning of the state using an infinite sequence of review phases like the one in the proof of Proposition 3, increasing in length through time. For ease of exposition we will refer to the mediator recommending an action to a player and to the belief induced by the mediator's message interchangeably. The former equals the latter minus  $\alpha$  so that no ambiguity will ensue.

Fix a small number  $\varepsilon$ . Number the phases as 1 through to  $n$  and beyond, and let  $T_n$  be the length of each phase. Recall that  $\delta$ ,  $p$  and  $r$  are given, and that, by assumption, the condition  $\alpha < p - 1/2$  of Proposition 3 is also satisfied here.

All players in each review phase, say  $n$ , receive a message from the mediator recommending an action based on the outcome of review phase  $n - 1$  (the empty message in the case of review phase 1), and report truthfully to the mediator the signal they observe.

Each review phase  $n$  is sufficiently long (but may exceed the minimum length required for this as will be clear below) to ensure that, as in the proof of Proposition 3, via the Weak Law of Large Numbers, the belief of players in review phase  $n + 1$  are within  $\varepsilon^n$  of the true state with probability at least  $1 - \varepsilon^n$ . This can be done by picking  $T_n$  large enough, just as in the proof of Proposition 3. For ease of reference below, call this the precision requirement for  $T_n$ . Notice that of course as  $n$  grows large and hence  $\varepsilon^n$  shrinks to zero this is sufficient to achieve full learning of the state as required.

We then need to check that no player wants to deviate from any of the prescriptions of  $\mathcal{M}^*$  in any of the review phases. In one respect, this is simpler than in the case of Proposition 3. This is so because no player can affect the actions of other players in the same review phase. The mediator does not reveal to them any information pertaining to the current review phase. So, the incentives to reveal truthfully his signal for a player in review phase  $n - 1$  only depend on the actions of players in review phases  $n$  and beyond. It then follows that by picking  $T_n$  sufficiently large we can ensure that the incentives for players in review phase  $n - 1$  are determined by their (strict) preferences over actions that will be taken during review phase  $n$ , regardless of the fixed value of the discount factor  $\delta$ . Call this the incentive requirement for  $T_n$ . We then pick each  $T_n$  to be large enough so that both the precision requirement and the incentive requirement are met, and, aside from one caveat, the argument is complete.

The caveat is that in our sketch of the proof of Proposition 3 we dealt with a single symmetric review phase that for simplicity was assumed to begin with a prior  $r$  equal to one half. Clearly, with multiple review phases, as  $n$  increases the starting point of each review

phase will be further and further away from the symmetry guaranteed by  $r = 1/2$ . This point can be dealt in the same way as in the formal proof of Proposition 3 presented in the Appendix. The outcome of each review phase will not be determined just by simple majority, and the length requirements of  $T_n$  mentioned above will have to take these asymmetries into account.

## 9. Conclusions and Extensions

### 9.1. More General Payoffs

Working with a specific functional form for payoffs as in (2) is clearly more restrictive than one would like, and hence begs the question of how far this can be relaxed. To fix ideas, imagine that we wrote a general form of (2) as

$$(1 - \delta) \left[ V(a_t, \omega) + \sum_{\tau=1}^{\infty} \delta^\tau U(a_{t+\tau}, \omega) \right] \quad (6)$$

which of course reduces to (2) when  $V(a_t, \omega) = -(\omega - a_t)^2 - 2\alpha a_t$  and  $U(a_{t+\tau}, \omega) = -(\omega - a_{t+\tau})^2$ . The question we are pursuing in this subsection is then what forms of  $V(\cdot, \cdot)$  and  $U(\cdot, \cdot)$  can we use and ensure that our results are still viable.

Although the assumption embodied in (2) can be relaxed in both cases, the degree to which we know this to be true depends on whether we are considering the case of private or public communication. Everything we say here about this issue depends on the structure of the proofs of our claims as they currently are. It follows that further generalizations cannot be ruled out, but depend on insights that would have to stem from proofs that are different from the ones we presented above.

In the case of private communication, our arguments for Propositions 1 and 3 generalize in a fairly direct way to a very broad class of possible payoffs. A set of sufficient conditions are as follows. Both  $U$  and  $V$  in (6) are continuous and concave over the appropriate range, and the maximizing values of  $a_t$  for both  $U$  and  $V$  exist finite for both  $\omega = 1$  and  $\omega = 0$ . The bias term works consistently in one direction so that, regardless of the state, the maximizing value of  $a_t$  for  $U$  must be larger than that for  $V$ . Finally, to dispose of some uninteresting cases we need the equivalent of assuming  $\alpha \leq 1/2$  when the payoffs are as in (2).<sup>19</sup> To this end it is enough to require that the value of  $U$  given  $\omega = 0$  is larger if we plug in the value

---

<sup>19</sup>See footnote 6 above.

of  $a_t$  that maximizes  $V$  given  $\omega = 0$  than if we use the  $a_t$  that maximizes  $V$  given  $\omega = 1$ .<sup>20</sup>

The conditions we know to be sufficient for our arguments in the case of public communication to be viable for more general payoffs are more stringent and less transparent than in the case of private communication. As before, we require both  $U$  and  $V$  in (6) to be continuous and concave over the appropriate range, and the maximizing values of  $a_t$  for both  $U$  and  $V$  to exist finite for both  $\omega = 1$  and  $\omega = 0$ . The bias term should be “separable” in the sense that for some  $f(\cdot)$  we can write  $U(a_t, \omega) = V(a_t, \omega) + f(a_t)$  for all  $a_t$  and  $\omega \in \{0, 1\}$ , with  $f(\cdot)$  (weakly) increasing and convex. Finally, if we let  $\mathbf{a}^V(x)$  be the maximizing value of  $a_t$  for  $xV(a_t, 1) + (1 - x)V(a_t, 0)$ , then  $\mathbf{a}^V(x)$  should be (weakly) convex.

The conditions for the public communication case that we have described are clearly met in the linear-quadratic case in (2) and in a “neighborhood” (appropriately defined) of it. It is not hard to see that some other explicitly defined functional forms will satisfy them too.

On balance, the sufficient conditions we have described so far in the two cases of private and public communication justify proceeding with the linear-quadratic payoffs as we do here. The reason is two-fold. First, the conditions for the public communication case that we have outlined are not “primitive” since they refer to the maximizer  $\mathbf{a}^V(x)$ ,<sup>21</sup> and not particularly transparent or easy to interpret. This should be compared with the elegance and ease of use of the quadratic-linear form in (2). Second, the linear-quadratic form in (2) easily allows some key parametric comparisons between the public and private communication cases like our remarks at the end of Subsection 7.3. Because of the different sufficient conditions we have outlined in the private and public communication cases, these comparisons would be much less transparent without the specific payoffs we postulated in (2).

### 9.2. Concluding Remarks and Further Extensions

We examine large, complex ongoing organizations in which information flows through players who are both senders and receivers of it. In particular, we analyze a model that fits an ongoing firm or other organization that is led by a sequence of imperfectly informed decision makers, with each decision maker getting information from his predecessors and transmitting it down the line to future decision makers. To take good decisions our players must, to borrow from Isaac Newton’s famous quote, be “standing on the shoulders of giants.”<sup>22</sup> We analyze

---

<sup>20</sup>The same conditions are sufficient to generalize the argument we use to prove Proposition 5.

<sup>21</sup>Of course, these could translated into conditions on  $V$ , which for a differentiable  $V$  would be statements involving its third derivative.

<sup>22</sup>In a 1676 letter to Robert Hooke. A popularized account of the context for this exceedingly well known quote and of the relationship between Newton and Hooke can be found in the “Introduction” and the “Sir

the limits to this process when preferences are not perfectly aligned and hence players act strategically in the transmission of information.

We examine both polar cases of private and public communication from one decision maker to his immediate successor in one case, and to all subsequent players in the other. The main results show that in either model, full learning is impossible.

Demonstrating the impossibility of full learning involves a number of difficulties in both cases. Even under a set of simplifying assumptions, in most cases a direct evaluation of continuation payoffs proves an intractable route to follow in our model. As a consequence our main results are proved using a variety of techniques to bound these payoffs appropriately.

Our proof techniques do not allow a strict characterization of the features of the equilibria of the model in addition to the no full learning claims. This is because of the bounding arguments we have just referred to. It is worth noticing that in the case of public communication we do know that in equilibrium there are “uniform” bounds that learning cannot exceed, and that periods of babbling and learning can alternate. In the case of private communication we know that babbling and learning can alternate and that there are equilibria in which all past is “erased” at exogenously give dates and behavior starts afresh as in period 0.

