Strategic Learning and Information Transmission

Alexander Frug†

November 5, 2014

Abstract

This paper addresses the problem of information transmission between a biased expert and a decision maker in an environment where the expert’s learning is gradual and strategic. It is shown how the gradualness of the expert’s learning can be exploited to enhance informativeness in communication. The result suggests that even in the absence of an “objective” reason to expedite information transmission, putting the expert under “strategic pressure” can be beneficial to both players.

1 Introduction

Consider the following scenario. A patient (decision maker) can be healthy (state 0) or ill, and in the latter case he can have either a mild disease (state 1) or a severe disorder (state 2). The three states are equally likely. The patient has to determine the level of his health care expenditure \( a \). His first-best level of health care expenditure at state \( \theta \) is \( a = \theta \). The doctor (expert), who stands to benefit from the treatment, can advise the patient. If the patient is healthy, the interests of the players are aligned. If the patient is ill, the doctor is upwardly biased and her most desired level of patient expenditure in state \( \theta \in \{1, 2\} \) is \( a = \theta + 1 \). Both players minimize the (expected) quadratic loss.

The case in which the doctor is perfectly informed and the patient is uninformed corresponds to the model of cheap talk analyzed in the seminal contribution of Crawford and Sobel (1982). Under this assumption on the players’ information, in the optimal equilibrium, the patient can only learn whether he is healthy (\( \theta = 0 \)) or ill (\( \theta \in \{1, 2\} \)).

We now enrich the standard cheap-talk environment with an explicit description of the process according to which the information asymmetry between the players arises. Assume that both players are initially uninformed but the doctor has the ability to explore the patient’s condition as follows. Two medical examinations are available: \( \epsilon_1 \) and \( \epsilon_2 \). Examination \( \epsilon_i \) is costless and it privately reveals to the doctor whether \( \theta = i \) or not. The patient knows which examinations are available to the doctor but does not

---

* I thank Elchanan Ben-Porath, Eddie Dekel, Kfir Eliaz, Ariel Rubinstein, and Ran Spiegler for useful comments and suggestions. I am also grateful to Yair Antler, Daniel Bird, and Ran Eilat for long fruitful discussions. I owe a special acknowledgment to my advisor Ady Pauzner for his guidance.

†Tel-Aviv University. Email: frugalex@post.tau.ac.il.
not observe which is actually performed. It is known that each examination takes one unit of time. Yet, the time available for the doctor before the patient has to make a decision, enough to fully explore the patient’s condition.

In this case, it turns out that the patient can accurately learns his health condition. The doctor’s strategy described below is consistent with an equilibrium. First, the doctor performs examination $\epsilon_2$. If the result is positive, she reports this fact immediately. Otherwise, she proceeds to examination $\epsilon_1$. This takes another unit of time. When the result of $\epsilon_1$ is obtained the doctor (truthfully) reports it.

To see why this is an equilibrium, it is important to note that when the doctor chooses her learning strategy, she takes into account that she has only one opportunity to induce the highest possible expenditure of the patient ($a = 2$). This can be done only by reporting that “$\theta = 2$” at the end of period 1. Since only one examination can be performed at each unit of time, the doctor has to choose which to perform first. It turns out that it is most desirable from her perspective to begin with $\epsilon_2$ and report that “$\theta = 2$” only if the result of $\epsilon_2$ is positive. If the outcome of $\epsilon_2$ is negative, inducing $a = 2$ is not desirable because, at this point, she does not know whether $\theta = 1$ or $\theta = 0$. She prefers to continue learning and report truthfully at the end of period 2.

The timing of the report provides the patient with some information regarding the doctor’s information structure. The key element in the equilibrium presented above is that if a report is submitted at the end of period 1, the patient infers that the doctor could not perform both medical examinations. Although, there is no objective reason to expedite information transmission, in the most informative equilibrium, the doctor is put under strategic pressure which induces a problem of strategic scheduling of the available examinations. Ex-ante, the fully revealing equilibrium is strictly preferred by both players over any other equilibrium.

This paper combines a dynamic cheap-talk environment with gradual learning on the part of an expert that is strategic, unobservable, and unverifiable. At the beginning, both the expert and the decision maker are uninformed about the state of the world but the expert has the ability to explore it. The expert’s learning is carried out under some technological limitations; certain experiments can be performed over time. Nevertheless, as is often the case, some aspects of the exploration process are left to the expert’s discretion, e.g. the selection of a particular collection of experiments to be performed and the ordering of these experiments. To emphasize the particular effect of the gradualness in the expert’s learning I assume that the learning is costless. The purpose of this analysis is to show how the gradualness of an expert’s learning can be exploited for the sake of more informative communication.

**Related Literature**

The literature on costless communication (cheap talk) between informed experts and uninformed decision makers began with the contributions of Crawford and Sobel (1982) and Green and Stokey (1980).\(^1\) Since then, many authors have studied a variety of cheap-talk environments under different specifications. Sobel

---

\(^1\)Earlier literature on costly signaling can be traced back to Spence (1973).
(2010) provides a comprehensive literature review of the communication literature. Several previous works consider environments where the sender is imperfectly informed. Fisher and Stocken (2001) showed that, in equilibrium, the accuracy of the receiver’s information is not monotonic in the quality of the sender’s information. This was extended by Ivanov (2010) who characterized the optimal static information structure from the receiver’s perspective for the leading uniform-quadratic case. In a recent contribution by Ivanov (2012) an optimal two-stage learning protocol is suggested. The author assumes that the receiver can perfectly control the dynamic structure of the sender’s information. By coupling distant separable elements at the early stage and by conditioning future experiments on early truthful communication, the receiver can successfully elicit full information. The sender is initially uninformed; however, her learning is not strategic but designed by the receiver in these models.

Several works have also considered certain aspects of the sender’s strategic learning. Austen-Smith (1994) proposed a model in which the sender chooses whether to learn (perfectly) the state and has the ability to prove information acquisition. The choice whether to become informed depends on the realization of the cost of learning that is privately observed. The author shows that, relative to Crawford and Sobel (1982), more information can be transmitted in equilibrium. As in our model, the choice whether to become informed is left to the sender’s discretion. Unlike in our model, the learning is costly, verifiable, and non-gradual. In Argenziano et al. (2011) the sender can affect the quality of her information by selecting how many Bernoulli trials to perform. The trials are costly for the sender. The authors compare “covert” and “overt” selection of the number of trials to be performed and show that under the overt regime it is possible to force the sender to over invest in learning. In Kamenica and Gentzkow (2011) the sender selects a “signal,” i.e., an information structure used to persuade a decision maker. While the selection of the information structure is costless and left to the sender’s discretion, the receiver observes both the information structure and its realization. In Gentzkow and Kamenica (2012) the sender also publicly selects the information structure but now it is no longer costless. A more informative information structure is associated with a higher cost. Since the information is costly for the sender, endogenous information will always be disclosed in equilibrium and so disclosure requirements have no effect on the set of equilibrium outcomes. A related result appears in Di Pei (2013). In that paper the sender gathers costly information before advising the receiver. The sender communicates all her information and all equilibria are less informative than the most informative one in Crawford and Sobel (1982). As in our model, learning is strategic and unobservable by the receiver. Unlike in our model, information is costly.