We also ask whether limit learning can be sustained in the sense of a sequence of equilibria that yields full learning in the limit, as discounting shrinks to zero. The two cases of private and public communication yield different results in this case. For some configurations of parameters, limit learning is possible with private communication but it is impossible in the public communication case.

We also examined the theoretical benchmark case of mediated communication, and showed full learning can be sustained in this case.

While we have restricted attention to pure strategies throughout, we strongly conjecture that our results carry over when mixing is in fact allowed.

As we have remarked already (see Subsection 3.1), mixed action strategies can be ruled out in any equilibrium by a simple use of concavity. To see why we consider it extremely unlikely that our main results would change when mixed message strategies are allowed, it is useful to consider again separately the one-period-ahead problem and the players’ continuation payoffs of Section 4 above. In the one-period-ahead problem, for any given  $p$  and  $x$ , with mixed messages (with overlapping supports) the difference in beliefs for the following player is clearly smaller than when mixing in messages is not allowed. Hence the deviation hypothesized there

---

Isaac Newton” chapter of Hawking (2003).

may become profitable when it would not have been otherwise. In other words, in the one-period-ahead problem mixed message strategies impede learning if anything at all.

As we have just recalled above, to go from the one-period-ahead to our main results we used bounds on the players' continuation payoffs. In the case of public communication, these were obtained by minimizing (or maximizing, as appropriate) continuation payoffs over all possible (message) strategy profiles. We believe that when mixed strategies are allowed it must be possible, if not easily tractable, to pin down similar bounds and hence establish that no full learning can occur.

Our results in this paper generalize to the case in which the state takes values in a finite set, and to correspondingly richer signal structures, provided that the imperfect signals are still informative and i.i.d.

We do not know how far the results generalize to models in which the signal quality can change through time. These changes could take place either exogenously because of, say, technological advances, or endogenously, because of investment choices on the part of the players. Both cases are worth exploring, but are obviously well beyond the scope of the present paper.

## Appendix

### A.1. The Proof of Proposition 1

**Definition A.1:** Fix any  $\delta, p, \alpha$  and  $r$ . Consider  $M$  and  $\hat{M}$ . Let  $\sigma^*$  and  $\hat{\sigma}^*$  be equilibrium strategy profiles when the message spaces are  $M$  and  $\hat{M}$  respectively.

We say that  $\sigma^*$  and  $\hat{\sigma}^*$  are outcome equivalent if and only if they induce the same stochastic process  $\{x_t\}_{t=1}^\infty$  on the players' posteriors for every realization of the state  $\omega$ .

**Lemma A.1:** Fix any  $M, \delta, p, \alpha, r$  and an equilibrium strategy profile  $\sigma^*$ . Then there exist an  $\hat{M} = \{\hat{M}_t\}_{t=0}^\infty$  with  $\hat{M}_t \subseteq M_t$  for every  $t$  and a  $\hat{\sigma}^*$  which is an equilibrium given  $\delta, p, \alpha, r$  and  $\hat{M}$  as follows.

Given  $\hat{\sigma}^*$  and  $\hat{M}$ , all messages are sent with positive probability for both values of  $\omega \in \{0, 1\}$ , and  $\hat{\sigma}^*$  is outcome equivalent to  $\sigma^*$  according to Definition A.1.

**Proof:** Recall that we are only considering pure strategies. It follows that the set of messages in each  $M_t$  that are sent with positive probability in any equilibrium for either value of  $\omega \in \{0, 1\}$  is finite. This is because it has at most the cardinality of all possible realizations of (binary) signal sequences up to and including  $t$ .

Now proceed to construct  $\hat{M} = \{\hat{M}_t\}_{t=0}^\infty$  from  $M = \{M_t\}_{t=0}^\infty$  by deleting all messages that are sent with probability zero for both values of  $\omega \in \{0, 1\}$  given  $\sigma^*$ . Construct  $\hat{\sigma}^*$  by letting  $\hat{\sigma}_t^*(m_{t-1}, s_t) = \sigma_t^*(m_{t-1}, s_t)$  for every  $t$ , every  $m_{t-1} \in \hat{M}_{t-1}$  and every  $s_t = \{0, 1\}$ .

Clearly,  $\hat{\sigma}^*$  is an equilibrium when the message spaces are  $\hat{M}$  since we have simply reduced the set of possible deviations. Also, by construction, the strategy profile  $\hat{\sigma}^*$  is outcome equivalent to the original strategy profile  $\sigma^*$ . ■

**Remark A.1:** In view of Lemma A.1, we only consider specifications of  $M$  and equilibria  $\sigma^*$  in which all messages are sent with positive probability for both values of  $\omega \in \{0, 1\}$  throughout the rest of the proof of Proposition 1.

Before proceeding further, some extra notation is needed. The entire batch is defined taking as given an  $M$  and an equilibrium strategy profile  $\sigma^*$  as in Remark A.1.

To begin with, define

$$X_t^B = \bigcup_{m_{t-1} \in M_{t-1}} x_t(m_{t-1}) \quad (\text{A.1})$$

so that  $X_t^B$  is the set of all possible beginning-of-period beliefs of player  $t$  that  $\omega = 1$ . Because of Remark A.1 these can all be computed using Bayes' rule.

We let  $x_t^E(m_{t-1}, s_t)$  denote the end-of-period belief of player  $t$  that  $\omega = 1$  when he receives message  $m_{t-1}$  and observes signal  $s_t = 0, 1$  during period  $t$ . Notice that using Bayes rule

$$x_t^E(m_{t-1}, s_t) = \frac{x_t(m_{t-1})\Pr(s_t|\omega = 1)}{x_t(m_{t-1})\Pr(s_t|\omega = 1) + [1 - x_t(m_{t-1})]\Pr(s_t|\omega = 0)} \quad (\text{A.2})$$

Consider player  $t$  and any message  $m_{t-1} \in M_{t-1}$  that player  $t$  can receive. We let  $U_t(m_{t-1}, \omega)$  be the discounted continuation payoff of player  $t$  if the state is  $\omega$  and he behaves as if he has received message  $m_{t-1}$  (in this case, we say player  $t$  *behaves according to*  $m_{t-1}$ ). The payoff is computed after player  $t$  has chosen his action and before he observes his signal  $s_t$ . The continuation payoff is normalized so that weight  $(1 - \delta)$  is given to the expected payoff in period  $t + 1$  and weight  $(1 - \delta)\delta^\tau$  is given to the expected payoff in each period  $t + \tau + 1$ .

Formally, for any  $\tau \geq 1$ , let  $a_{t+\tau}(m_{t-1}, s^{t,t+\tau-1})$  denote the action that player  $t + \tau$  will chose if player  $t$  behaves according to  $m_{t-1}$  and the sequence of signals between period  $t$  and period  $t + \tau - 1$  is  $s^{t,t+\tau-1} = (s_t, s_{t+1}, \dots, s_{t+\tau-1})$ .<sup>23</sup> Then  $U_t(m_{t-1}, \omega)$  is given by

$$U_t(m_{t-1}, \omega) = -(1 - \delta) \sum_{\tau=1}^{\infty} \delta^{\tau-1} \sum_{s^{t,t+\tau-1} \in \{0,1\}^\tau} \Pr(s^{t,t+\tau-1}|\omega) [\omega - a_{t+\tau}(m_{t-1}, s^{t,t+\tau-1})]^2 \quad (\text{A.3})$$

Lastly, notice that  $U_t(m_{t-1}, \omega)$  as defined in (A.3) satisfies the recursive relationship

$$U_t(m_{t-1}, \omega) = \Pr(s_t = 0|\omega) \left\{ -(1 - \delta) [\omega - a_{t+1}(m_{t-1}, s^{t,t} = 0)]^2 + \delta U_{t+1}[\sigma_t(m_{t-1}, s_t = 0), \omega] \right\} + \Pr(s_t = 1|\omega) \left\{ -(1 - \delta) [\omega - a_{t+1}(m_{t-1}, s^{t,t} = 1)]^2 + \delta U_{t+1}[\sigma_t(m_{t-1}, s_t = 1), \omega] \right\} \quad (\text{A.4})$$

**Lemma A.2:** Fix an  $M$  and  $\sigma^*$  as in Remark A.1. For any  $\eta > 0$  there exists an  $\varepsilon > 0$  such that the following is true for every  $t$ . Suppose that  $m_{t-1}$  and  $m'_{t-1}$  are two messages in  $M_{t-1}$  with  $\min\{x_t(m_{t-1}), x_t(m'_{t-1})\} > 1 - \varepsilon$  and  $x_t(m_{t-1}) \neq x_t(m'_{t-1})$ . Then

$$|U_t(m_{t-1}, 1) - U_t(m'_{t-1}, 1) - U_t(m_{t-1}, 0) + U_t(m'_{t-1}, 0)| < \eta \quad (\text{A.5})$$

**Proof:** Omitted. Available at <http://www.anderlini.net/learning-omitted-proofs.pdf>

We are now ready to proceed with the main part of the argument for Proposition 1.