All of the above models that contain strategic learning are essentially static. However, the key feature of the model studied in this paper is the dynamic nature of the sender’s learning. The timing of the sender’s reports provides a signal to the receiver about the quality of the sender’s information. Frug (2014a) considers a model of gradual exogenous learning and communication. In this model, by designing an appropriate reporting protocol, the players manage to exploit the gradualness of the sender’s information arrival for the sake of more informative communication. The present study suggests that players can benefit from the gradualness of the expert’s learning even when the expert’s learning is strategic (endogenous),

\[^2\text{A similar idea appears in Golosov et al. (2013).}\]
The rest of the paper is organized as follows. Section 2 presents the model. In Section 3 I analyze the model in a convenient state space and discuss the interpretation of strategic pressure and the manner in which an upwardly biased expert explores the state of the world. Section 4 discusses the role of a special component of the convenient state space addressed earlier and develops a related result in a more standard state space. In Section 5 the results given in previous Sections are compared with the informed sender benchmark. Sections 6 and 7 illustrate how the ideas presented apply to other environments. Section 6 provides an example of a case where the expert is biased towards extreme actions and Section 7 discusses alternative learning technologies. Section 8 concludes.

2 Model

There are two players - sender and receiver, \( N = \{S,R\} \).

State Space To present the forces behind the main result in the most transparent way, I begin the analysis with a convenient state space \( \Theta = \{\phi, 1, 2, \ldots, n\} \) where \( \phi \) represents a “categorically” different state (intuitively, \( \phi \) corresponds to the state in which the patient is healthy in our doctor-patient example). Let \( P \) denote the common prior on \( \Theta \).

The inclusion of \( \phi \), along with the associated assumptions given below, has clear technical implications and this was my primary motivation. I discuss the role of \( \phi \) and develop a related result in a state space that does not contain \( \phi \) in Section 4.

Sender’s Learning For each \( k \in \{1, \ldots, n\} \), there is a potential experiment \( \epsilon_k = 1_{\theta=k} \) that reveals whether \( \theta = k \) or not. All of the experiments entail no direct costs for none of the players; however, conducting an experiment takes one unit of time (parallel experimentation is not possible). Only the sender has access to the set of experiments \( \{\epsilon_k|k \in \{1, \ldots, n\}\} \). Which experiments she has conducted and in what order is unobservable and non-verifiable.

Notice that there does not exist an experiment \( \epsilon_\phi \). In other words, the realization \( \theta = \phi \) can be learned by the sender only by means of ruling out all other states (In our doctor-patient example, the doctor learns that the patient is healthy by ruling out all diseases).

Sender’s Reporting I assume that the sender is allowed to submit only one report but she can choose the timing of the report. Formally, let \( M \) be a set of available reports and let \( \sigma \in M \) denote the “empty report,” corresponding to the sender’s choice to “remaining silent.” Assume that the set \( M \) is rich enough to transmit any information available for the sender.

Let \( T = \{1, 2, \ldots, T\} \) be a discrete time line, and assume \( T > n \). At each \( t \in T \), the sender may remain silent (i.e. submit an “empty report”) or submit an active report \( m_t \in M - \{\sigma\} \). Any terminal history of reports contains at most one “non-empty report.” The reports are cheap talk in the sense that they have no direct effect on players’ utility.

It is worthwhile to note that the main result of the paper does not rely on the selection of this particular reporting protocol. In other words, every equilibrium under the selected single report protocol
can be recovered in an environment that places no restrictions on the sender’s reporting opportunities. The assumed reporting protocol highlights the main aspect of the reporting strategy - the timing of the report.

Receiver’s Action  At period $T$, the receiver has to choose an action $a \in \{ \phi \} \cup \mathbb{R}_+$. It is convenient to interpret the action $\phi$ as “the choice of avoiding to take an active action”. The assumption that $T > n$ is to capture the idea that, in principle, the time available for learning places no restrictions on the expert’s ability to become fully informed.

Utility Functions  Player $i$’s utility function is

$$U^i(a, \theta) = \begin{cases} u^i(a, \theta) & a, \theta \neq \phi \\ 0 & a = \theta = \phi \\ -\infty & a = \phi \oplus \theta = \phi \end{cases}$$

where the symbol $\oplus$ denotes the “exclusive or” logical operator. The assumption that $U^i(a, \theta) = -\infty$ if $a = \phi \oplus \theta = \phi$ captures the intuition that any mismatch with respect to $\phi$ is very costly for the players (intuitively, operating a healthy patient and avoiding to take care of an ill patient are both strongly undesirable). The simplifying assumption of the infinite cost is stronger than required for the main result and it is discussed at the end of the analysis in Section 3.

The functions $u^i : \mathbb{R}_+ \times \{1, 2, \ldots, n\} \rightarrow \mathbb{R}$ for $i \in \{S, R\}$, satisfy the following assumptions. For each $\theta$, the function $u^i(\cdot, \theta) \leq 0$ represents single-peaked preferences, $a^i(\theta) = \arg\max_a u^i(a, \theta)$ is (strictly) increasing in $\theta$, and the sender is upwardly biased in the sense that for all $\theta < n$,

$$u^S(a^R(\theta), \theta) < u^S(a^R(\theta + 1), \theta)$$

Strategies and Equilibrium  The sender’s strategy is twofold. She chooses both how to explore and how to report. Clearly, the flexibility of the reporting component of the strategy, $m = \{m_t\}_{t \in \mathbb{T}}$, depends on its learning component $f$ because the report at each period can depend only on the information collected up to this period.

To make it precise let $f : \mathbb{T} \rightarrow \{0, 1, \ldots, n\}$ be a given plan of learning, specifying which experiment $\epsilon_{f(t)}$ to be conducted at date $t \in \mathbb{T}$, where $\epsilon_0$ denotes “do nothing”. Denote by $\mathcal{F}_i(f)$ the sender’s information structure at period $t$ if she follows the learning plan $f$. In particular, $\mathcal{F}_i(f) = \{ \{ f(s) \}, (\cup \{ f(s) \})^c | s \leq t \}$. A report choice in period $t$, $m_t$, must be measurable with respect to $\mathcal{F}_i(f)$. The reporting component of the sender’s strategy, $m = \{m_t\}_{t \in \mathbb{T}}$, is “$f$ – measurable” if $m_t$ is measurable with respect to $\mathcal{F}_i(f)$ for all $t$.

Note that the continuation of learning is relevant only if previous experiments have not revealed $\theta$ (and a “non-empty report” has not been submitted), Thus, it is without loss to assume that the whole plan of learning $f$ is selected at the beginning of the interaction and is not periodically revised. To summarize, a sender’s strategy is $s \in \Delta\{ (f, m) | f \in \{0, 1, \ldots, n\}^\mathbb{T} \text{ and } m \text{ is } f – \text{measurable} \}$. Whenever $f$ is not random,
simply denote $s = (f, m^f)$, where $m^f$ is $f$-measurable and possibly mixed.

The receiver chooses an action rule $\alpha(m) \in \Delta A$ that attaches to any terminal history of reports a distribution over actions. Receiver’s pure strategy is denoted by $a(m)$.

The applied solution concept is sequential equilibrium.

3 Analysis

In the main part of the analysis I illustrate how strategic pressure can increase the amount of information transmitted by showing that, under some conditions, there exists an equilibrium in which the receiver reveals the state of the world. An equilibrium in which the sender acquires complete information and transmits it accurately is termed here a fully informative equilibrium. Proposition 1 characterizes the essentially unique sender’s strategy in any fully informative equilibrium. Qualitatively, this has two implications. First, due to strategic pressure, the sender will perform her exploration as fast as possible. Strategic pressure implies that certain actions can be induced only during the early stages of the game. To decide whether to induce a particular action, the sender would like to proceed as much as she can with the most effective exploration program (from her perspective). Second, in a fully informative equilibrium, an upwardly biased sender will necessarily explore the states downwards (against the bias direction). Actions towards which she is “most biased” are the first to become no longer inducible.

Proposition 1 In any fully informative equilibrium, the sender’s strategy $(f, m^f)$ satisfies, for every $t < n$,

- $f(t) = n + 1 - t$, and
- $m^f_t \neq \sigma$ if and only if $\epsilon_{f(t)} = 1$.