**Proof of Proposition 1:** Fix any  $\delta, p, \alpha$  and  $r$ . Let also an  $M$  and  $\sigma^*$  as in Remark A.1. Suppose by way of contradiction that the equilibrium  $\sigma^*$  induces full learning as in Definition 1. Distinguish between the following two possibilities.

<sup>23</sup>Throughout the rest of the paper, for any  $t' \geq t$ , the array  $(s^t, \dots, s^{t'})$  will be denoted by  $s^{t,t'}$ .

Case (a). For every  $\varepsilon > 0$ , we can find a  $\bar{t}$  and a pair of messages  $m_{\bar{t}-1}$  and  $m'_{\bar{t}-1}$  with  $\min\{x_{\bar{t}}(m_{\bar{t}-1}), x_{\bar{t}}(m'_{\bar{t}-1})\} > 1 - \varepsilon$  and  $x_{\bar{t}}(m_{\bar{t}-1}) \neq x_{\bar{t}}(m'_{\bar{t}-1})$ .

and

Case (b). There exists  $\tilde{\varepsilon}$  such that for every  $t$  there is at most one element of  $X_t^B$  (the set of all possible beginning-of-period equilibrium beliefs of player  $t$  as in (A.1)) strictly greater than  $1 - \tilde{\varepsilon}$ .

Clearly one (and only one) of Case (a) and (b) must hold.

We begin with Case (a). Let  $\eta = (1 - \delta)\alpha/2\delta$ . Using Lemma A.2, we can pick an  $\varepsilon \in (0, \alpha/2)$ , a  $\bar{t}$  and a pair of messages  $m_{\bar{t}-1}$  and  $m'_{\bar{t}-1}$  as specified in Case (a) such that

$$|U_{\bar{t}}(m_{\bar{t}-1}, 1) - U_{\bar{t}}(m'_{\bar{t}-1}, 1) - U_{\bar{t}}(m_{\bar{t}-1}, 0) + U_{\bar{t}}(m'_{\bar{t}-1}, 0)| < \eta = \frac{(1 - \delta)\alpha}{2\delta} \quad (\text{A.6})$$

To keep notation usage down, during this part of the argument we let  $m_{\bar{t}-1} = m$ ,  $m'_{\bar{t}-1} = m'$ ,  $x_{\bar{t}}(m_{\bar{t}-1}) = x$  and  $x_{\bar{t}}(m'_{\bar{t}-1}) = x'$ . Without loss of generality, assume  $x > x'$ .

Clearly, there exists a type  $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1})$  of player  $\bar{t}-1$  who has the end-of-period belief  $x_t^E(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1}) \geq x'$  and sends message  $m'$ . We now show that type  $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1})$  has an incentive to deviate and send message  $m$ . To keep notation usage down, during this part of the argument we let  $\hat{x} = x_t^E(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1})$ .

Let  $\varphi(\hat{x})$  denote the difference between the continuation payoff of type  $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1})$  after he sends message  $m$  and the continuation payoff after he sends the equilibrium message  $m'$ . Using (2) and (A.4), the quantity  $\varphi(\hat{x})$  can be written as

$$\varphi(\hat{x}) = (1 - \delta)\bar{\varphi}(\hat{x}) + \delta(A\hat{x} + B) \quad (\text{A.7})$$

where

$$\bar{\varphi}(\hat{x}) = -\hat{x}(1 - x + \alpha)^2 - (1 - \hat{x})(x - \alpha)^2 + \hat{x}(1 - x' + \alpha)^2 + (1 - \hat{x})(x' - \alpha)^2$$

and

$$A = U_{\bar{t}}(m, 1) - U_{\bar{t}}(m', 1) + U_{\bar{t}}(m', 0) - U_{\bar{t}}(m, 0) \quad \text{and} \quad B = -U_{\bar{t}}(m', 0) + U_{\bar{t}}(m, 0) \quad (\text{A.8})$$

Notice that it follows from the incentive constraints of player  $\bar{t}$  (to whom the second term of the payoff difference (A.7) applies) that

$$Ax + B \geq 0 \geq Ax' + B \quad (\text{A.9})$$

which, since  $x > x'$ , implies that  $A \geq 0$ .

We now show that  $\varphi(\hat{x})$  is strictly positive and, therefore that type  $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1})$  of player  $\bar{t}-1$  has an incentive to deviate from his equilibrium strategy. We need to distinguish between two cases depending on whether  $\hat{x}$  is smaller or greater than  $x$ . We start with the case  $\hat{x} \in [x', x)$ .

A lower bound for the expression  $\bar{\varphi}(\hat{x})$  can be computed in the following way

$$\bar{\varphi}(\hat{x}) = (x - x')(2\alpha + 2\hat{x} - x - x') > (x - x')2(\alpha - \varepsilon) > (x - \hat{x})\alpha \quad (\text{A.10})$$

where the first inequality follows from  $\hat{x} > 1 - \varepsilon$  (and  $x < 1$  and  $x' < 1$ ), while the second inequality follows from  $\varepsilon < \alpha/2$ .

Consider now again the entire payoff difference  $\varphi(\hat{x})$ . We have

$$\begin{aligned} \varphi(\hat{x}) &> (1 - \delta)(x - \hat{x})\alpha + \delta(A\hat{x} + B) = \\ &(1 - \delta)(x - \hat{x})\alpha - \delta(x - \hat{x})A + \delta(Ax + B) \geq \\ &(x - \hat{x})[(1 - \delta)\alpha - \delta A] > (x - \hat{x})(1 - \delta)\frac{\alpha}{2} > 0 \end{aligned}$$

where the first inequality comes from (A.7) and (A.10), the second follows from (A.9), the third follows from (A.6) and the definition of  $A$  in (A.8), and the last one from the fact that we are considering the case in which  $\hat{x} \in [x', x]$ .

Consider now the case  $\hat{x} \geq x$ . We have that (recall that  $A \geq 0$ )

$$\begin{aligned} \varphi(\hat{x}) &= (1 - \delta)(x - x')(2\alpha + 2\hat{x} - x - x') + \delta(A\hat{x} + B) \geq \\ &(1 - \delta)(x - x')(2\alpha + x - x') + \delta(Ax + B) > 0 \end{aligned}$$

This proves that type  $(\hat{m}_{\bar{t}-2}, \hat{s}_{\bar{t}-1})$  of player  $\bar{t}-1$  has an incentive to deviate from the equilibrium prescription of sending message  $m'$ , and hence concludes the argument for Case (a).

We now turn to Case (b). By our contradiction hypothesis,  $\sigma^*$  induces full learning of the state. Therefore, since we are in Case (b) as identified above, we can find  $\tilde{t}$  such that for every  $t > \tilde{t}$  there exists a unique beginning-of-period belief for player  $t$  above  $1 - \tilde{\varepsilon}$ . This belief must be the largest element in  $X_t^B$  and we denote it by  $x_t^{\max}$ .