Proof Let $e$ be a fully informative equilibrium and denote by $(f, m^f)$ the sender’s strategy. For each $k \in \Theta - \{\phi\}$, denote by $t(k)$ the last period in which the information that $\theta = k$ is transmitted to the receiver in this equilibrium. A necessary condition for a fully informative equilibrium is that for every $k > 1$, $\{k - 1\} \notin \mathcal{F}_t(k)(f)$. Otherwise, upon learning the fact that $\theta = k - 1$, the sender is better off pretending that $\theta = k$. Thus, it follows that $t(n) < t(n-1) < \ldots < t(2) < t(1)$. Moreover, if $t(k) > n+1-k$ for some $k > 1$, it follows that $t(2) \geq n$. In this case, the sender has a profitable deviation to a strategy by which she performs all the experiments by $t = n$ and induces the action $\arg\max_a u^R(a, \theta = 2)$ in case $\theta = 1$.

We established that for any $k > 1$, $t(k) = n + 1 - k$. Thus, for $t < n$ it must hold that $f(t) = n + 1 - t$. Moreover, different realizations of $\epsilon_k$ (an experiment that is performed at $t(k)$) must induce distinct reports at $t(k)$. Submitting $m_{t(k)} \neq \sigma$ terminates the game. Since $e$ is fully informative, $\epsilon_k = 0$ must be followed by $m_{t(k)} = \sigma$. QED
Next, a sufficient condition for a fully informative equilibrium is provided. This condition relates the sender’s preferences to the prior distribution. It replaces the standard supermodularity condition that is often assumed in information economics when considering interactions with informed agents.

The sender’s preferences satisfy \emph{p-supermodularity} if, for every \( \bar{a} > a \) and \( \bar{\theta} > \theta \), the following inequality holds:

\[
p(\bar{\theta})[u^S(\bar{a}, \bar{\theta}) - u^S(a, \bar{\theta})] > p(\theta)[u^S(\bar{a}, \theta) - u^S(a, \theta)].
\]

**Proposition 2** If the sender’s preferences satisfy p-supermodularity, a fully informative equilibrium exists.

**Proof** We begin the proof with the following observation.

**Observation 1** Since the learning is unobservable and nonverifiable, and it consists of deterministic and costless experiments, we can assume without loss of generality that the sender’s (pure) learning plan is a permutation \( f \in S_n \).

Consider the sender’s strategy \( s^* = (f, m^f) \), where \( f(t) = n + 1 - t \) and \( m^f_t \neq \sigma \) if and only if \( \epsilon_{f(t)} = 1 \). The observable part of the sender’s strategy (the reporting component) has full support (whenever \( m_t \neq \sigma \) let the sender report a random message in \( M \)). Therefore, the whole system of the receiver’s beliefs is pinned down by the Bayes’ rule. Accordingly, the receiver’s best response is to select \( a^*([m_t]_{t \in T}) = a^R(n + 1 - \tau) \) if \( m_\tau \neq \sigma \) for some \( \tau \leq n \), and \( a^*(\sigma) = \phi \) where \( \sigma \) denotes \( m_t = \sigma \) for all \( t \leq n \). It is left to show that \( (f, m^f) \) is the sender’s best response to \( a^*(\cdot) \).

First, we show that \( m^f \) is optimal given \( f \). Since \( \phi \in \Theta \), it cannot be optimal for the sender to submit a “false positive” report (to induce \( a \neq \phi \) as long as the possibility that \( \theta = \phi \) is not ruled out). On the other hand, once \( \theta \) is revealed to the sender, she cannot benefit from delaying the report because \( u^S(a, \theta) \) represents single-peaked preferences for each \( \theta \) and \( a^R(\theta) < a^S(\theta) \).

We now show that there does not exist an alternative learning plan \( g \), and a “\( g \)-measurable” reporting strategy \( m^g \) such that \( (g, m^g) \) constitutes a profitable deviation.