Next, fix  $\varepsilon'$  such that

$$\frac{(1 - \tilde{\varepsilon})p}{(1 - \tilde{\varepsilon})p + \tilde{\varepsilon}(1 - p)} < 1 - \varepsilon' \tag{A.11}$$

Given  $\varepsilon'$  we can find a period  $t(\varepsilon') > \tilde{t}$  such that for every  $t > t(\varepsilon')$

$$\frac{x_t^{\max}(1 - p)}{x_t^{\max}(1 - p) + (1 - x_t^{\max})p} > 1 - \varepsilon' \tag{A.12}$$

For every  $t$  let  $M_{t-1}^{\max} \subset M_{t-1}$  be the set of messages that induce belief  $x_t^{\max}$ , so that

$$M_{t-1}^{\max} = \{m_{t-1} \in M_{t-1} \mid x_t(m_{t-1}) = x_t^{\max}\}$$

Because  $\sigma^*$  induces full learning, the sequence  $\{x_t^{\max}\}_{t=1}^{\infty}$  must converge to one. Using (A.11) and (A.12) it is then easy to show that there must exist a period  $\bar{t}-1 > t(\varepsilon')$  and two types  $(m_{\bar{t}-2}, s_{\bar{t}-1})$  and  $(m'_{\bar{t}-2}, s'_{\bar{t}-1})$  in  $M_{\bar{t}-2}^{\max} \times \{0, 1\}$  which send two distinct messages, say  $m$  and  $m'$  respectively, such that

$$x_{\bar{t}}(m) > 1 - \varepsilon' \quad \text{and} \quad x_{\bar{t}}(m') < 1 - \tilde{\varepsilon}$$

It is also easy to check that when  $\varepsilon'$  is sufficiently small type  $(m'_{\bar{t}-2}, s'_{\bar{t}-1})$  has an incentive to deviate and send message  $m$  instead of  $m'$  as the equilibrium prescribes. The analysis is very similar to the analysis of Case (a) and we omit the details. The argument in Case (b) is therefore concluded. Hence, the proof of the proposition is now complete. ■

## A.2. The Proof of Proposition 2

We begin with some observations that do not require proof.

**Remark A.2:** Given that we restrict attention to pure strategies, there is no loss of generality in assuming that the set of messages  $M_t$  available to player  $t = 1, 2, \dots$  is  $\{0, 1\}$ . This is what we do in the remainder of the proof.

Hence, let  $M^0 = \emptyset$  and define, for each  $t \geq 1$ ,  $M^t = \{0, 1\}^t$ .  $M^t$  is the set of messages profiles of length  $t$ . Denote an arbitrary element by  $m^t = (m_1, \dots, m_t) \in M^t$ . Note that  $M^t$  contains the message profiles observed by player  $t + 1$ .

**Remark A.3:** Given the restriction to pure strategies, it is without loss of generality to restrict attention to strategy profiles  $\sigma$  such that:

$$\sigma_t(m^{t-1}, s_t = 1) = 1 \quad (\text{A.13})$$

for every  $t = 1, 2, \dots$ , and every  $m^{t-1} \in M^{t-1}$ . Thus, if the signal  $s_t$  is revealed, then this done by matching messages and signals. Moreover if the the signal is not revealed, then the message “1” is sent regardless of the signal observed.

We say that player  $t$  adopts a truthful strategy if

$$\sigma_t(m^{t-1}, s_t = 0) = 0$$

for every  $m^{t-1} \in M^{t-1}$ . We say that a strategy profile is truthful if each player adopts a truthful strategy. We denote the truthful strategy profile by  $\sigma^{TR}$ .

We also say that player  $t$  adopts a babbling strategy if

$$\sigma_t(m^{t-1}, s_t = 0) = 1$$

for every  $m^{t-1} \in M^{t-1}$ , so that, using (A.13) he sends message “1” regardless of  $s_t$ . We say that a strategy profile is babbling if each player adopts a babbling strategy. We denote the babbling strategy profile by  $\sigma^B$ .

**Remark A.4:** Fix a strategy profile  $\sigma$ . For every  $t \geq 1$  and every  $m^t \in M^t$ , recall that  $\Pr_\sigma(m^t | \omega)$  is the probability induced by strategy profile  $\sigma$  that player  $t + 1$  observes the message profile  $m^t$  when the state is  $\omega$ . We also let  $m^t(s^t, \sigma) \in M^t$  denote the message profile generated by  $\sigma$  when the signal profile is  $s^t$ .

Let

$$\mathcal{M}_\sigma^t = \bigcup_{s^t \in \{0,1\}^t} m^t(s^t, \sigma) \quad (\text{A.14})$$

Notice that  $\mathcal{M}_\sigma^t$  is the set of messages that are observed by  $t + 1$  with positive probability under  $\sigma$ , and that this set is independent of  $\omega$ . This is simply because the quality of the signals  $p$  is less than one and therefore every sequence of signals  $s^t$  has positive probability, regardless of the state.

Before proceeding further, some extra notation is needed. In the remainder of the proof, we will often use the notation  $x \in [0, 1]$ , and variations of it like  $x'$  to denote a player’s *prior*, in other words, the probability that this player attaches to the *original draw by nature* of  $\omega$  being equal to 1. While in the main body of the paper this is just defined as  $r \in (0, 1)$ , we find it convenient to proceed in this way here because of the way this prior will be made to change at various points in the proof. This will turn out to be a convenient way to take into account the players’ possible beliefs.

Fix a profile  $\sigma$ , and a prior for player  $t$  equal to  $x \in [0, 1]$ . By Bayes’ rule, after observing any  $m^{t-1} \in \mathcal{M}_\sigma^{t-1}$ , player  $t$ ’s beginning-of-period belief that  $\omega = 1$  must be

$$\Pr_\sigma(\omega = 1 | m^{t-1}, x) = \frac{x \Pr_\sigma(m^{t-1} | \omega = 1)}{x \Pr_\sigma(m^{t-1} | \omega = 1) + (1 - x) \Pr_\sigma(m^{t-1} | \omega = 0)} \quad (\text{A.15})$$

Recall that by (3), if his belief is as in (A.15), player  $t$  then chooses action  $a_t = \Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^{t-1}, x) - \alpha$ . To streamline the notation further, we let

$$\beta_t(s^{t-1}, \boldsymbol{\sigma}, x) = \Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^{t-1}(s^{t-1}, \boldsymbol{\sigma}), x) \quad (\text{A.16})$$

This is the beginning-of-period belief of player  $t$  if he has prior  $x$ , the players use strategy profile  $\boldsymbol{\sigma}$ , and the sequence of realized signals is  $s^{t-1}$ .

Some of our steps below are more easily conceptualized thinking by of a fictitious “player 0” added to the game. Player 0 does not make any choice, his payoff depends on the actions of future players and on  $\omega$  in the same way as all other players. Given a sequence of actions  $\{a_1, \dots, a_t, \dots\}$  and a state  $\omega$ , player 0’s payoff is

$$-(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} (\omega - a_t)^2$$

Our fictitious player 0 is also allowed to have a *different prior* from all other players  $1, 2, \dots$ . This will be particularly helpful below when possible deviations are entertained.

Fix a prior  $x$  for player 0 and let the prior of all other players be  $x'$ . Fix also a profile  $\boldsymbol{\sigma}$ . Given these we know what (contingent) actions future players will take. We then let  $V_x(\boldsymbol{\sigma}, x', t)$  be the expected one-period payoff that player 0 will receive in period  $t$ . Using (A.16) and (3), this can be computed as

$$\begin{aligned} V_x(\boldsymbol{\sigma}, x', t) = & - \sum_{s^{t-1} \in \{0,1\}^{t-1}} [x \Pr(s^{t-1} | \omega = 1) (1 - \beta_t(s^{t-1}, \boldsymbol{\sigma}, x') + \alpha)^2 \\ & + (1 - x) \Pr(s^{t-1} | \omega = 0) (\beta_t(s^{t-1}, \boldsymbol{\sigma}, x') - \alpha)^2] \end{aligned} \quad (\text{A.17})$$

When  $t = 1$ , (A.17) reduces to

$$V_x(\boldsymbol{\sigma}, x', 1) = -x(1 - x' + \alpha)^2 - (1 - x)(x' - \alpha)^2 \quad (\text{A.18})$$

Fix  $x, x'$  and  $\boldsymbol{\sigma}$  as above. Given any  $T \geq 1$  we let  $V_x^T(\boldsymbol{\sigma}, x')$  denote the discounted expected payoff to player zero in the first  $T$  periods. Hence, using (A.17)

$$V_x^T(\boldsymbol{\sigma}, x') = (1 - \delta) \sum_{t=1}^T \delta^{t-1} V_x(\boldsymbol{\sigma}, x', t) \quad (\text{A.19})$$

Still fixing  $x, x'$  and  $\boldsymbol{\sigma}$  as above, we let  $V_x(\boldsymbol{\sigma}, x')$  denote the overall discounted expected payoff of player zero. Hence, using (A.19)

$$V_x(\boldsymbol{\sigma}, x') = \lim_{T \rightarrow \infty} V_x^T(\boldsymbol{\sigma}, x') \quad (\text{A.20})$$

Finally, observe that, since this is true period-by-period, all three quantities in (A.17), (A.19) and (A.20) are linear in  $x$ .

**Lemma A.3:** Fix any  $p, \alpha$  and  $\delta$ . For any  $x \in [0, 1]$  and any strategy profile  $\boldsymbol{\sigma}$ ,

$$V_x(\boldsymbol{\sigma}^{TR}, x) \geq V_x(\boldsymbol{\sigma}, x).$$

**Proof:** It suffices to show that for every  $t \geq 1$

$$V_x(\boldsymbol{\sigma}^{TR}, x, t) \geq V_x(\boldsymbol{\sigma}, x, t). \quad (\text{A.21})$$

Consider first  $t = 1$ . It is apparent from (A.18) that  $V_x(\boldsymbol{\sigma}, x, 1)$  does not vary with  $\boldsymbol{\sigma}$ . The profile  $\boldsymbol{\sigma}$  only enters payoffs from date  $t = 2$  onward. Hence, (A.21) is trivially satisfied for  $t = 1$ .

Next, consider any date  $t + 1$  with  $t \geq 1$ . For any profile  $\boldsymbol{\sigma}$ , (A.16) and (A.17) imply

$$V_x(\boldsymbol{\sigma}, x, t + 1) = - \sum_{m^t \in \mathcal{M}_{\boldsymbol{\sigma}}^t} [x \Pr_{\boldsymbol{\sigma}}(m^t | \omega = 1)(1 - \Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^t, x) + \alpha)^2 + (1 - x) \Pr_{\boldsymbol{\sigma}}(m^t | \omega = 0)(\Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^t, x) - \alpha)^2] \quad (\text{A.22})$$

Using (A.15) and some simple algebra we get

$$x \sum_{m^t \in \mathcal{M}_{\boldsymbol{\sigma}}^t} \Pr_{\boldsymbol{\sigma}}(m^t | \omega = 1)(1 - \Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^t, x)) = (1 - x) \sum_{m^t \in \mathcal{M}_{\boldsymbol{\sigma}}^t} \Pr_{\boldsymbol{\sigma}}(m^t | \omega = 0) \Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^t, x) \quad (\text{A.23})$$

Expanding the square terms of the right-hand side of (A.22) and using (A.23), we can further rewrite  $V_x(\boldsymbol{\sigma}, x, t + 1)$  as

$$V_x(\boldsymbol{\sigma}, x, t + 1) = - \sum_{m^t \in \mathcal{M}_{\boldsymbol{\sigma}}^t} [x \Pr_{\boldsymbol{\sigma}}(m^t | \omega = 1)(1 - \Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^t, x))^2 + (1 - x) \Pr_{\boldsymbol{\sigma}}(m^t | \omega = 0)(\Pr_{\boldsymbol{\sigma}}(\omega = 1 | m^t, x))^2] - \alpha^2$$

And finally, using (A.16), we get

$$V_x(\boldsymbol{\sigma}, x, t + 1) = - \sum_{s^t \in \{0,1\}^t} [x \Pr(s^t | \omega = 1)(1 - \beta_{t+1}(s^t, \boldsymbol{\sigma}, x))^2 + (1 - x) \Pr(s^t | \omega = 0)(\beta_{t+1}(s^t, \boldsymbol{\sigma}, x))^2] - \alpha^2 \quad (\text{A.24})$$

Equation (A.24) is particularly useful to evaluate the difference between time- $t + 1$  payoffs induced  $\boldsymbol{\sigma}^{TR}$  and  $\boldsymbol{\sigma}$ . Under the truthful strategy profile  $\boldsymbol{\sigma}^{TR}$ , equality (A.16) reduces to

$$\beta_{t+1}(s^t, \boldsymbol{\sigma}^{TR}, x) = \frac{x \Pr(s^t | \omega = 1)}{x \Pr(s^t | \omega = 1) + (1 - x) \Pr(s^t | \omega = 0)} = \Pr(\omega = 1 | s^t, x) \quad (\text{A.25})$$

Then using (A.24) and (A.25) we can write,

$$V_x(\boldsymbol{\sigma}^{TR}, x, t + 1) - V_x(\boldsymbol{\sigma}, x, t + 1) = \sum_{s^t \in \{0,1\}^t} \left\{ - [x \Pr(s^t | \omega = 1)(1 - \Pr(\omega = 1 | s^t, x))^2 + (1 - x) \Pr(s^t | \omega = 0)(\Pr(\omega = 1 | s^t, x))^2] + [x \Pr(s^t | \omega = 1)(1 - \beta_{t+1}(s^t, \boldsymbol{\sigma}, x))^2 + (1 - x) \Pr(s^t | \omega = 0)(\beta_{t+1}(s^t, \boldsymbol{\sigma}, x))^2] \right\} \quad (\text{A.26})$$

Now fix any  $s^t \in \{0, 1\}^t$  and consider the corresponding element in the summation sign in (A.26). Dividing though by  $x\Pr(s^t|\omega = 1) + (1-x)\Pr(s^t|\omega = 0)$  we get

$$- [\Pr(\omega = 1|s^t, x)(1 - \Pr(\omega = 1|s^t, x))^2 + \Pr(\omega = 0|s^t, x)(\Pr(\omega = 1|s^t, x))^2] + [\Pr(\omega = 1|s^t, x)(1 - \beta_{t+1}(s^t, \sigma, x))^2 + \Pr(\omega = 0|s^t, x)(\beta_{t+1}(s^t, \sigma, x))^2] \quad (\text{A.27})$$

To check that this difference is nonnegative, observe that the quantity

$$\Pr(\omega = 1|s^t, x)(1 - z)^2 + \Pr(\omega = 0|s^t, x)z^2$$

as a function of  $z$  achieves a minimum at  $z = \Pr(\omega = 1|s^t, x)$ . Consequently,  $V_x(\sigma^{TR}, x, t+1) \geq V_x(\sigma, x, t+1)$  for  $t \geq 1$ , and hence the proof of Lemma A.3 is now complete. ■

**Lemma A.4:** Fix any  $p, \alpha$  and  $\delta$ . For any  $x \in [0, 1]$  and any strategy profile  $\sigma$ ,

$$V_x(\sigma, x) \geq V_x(\sigma^B, x).$$

**Proof:** The comparison between  $V_x(\sigma, x)$  and  $V_x(\sigma^B, x)$  can be carried out manipulating the former as in the proof of Lemma A.3, and evaluating the latter observing that the players' beliefs do not change through time under the babbling profile  $\sigma^B$ . The details are omitted for the sake of brevity. ■

**Remark A.5:** Fix a prior  $r$  and a signal quality  $p$ . Suppose that someone could observe a signal profile of arbitrary length, comprising  $k_1$  realizations of value 1 and  $k_0$  realizations of value 0. To state the next result we need to introduce two further pieces of notation. Let  $k = k_1 - k_0$ . Then, using Bayes' rule and simple algebra, the belief of this observer that  $\omega = 1$  would be given by

$$\gamma(r, p, k) = \frac{1}{1 + \frac{1-r}{r} \left( \frac{1-p}{p} \right)^k}$$

Hence,

$$\Gamma(r, p) = \{x \in (0, 1) \mid x = \gamma(r, p, k) \text{ for some } k = 0, \pm 1, \pm 2, \dots\} \quad (\text{A.28})$$

denotes the set of beliefs that can be generated by observing arbitrary signal profiles of any length given  $r$  and  $p$ .

**Lemma A.5:** Fix any  $p, \alpha$  and  $\delta$ . Then there exists an  $\bar{x} < 1$  such that for every  $x \in [\bar{x}, 1] \cap \Gamma(r, p)$  and every strategy profile  $\sigma$ ,

$$V_0(\sigma, x) \geq V_0(\sigma^B, x).$$

**Proof:** Omitted. Available at <http://www.anderlini.net/learning-omitted-proofs.pdf>

**Lemma A.6:** Fix  $p, \alpha$ , and  $\delta$ . There exists  $\bar{x} < 1$  such that for every  $x' \in [\bar{x}, 1] \cap \Gamma(r, p)$ , every  $x \leq x'$  and every strategy profile  $\sigma$ ,

$$V_x(\sigma, x') \geq V_x(\sigma^B, x')$$

**Proof:** Recall that for any pair  $(\sigma, x')$  the payoff  $V_x(\sigma, x')$  is linear in  $x$ . Therefore, the claim is an immediate consequence of Lemmas A.4 and A.5. ■

**Proof of Proposition 2:** Consider a player  $t$ , who observes the sequence of messages  $m^{t-1}$ , the signal  $s_t = 0$  and holds the associated end-of-period belief  $x$ . Using Bayes' rule, if instead of observing  $s_t = 0$  he were to observe  $s_t = 1$ , his posterior would be equal to

$$h(x) = \frac{xp^2}{xp^2 + (1-x)(1-p)^2} \quad (\text{A.29})$$

Clearly  $h(x) > x$  for every  $x \in (0, 1)$ .