Let \( g \in S_n \) such that \( g \neq f \). Let \( k \) be the minimal integer with the property that \( g(k) > f(k) \). Define \( g' = (g(k), g(k - 1)) \circ g \) and let \( m^g \) and \( m^{g'} \) denote some optimal reporting plans given \( g \) and \( g' \) respectively.\(^3\) We now show that the sender prefers \( (g', m^{g'}) \) over \( (g, m^g) \). Since at least one of the pure learning plans in \( S_n \) is consistent with a best response, this concludes the proof.

Since \( \phi \in \Theta \), the reporting strategies \( m^g \) and \( m^{g'} \) are essentially the same (they induce identical receiver’s actions) whenever \( \theta \notin \{g(k), g(k - 1)\} \) the sender never induces \( a \neq \phi \) as long as \( \theta = \phi \) is not

\(^3\)For a permutation \( f \in S_n \) and \( i, j \in \{1, \ldots, n\} \), \((i, j) \circ f \) denotes the permutation obtained by swapping the elements \( i \) and \( j \) in \( f \).
ruled out, and \( g^{-1}(\theta) = g^{-1}(\theta) \) for every \( \theta \notin \{g(k), g(k - 1)\} \). That is, every such \( \theta \) is learned by the sender at identical times under both learning plans \( g \) and \( g' \). Thus, to compare \( g \) and \( g' \) it suffices to evaluate the change in the utilities at the states \( \theta \in \{g(k), g(k - 1)\} \).

From the definition of \( k \), it follows that \( g(k) > g(k - 1) \). Otherwise, \( g(k - 1) \geq g(k) + 1 > f(k) + 1 = f(k - 1) \).

Let \( \bar{a} \) and \( \bar{a} \) denote the actions that are taken by the receiver after \( m_t \neq \sigma \) at \( t = k \) and \( t = k - 1 \) respectively. Recall that \( \bar{a} < \bar{a} \). As \( g(k) > f(k) \), expediting \( \epsilon_{g(k)} \) from \( t = k \) to \( t = k - 1 \) generates an (expected) improvement of

\[
p(\theta = g(k)) \cdot [u^S(\bar{a}, g(k)) - u^S(\bar{a}, g(k))].
\]

On the other hand, a delay of \( \epsilon_{g(k-1)} \) from \( t = k - 1 \) to \( t = k \) can have one of two possible consequences. If \( u^s(\bar{a}, g(k - 1)) < u^s(\bar{a}, g(k - 1)) \), such a delay does not affect the sender’s utility because the optimality of \( m_0 \) implies that if \( \epsilon_{g(k-1)} = 1 \), then \( m_0 = \sigma \) for some \( \tau \geq k \) (it induces an action lower than \( \bar{a} \)). This means that under \( g' \), the sender can induce the same action at \( \theta = g(k - 1) \) since, as in the case according to \( g \), the experiment \( \epsilon_{g(k-1)} \) is performed no later than \( \tau \).

Alternatively, if \( u^s(\bar{a}, g(k - 1)) \geq u^s(\bar{a}, g(k - 1)) \) a delay of \( \epsilon_{g(k-1)} \) from \( t = k - 1 \) to \( t = k \) amounts to the expected loss of

\[
p(\theta = g(k)) \cdot [u^S(\bar{a}, g(k - 1)) - u^S(\bar{a}, g(k - 1))].
\]

In this case, if \( \epsilon_{g(k - 1)} = 1 \) then \( m_0' = \sigma \), because \( u^s(\cdot, g(k - 1)) \) represents single-peaked preferences.

To summarize: the net value of the swap is

\[
p(\theta = g(k)) \cdot [u^S(\bar{a}, g(k)) - u^S(\bar{a}, g(k))] + p(\theta = g(k - 1)) \cdot \max\{0, [u^S(\bar{a}, g(k)) - u^S(\bar{a}, g(k))]\} \geq p(\theta = g(k)) \cdot [u^S(\bar{a}, g(k)) - u^S(\bar{a}, g(k))] + p(\theta = g(k - 1)) \cdot [u^S(\bar{a}, g(k)) - u^S(\bar{a}, g(k))] > 0.
\]

where the last inequality follows from the \( p \)-supermodularity of \( u^S \).