Now suppose by way of contradiction that there is an equilibrium  $\sigma^*$  which induces full learning of the state as in Definition 1. Then, for any  $\varepsilon > 0$  we can find a  $t$  and a sequence of messages  $m^{t-1} \in \mathcal{M}_{\sigma^*}^{t-1}$  such that

$$\Pr_{\sigma^*}(\omega = 1 | m^{t-1}, r) > 1 - \varepsilon \quad (\text{A.30})$$

and

$$\sigma_t^*(m^{t-1}, s_t = 0) = 0 \quad (\text{A.31})$$

Let  $x$  denote the end-of-period belief of player  $t$  if he observes  $m^{t-1}$  and the signal  $s_t = 0$ . Clearly this can be computed applying Bayes' Rule to the left-hand side of (A.30). We take  $\varepsilon$  sufficiently small so that  $x > \bar{x}$ , where  $\bar{x}$  is as in Lemma A.6.

Suppose that type  $(m^{t-1}, s_t = 0)$  follows the equilibrium strategy and reveals his signal as required by (A.31). It then follows from Lemma A.3 that the continuation payoff (at the message stage) of type  $(m^{t-1}, s_t = 0)$  is bounded above by  $V_x(\sigma^{TR}, x)$ .

Suppose now that instead type  $(m^{t-1}, s_t = 0)$  deviates and sends message one. Using  $h(x) > x > \bar{x}$ , in this case it follows from Lemma A.6 that his continuation payoff (at the message stage) is bounded below by  $V_x(\sigma^B, h(x))$ .

It follows that  $\psi(x) = V_x(\sigma^B, h(x)) - V_x(\sigma^{TR}, x)$  is a lower bound on the gain from deviating from equilibrium at the message stage for type  $(m^{t-1}, s_t = 0)$ . Using simple algebra we can write  $\psi(x)$  as

$$\begin{aligned} & -x(1-h(x)+\alpha)^2 - (1-x)(h(x)-\alpha)^2 + (1-\delta)[x(1-x+\alpha)^2 + (1-x)(x-\alpha)^2] \\ & + (1-\delta) \sum_{t=1}^{\infty} \delta^t \sum_{s^t \in \{0,1\}^t} \left\{ x \Pr(s^t | \omega = 1) \left[ 1 - \frac{x \Pr(s^t | \omega = 1)}{x \Pr(s^t | \omega = 1) + (1-x) \Pr(s^t | \omega = 0)} + \alpha \right]^2 \right. \\ & \left. + (1-x) \Pr(s^t | \omega = 0) \left[ \frac{x \Pr(s^t | \omega = 1)}{x \Pr(s^t | \omega = 1) + (1-x) \Pr(s^t | \omega = 0)} - \alpha \right]^2 \right\} \end{aligned}$$

Notice next that  $\psi(1) = 0$  and  $h'(1) = [(1-p)/p]^2$ . Therefore  $\psi'(1)$  can be computed as

$$\begin{aligned} \psi'(1) &= -\alpha^2 + 2\alpha \left( \frac{1-p}{p} \right)^2 + (1-\alpha)^2 + (1-\delta)[\alpha^2 - 2\alpha - (1-\alpha)^2] + \\ & (1-\delta) \sum_{t=1}^{\infty} \delta^t \sum_{s^t \in \{0,1\}^t} \left[ \Pr(s^t | \omega = 1) \alpha^2 - 2\alpha \Pr(s^t | \omega = 1) \frac{\Pr(s^t | \omega = 0)}{\Pr(s^t | \omega = 1)} - \Pr(s^t | \omega = 0) (1-\alpha)^2 \right] \end{aligned}$$

which after simplification yields

$$\psi'(1) = 2\alpha \left[ \left( \frac{1-p}{p} \right)^2 - 1 \right] < 0$$

The latter clearly implies that for  $x$  sufficiently close to one we must have that  $\psi(x) > 0$ . Hence we have shown that, for  $x$  sufficiently close to one, type  $(m^{t-1}, s_t = 0)$  has an incentive to deviate from the putative equilibrium  $\sigma^*$  at the message stage. Therefore the proof of the proposition is now complete. ■

### A.3. Proof of Proposition 3

Fix  $r$  and  $p$ , and recall from (A.28) that  $\Gamma(r, p)$  is the set of beliefs that can be generated by observing arbitrary signal profiles of any length.

Let  $z$  denote the largest element of  $\Gamma(r, p)$  not exceeding  $1/2$ . In other words, let  $z = \max\{x \in \Gamma(r, p) \mid x \in (0, 1/2]\}$ . Consider a player with a beginning-of-period belief equal to  $z$ . After observing signals zero and one his beliefs are denoted  $z_0$  and  $z_1$  respectively. Using Bayes' rule, these can be trivially computed as

$$z_0 = \frac{z(1-p)}{z(1-p) + (1-z)p} \quad \text{and} \quad z_1 = \frac{zp}{zp + (1-z)(1-p)} \quad (\text{A.32})$$

From (A.32) and the way we have defined  $z$  it is evident that  $0 < z_0 \leq 1-p$  and  $1/2 < z_1 \leq p$ . Since by assumption  $\alpha < p - 1/2$  it then follows easily that for some  $\eta > 0$  it must be that for every  $a' \in [-\alpha, \eta - \alpha]$  and every  $a'' \in [1 - \alpha - \eta, 1 - \alpha]$  we have that both of the following inequalities hold

$$\begin{aligned} -z_0(1-a')^2 - (1-z_0)a'^2 &> -z_0(1-a'')^2 - (1-z_0)a''^2 \\ -z_1(1-a')^2 - (1-z_1)a'^2 &< -z_1(1-a'')^2 - (1-z_1)a''^2 \end{aligned} \quad (\text{A.33})$$

In other words, any player who believes that  $\omega = 1$  with probability  $z_0$  prefers any  $a' \in [-\alpha, \eta - \alpha]$  to any  $a'' \in [1 - \alpha - \eta, 1 - \alpha]$ . Moreover these preferences are reversed for any player who believes that  $\omega = 1$  with probability  $z_1$ .

It useful to observe at this point that since  $z \in \Gamma(r, p)$ , there must exist an integer  $K$  (positive or negative) that satisfies

$$z = \frac{rp^K}{rp^K + (1-r)(1-p)^K} \quad (\text{A.34})$$

Intuitively,  $K$  has the following property. Suppose that, beginning with a prior  $r$ , a player can observe a sequence of signals with the property that the number of signals equal to one minus the number of signals equal to zero is  $K$ . Then the player's belief that  $\omega = 1$  is precisely  $z$ . For the rest of the proof we let  $T$  be an integer such that  $T - 1 + K$  is an even positive integer.

Our construction relies on strategies that embody a “review phase” of length  $T$ , with  $T$  large. During this review phase the strategies keep track of the difference between the number of signals equal to one minus the number of signals equal to zero. If this difference is more than  $K$ , then the “outcome” of the review phase is “one,” while if the difference is less than or equal to  $K$  then the “outcome” of the review is “zero.”

It turns out that it is convenient for the argument that follows to write the basics of the review phase in terms of  $T$  and the number of signals equal to one, rather than in terms of the difference we have just mentioned. This is why we define the review phase “outcome function”  $R_T : \{0, 1\}^T \rightarrow \{0, 1\}$  as follows.