Thus, \( s^* = (f, m') \) is a best response to \( a^*(\cdot) \) and the proposition follows. QED

I interpret this result as follows. Even though there is no “objective” reason to expedite information transmission, sometimes it is possible to put the expert under an effective form of “strategic” pressure. In particular, for the expert not to have an opportunity to inflate the report in the direction of her bias, it is important that the set of inducible actions shrink sufficiently fast over time in the direction opposite to the bias. In other words, high actions can be induced by the sender only during the early stages of the interaction. Of course, for the communication to be efficient, it is also important that the set of inducible actions shrink at a pace that is in keeping with the expert’s physical ability to acquire information.

In Sections 6 and 7, it is shown that the intuitions of strategic pressure and exploration manner (in terms of direction) also apply to other situations in which the sender’s information is acquired through a different learning technology and the bias is not of constant direction.

It is worthwhile to note that the assumption that \( U^s(a, \theta) = -\infty \) if \( a = \phi \oplus \theta = \phi \) is clearly stronger...
than required. As the example in the Introduction suggests, often a much weaker assumption will suffice to support this equilibrium. I include this assumption in order to clear the proof of additional elements that depend on the fine specifications of the game. It simplifies the exposition, highlights the general aspect, and helps in the later comparison with the informed sender benchmark in section 5.

4 The Role of $\phi$

I now discuss the role of the special state $\phi$ and develop a related result in a state space not containing $\phi$ for the discrete uniform quadratic case. The absence of $\phi$ might severely complicate the derivation of most informative equilibria. On the other hand, it is shown below that the effect of $\phi$ on the optimal equilibrium becomes insignificant if the state space is large relative to the divergence between the players’ interests.

Consider again the sender’s strategy in Proposition 2. As long as she has not learned $\theta$, the mere possibility of $\theta = \phi$ prevents her from inducing an “active” action ($a \neq \phi$). Assume now that $\Theta = \{1, 2, \ldots, n\}$ and that the sender follows the same learning plan as in Proposition 2. If the divergence in interests is not too great, during the early periods the sender will not have an incentive to submit a “false positive” report. By doing so she will induce a high action and benefit in the case where $\theta$ is sufficiently high (below the receiver’s belief). On the other hand, if $\theta$ turns out to be very low, the induced action will be too high from the sender’s perspective. Thus, during the early stages of the game, the sender’s behavior as in Proposition 2 seems plausible in this environment as well. However, as time goes by, the risk of “far low states” is reduced. For example, if $\epsilon_k = 0$ for all $k > 1$, then rather than wait another period, the sender is better off submitting a false positive report that corresponds to “$\epsilon_2 = 1$.”

This suggests that if $\Theta = \{1, \ldots, n\}$ one can proceed with the logic of the equilibrium as in Proposition 2, but not all the way down the state space. Instead, at some point, the players must engage in another reporting phase, which will not be fully revealing. In Proposition 3, I prove that for the discrete uniform quadratic case, there exists a threshold that depends on $b$ (and does not depend on $n$), such that, in equilibrium, the receiver can discover every state above this threshold.

A discrete uniform quadratic case: $\Theta = \{1, \ldots, n\}$, $p(k) = \frac{1}{n}$ for each $k \in \Theta$, the players’ preferences are represented by $u^R(a, \theta) = -(\theta - a)^2$, and $u^S(a, \theta, b) = -(\theta + b - a)^2$, where $b \in \mathbb{N}$.

The set of available experiments is $\{\epsilon_k | k \in \Theta\}$. As before, the sender is allowed to submit only one report but she can choose the timing of the report’s release.