Consider an arbitrary sequence of signals  $s^T = (s_1, \dots, s_T)$  of length  $T$ , and let

$$R_T(s^T) = \begin{cases} 1 & \text{if } \sum_{t=1}^T s_t > \frac{T-1+K}{2} \\ 0 & \text{if } \sum_{t=1}^T s_t \leq \frac{T-1+K}{2} \end{cases}$$

The Weak Law of Large Numbers implies that for every  $\varepsilon > 0$ , there exists a  $\bar{T}$  such that for every  $T > \bar{T}$

$$\Pr[R_T(s^T) = i \mid \omega = i] > 1 - \varepsilon \quad \forall i \in \{0, 1\} \quad (\text{A.35})$$

and

$$\Pr[\omega = i \mid R_T(s^T) = i] > 1 - \varepsilon \quad \forall i \in \{0, 1\} \quad (\text{A.36})$$

We are now ready to specify the message spaces and strategies that induce limit learning of the state as required. Note that these will not depend on  $\delta$  even though this is allowed in principle by the statement of Proposition 3 (see footnote 17).

Fix an arbitrary  $\varepsilon > 0$  and an  $\eta > 0$  as in (A.33). Then using (A.36) find an integer  $T$  that satisfies (A.35) and such that

$$\bar{x}_1 = \Pr[\omega = 1 \mid R_T(s^T) = 1] > \max\{1 - \varepsilon, 1 - \eta\}$$

$$\bar{x}_0 = \Pr[\omega = 1 \mid R_T(s^T) = 0] < \min\{\varepsilon, \eta\}$$

and

$$\frac{\bar{x}_1(1-p)}{\bar{x}_1(1-p) + (1-\bar{x}_1)p} > z_1 \quad \frac{\bar{x}_0 p}{\bar{x}_0 p + (1-\bar{x}_0)(1-p)} < z_0 \quad (\text{A.37})$$

Recall that our claim is that limit learning can obtain in a Sequential Equilibrium in the sense of Kreps and Wilson (1982) (see Subsection 3.1), rather than just in a Perfect Bayesian Equilibrium. We accomplish this by constructing the message spaces in a particularly careful way, so that once the strategies are specified *all* messages will be sent in equilibrium with positive probability. In this way there will be nothing further to prove since off-path beliefs will not even need to be defined. With this in mind, we are now ready to proceed with the construction of message spaces.

In addition to the standard messages corresponding to integers, consider two additional ones labeled  $0^*$  and  $1^*$ . The integer messages are interpreted as counting the number of signals equal to one, while the  $0^*$  and  $1^*$  are used to “declare” the outcome of the review phase. The five possibilities for  $M_t$  so that no messages

are ever off-path are then as follows

- (i) If  $t < T$ ,  $t \leq \frac{T-1+K}{2}$  and  $t < \frac{T+1-K}{2}$  then  $M_t = \{0, 1, \dots, t\}$
- (ii) If  $t < T$ ,  $t > \frac{T-1+K}{2}$  and  $t < \frac{T+1-K}{2}$  then  $M_t = \{0, 1, \dots, \frac{T-1+K}{2}, 1^*\}$
- (iii) If  $t < T$ ,  $t \leq \frac{T-1+K}{2}$  and  $t \geq \frac{T+1-K}{2}$  then  $M_t = \{t - \frac{T-1-K}{2}, \dots, t, 0^*\}$
- (iv) If  $t < T$ ,  $t > \frac{T-1+K}{2}$  and  $t \geq \frac{T+1-K}{2}$  then  $M_t = \{t - \frac{T-1-K}{2}, \dots, \frac{T-1+K}{2}, 0^*, 1^*\}$
- (v) If  $t \geq T$  then  $M_t = \{0^*, 1^*\}$

We are now ready to construct the strategy profile  $\sigma^*$  that induces limit learning of the state as required. Intuitively, all players  $t \leq T$  truthfully participate in the review phase. They truthfully count the tally of signals equal to one they are given by their predecessors on the basis of the actual signal they observe. (Player 1 has no predecessor, so he just counts to one or to zero according to the signal he observes.) Once the review phase is declared either by  $0^*$  or  $1^*$  all players babble, indefinitely into the future.

Formally,  $\sigma_1^*(m_0, s_1) = s_1$  for all  $s_1 \in \{0, 1\}$ . For all  $t \geq 2$ , the strategy  $\sigma_t^*$  is then defined by

$$\sigma_t^*(m_{t-1}, s_t) = \begin{cases} m_{t-1} & \text{if } m_{t-1} \in \{0^*, 1^*\}, \text{ regardless of } s_t \\ m_{t-1} + 1 & \text{if } m_{t-1} + 1 \leq \frac{T-1+K}{2} \text{ and } s_t = 1 \\ 1^* & \text{if } m_{t-1} = \frac{T-1+K}{2} \text{ and } s_t = 1 \\ m_{t-1} & \text{if } m_{t-1} > t + \frac{K-T-1}{2} \text{ and } s_t = 0 \\ 0^* & \text{if } m_{t-1} \leq t + \frac{K-T-1}{2} \text{ and } s_t = 0 \end{cases} \quad (\text{A.38})$$

Observe that if the players use the strategy profile  $\sigma^*$  defined in (A.38), then, using (A.35) and (A.36), in any period  $t > T$  the beginning-of-period belief  $x_t$  will be greater than  $1 - \varepsilon$  (smaller than  $\varepsilon$ ) with probability at least  $1 - \varepsilon$  when the state is  $\omega = 1$  ( $\omega = 0$ ). Hence limit learning obtains under  $\sigma^*$ , as required.

It remains to check that  $\sigma^*$  is in fact an equilibrium for  $\delta$  sufficiently large.

Notice that in any period  $t \geq T$  only two messages will ever be sent, and only the two actions  $\bar{x}_0 - \alpha$  and  $\bar{x}_1 - \alpha$  will ever be chosen. It then follows easily from (A.33) and (A.37) that no player wants to deviate from the proposed equilibrium after  $T$ .

Finally, consider any player  $t \leq T$ . If he follows the proposed equilibrium strategy defined in (A.38), from  $T + 1$  onwards the following will take place. The action  $\bar{x}_1 - \alpha$  will be played if the prior  $r$  updated on the basis of the first  $T$  signal realizations yields a belief at least as large as  $z_1$ . Conversely,

$\bar{x}_0 - \alpha$  will be played if the prior  $r$  updated on the basis of the first  $T$  signal realizations yields a belief at most equal to  $z_0$ . These two actions are the only ones that will possibly be taken in period  $T + 1$  and all subsequent periods. Hence it follows from (A.33) that a player  $t$  who plays according to  $\sigma_t^*$  will induce his preferred action among these two with probability one in every period greater than  $T$ . Deviating from  $\sigma_t^*$  will necessarily reduce this probability to be strictly less than one. While following the equilibrium strategy might be costly for player  $t$  in periods before  $T$ , this is clearly sufficient to show that he will not want to deviate from the proposed equilibrium if  $\delta$  is close enough to one. ■

#### A.4. Proof of Proposition 4

In view of Remarks A.2 and A.3 we restrict attention to message spaces  $M_t = \{0, 1\}$  for every  $t$  and to strategy profiles that satisfy (A.13) since this is without loss of generality.

We begin by observing that, as we remarked already, inspection of the proof of Proposition 2 reveals that if we fix any equilibrium  $\sigma^*$  then there exists an  $\bar{x} < 1$  such that following any sequence of signals, the beginning-of-period belief  $x_t$  of any player  $t$  must satisfy  $x_t \leq \bar{x}$ , with equality holding for some  $t$ .

Using a similar line of argument it is possible to show that if we fix any equilibrium  $\sigma^*$  then there exists an  $\underline{x} > 0$  such that following any sequence of signals, the beginning-of-period belief  $x_t$  of any player  $t$  must satisfy  $x_t \geq \underline{x}$ , with equality holding for some  $t$ . We omit the details for the sake of brevity.

Recall that in (A.29) we defined  $h : [0, 1] \rightarrow [0, 1]$  as the belief  $x$  updated after the observation of two signals of one. Using the fact that by assumption  $p < \sqrt{2}/(1 + \sqrt{2})$  and Bayes' rule, simple algebra is then sufficient to show that there exists an  $\eta > 0$  such that for any  $x' \in (0, \eta)$  and any  $x \geq x'(1 - p)/[x'(1 - p) + (1 - x')p]$  it must be that

$$-x[1 - h(x) + \alpha]^2 - (1 - x)[h(x) - \alpha]^2 > -x[1 - x' + \alpha]^2 - (1 - x)[x' - \alpha]^2 \quad (\text{A.39})$$

Now suppose by way of contradiction that limit learning of the state is possible under the hypotheses of the proposition. Let  $\sigma^*(\delta)$  be the sequence of equilibria needed for limit learning to be possible as in Definition 2 as  $\varepsilon$  shrinks to zero and  $\delta$  approaches one. For each such equilibrium, let  $\underline{x}_\delta$  be the lower bound on beginning-of-period beliefs as we just defined above.

Since  $\eta$  as needed for (A.39) to hold is in fact fixed, it follows from our contradiction hypothesis that choosing  $\delta$  sufficiently large we can ensure that  $\underline{x}_\delta < \eta$ .

Hence there is at least one player, say  $t + 1$ , who with positive probability has beginning-period-belief  $\underline{x}_\delta < \eta$ . This implies that with positive probability player  $t$  has a beginning-of-period belief of  $\underline{x}_\delta p / [\underline{x}_\delta p + (1 - \underline{x}_\delta)(1 - p)]$  and that his strategy must be to truthfully reveal his signal in this case. This is because in this way player  $t$ 's belief after observing a signal of zero is precisely  $\underline{x}_\delta$ , and hence the lower bound on beginning-of-period beliefs is achieved in period  $t + 1$ .

Now consider the incentives of player  $t$  after observing  $s_t = 0$ . If he reports his signal truthfully as prescribed by the equilibrium then the action  $\underline{x}_\delta - \alpha$  will be chosen in every subsequent period. This is because the lower bound on beliefs  $\underline{x}_\delta$  will have been achieved, and this implies that all players will babble in the continuation equilibrium prescribed by  $\sigma^*(\delta)$ .

Suppose now instead that player  $t$  with beginning-of-period belief  $\underline{x}_\delta p / [\underline{x}_\delta p + (1 - \underline{x}_\delta)(1 - p)]$  deviates and reports a signal of one instead. This report is believed because the putative equilibrium  $\sigma^*(\delta)$  prescribes that  $t$  should tell the truth in this case.

Consider now an arbitrary period  $t + \tau$  with  $\tau \geq 1$  and an arbitrary sequence of signals  $s^{t, t+\tau-1} = (s_t, \dots, s_{t+\tau-1})$  with  $s_t = 0$ . Two cases are possible. The first case is that in period  $t + \tau$  the action  $\underline{x}_\delta - \alpha$  is chosen because the belief has reached its lower bound  $\underline{x}_\delta$ . In this case player  $t$  is clearly indifferent between telling the truth as prescribed by the equilibrium or deviating and reporting  $m_t = 1$ .

The second case is that  $s^{t, t+\tau-1}$  is such that by period  $t + \tau$  the lower bound for beliefs has not been reached and perpetual babbling has not necessarily begun. Let  $x$  be the belief that player  $t + \tau$  would have at this point if he knew that player  $t$  had deviated. Since  $t + \tau$  obviously does not know that  $t$  has deviated, his belief is simply  $h(x)$ . So, since the lower bound has not been reached we must have  $h(x) > \underline{x}_\delta$ . This implies that  $x \geq \underline{x}_\delta(1 - p) / [\underline{x}_\delta(1 - p) + (1 - \underline{x}_\delta)p]$ .

To sum up, in the second case in period  $t + \tau$  an action  $h(x) - \alpha$  is taken, where  $x$  is some belief satisfying  $x \geq \underline{x}_\delta(1 - p) / [\underline{x}_\delta(1 - p) + (1 - \underline{x}_\delta)p]$ . How does player  $t$  evaluate this situation? His updated belief must be in fact  $x$ . This is because in period  $t$  he observed a signal of zero, but in fact reported a signal of one, and

a belief of  $x$  updated on the basis of two signals of one is precisely  $h(x)$ . Given this belief, it follows from inequality (A.39) that player  $t$  strictly prefers action  $h(x) - \alpha$  to action  $\underline{x}_\delta - \alpha$ .

We conclude that the deviation is strictly profitable for  $t$  in every period  $t + 1, t + 2, \dots$  provided that the lower bound in beliefs has not been reached. Since the probability that the bound has not been reached in any period is obviously strictly positive, we can then conclude that the proposed deviation is strictly profitable. ■

### References

- ANDERLINI, L., D. GERARDI, AND R. LAGUNOFF (2008): “A ‘Super’ Folk Theorem for Dynastic Repeated Games,” *Economic Theory*, 37, 357–394.
- (2010): “Social Memory, Evidence and Conflict,” *Review of Economic Dynamics*, forthcoming.
- ANDERLINI, L., AND R. LAGUNOFF (2005): “Communication in Dynastic Repeated Games: ‘Whitewashes’ and ‘Coverups,’” *Economic Theory*, 26, 265–299.
- BANERJEE, A. (1992): “A Simple Model of Herd Behavior,” *Quarterly Journal of Economics*, 107, 797–817.
- CARRILLO, J., AND T. MARIOTTI (2000): “Strategic Ignorance as a Self-Disciplining Device,” *Review of Economic Studies*, 67, 529–44.
- CRAWFORD, V. P., AND J. SOBEL (1982): “Strategic Information Transmission,” *Econometrica*, 50, 1431–1351.
- FARRELL, J., AND M. RABIN (1996): “Cheap Talk,” *Journal of Economic Perspectives*, 10, 103–118.
- FORGES, F. (1986): “An Approach to Communication Equilibria,” *Econometrica*, 54, 1375–1385.
- FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. Cambridge, Massachusetts: The MIT Press.
- GALE, D. (1996): “What Have We Learned from Social Learning?,” *European Economic Review, Papers and Proceedings of the Tenth Annual Congress of the European Economic Association*, 40, 617–628.
- GOLOSOV, M., V. SKRETA, A. TSYVINSKI, AND A. WILSON (2008): “Dynamic Strategic Information Transmission,” MIT, NBER, & NYU, in preparation.

- HAWKING, S. W. (ed.) (2003): *On the Shoulders of Giants: The Great Works of Physics and Astronomy*. Philadelphia: Running Press.
- KARTIK, N. (2009): “Strategic Communication with Lying Costs,” *Review of Economics Studies*, Forthcoming.
- KOBAYASHI, H. (2007): “Folk Theorems for Infinitely Repeated Games Played by Organizations with Short-Lived Members,” *International Economic Review*, 48, 517–549.
- KREPS, D. M., AND R. WILSON (1982): “Sequential Equilibria,” *Econometrica*, 50, 863–894.
- KRISHNA, V., AND J. MORGAN (2001): “A Model of Expertise,” *Quarterly Journal of Economics*, 116, 747–775.
- (2008): “Cheap Talk,” in *The New Palgrave Dictionary of Economics*, ed. by L. Blume, and S. Durlauf, 2<sup>nd</sup> edn. New York, NY: McMillan.
- LAGUNOFF, R. (2006): “Credible Communication in Dynastic Government,” *Journal of Public Economics*, 90, 59–86.
- LAGUNOFF, R., AND A. MATSUI (2004): “Organizations and Overlapping Generations Games: Memory, Communication, and Altruism,” *Review of Economic Design*, 8, 383–411.
- LI, W. (2007): “Changing One’s Mind when the Facts Change: Incentives of Experts and the Design of Reporting Protocols,” *Review of Economic Studies*, 74, 1175–1194.
- MAILATH, G., AND L. SAMUELSON (2006): *Repeated Games and Reputations: Long Run Relationships*. Oxford: Oxford University Press.
- MOSCARINI, G., M. OTTAVIANI, AND L. SMITH (1998): “Social Learning in a Changing World,” *Economic Theory*, 11, 657–665.
- MYERSON, R. B. (1982): “Optimal Coordination Mechanisms in Generalized Principal-Agent Problems,” *Journal of Mathematical Economics*, 10, 67–81.
- OTTAVIANI, M., AND P. N. SØRENSEN (2006): “Professional Advice,” *Journal of Economic Theory*, 126, 120–142.

- SMITH, L., AND P. SØRENSEN (2008): “Information Herding and Optimal Experimentation,” working paper.
- SOBEL, J. (2000): “Economists’ Models of Learning,” *Journal of Economic Theory*, 94, 241–261.
- (2007): “Signaling Games,” UCSD, mimeo.
- SPECTOR, D. (2000): “Rational Debate and One-Dimensional Conflict,” *Quarterly Journal of Economics*, 115, 181–200.
- TAUB, B. (1997): “Dynamic Agency with Feedback,” *Rand Journal of Economics*, 28, 515–543.