**Proposition 3**  There exists an equilibrium that attains complete separation of every state $\theta \geq 10b$.

**Proof**  See Appendix. QED
The proof of this proposition is more involved than the proof of Proposition 2. An important feature in the proof of Proposition 2 is that, due to $\phi$, we can limit attention to strategies in which a nonempty report is submitted only if the sender knows $\theta$. In the proof of Proposition 3, however, we also have to consider strategies in which both the learning plan is changed and a false positive report is submitted. Moreover, an important assumption in Proposition 2 is that $\epsilon_\phi$ is not part of the sender’s arsenal. In contrast, now the sender can directly explore any of the states.

The sender knows that she has a given amount of time before choosing whether to induce each action. She might find it optimal to invest this time in exploring some “low” states and upon ruling them out to induce a relatively high action. Dealing with these kinds of deviations is central in the proof of Proposition 3 (see step 3 in the proof). The threshold selected in this proposition ensures that it will take the sender too long (and hence be too costly) to rule out a sufficient amount of low states.

Proposition 3 does not present the optimal equilibrium. Sometimes the players can benefit if they choose a more complicated equilibrium below the threshold (for example, playing some CS equilibrium in this region). The main purpose of this Proposition is to illustrate how the logic of the equilibrium in Proposition 2 can be applied in a more standard state space. Essentially, by giving up communication about states below a certain threshold, we create an artificial substitute for $\phi$. In the suggested equilibrium, we again see a similar form of strategic pressure as high actions can be induced only during the early stages of the game. Importantly, the threshold in Proposition 3 is independent of $n$. In this sense, the effect of $\phi$ on the optimal equilibrium becomes negligible as the state space becomes large.

I conclude this section with a minor comment on the threshold that is selected in Proposition 3. For the construction of an equilibrium as in the proof, the suggested threshold is pretty tight: for a threshold of the form $c \cdot b$, selecting $c \leq 8$ will not suffice for any $b \in \mathbb{N}$, and, for sufficiently small values of $b$, $c = 9$ is also insufficient.

5 The Informed Sender Benchmark

At this point, it is natural to compare the results in Propositions 2 and 3 with equilibria à la Crawford and Sobel (CS equilibria), namely, the completely informed sender benchmark. We begin with a general observation regarding the structure of Pareto efficient equilibria when $\phi$ is an element of the state space and the sender is completely informed. Then we focus on the optimal CS equilibrium in the discrete uniform quadratic case (with and without $\phi$).

Observation 2 In any (ex-ante) Pareto efficient equilibrium, $\{\phi\}$ is a completely isolated element in the receiver’s information structure.

---

4In the uniform quadratic case, both players rank different equilibria identically. Therefore, there exists an unambiguously ex-ante most-desirable equilibrium, which is termed the optimal equilibrium.
Proof  Note that the receiver’s information partition $\Theta_\phi = \{\{\phi\}, \Theta - \{\phi\}\}$ is consistent with equilibrium. Denote this equilibrium by $e_\phi$. Recall that whenever $a = \phi \oplus \theta = \phi$, $U^i(a, \theta) = -\infty$. Thus, the equilibrium $e_\phi$ is mutually preferred by both players over any equilibrium in which $\phi$ is not isolated. QED

It is easy to see that any equilibrium on the state space $\{1, 2, ..., n\}$ corresponds to an equilibrium on $\{\phi, 1, 2, ..., n\}$ in which $\phi$ is completely isolated (and vice versa). Thus, it is sufficient to determine the receiver’s information structure that is obtained in the optimal CS equilibrium on $\Theta = \{1, 2, ..., n\}$.

Essentially, CS equilibria in a discrete state space are similar to the continuous case. However, now the sender’s types have a strictly positive measure. Thus, in general, limiting attention to partitional equilibria is with loss of generality. Yet, in a separate technical report it is shown that in the discrete uniform quadratic case, the optimal equilibrium is always partitional and it can be derived according to the following procedure.$^5$

Procedure 1 Define $x_1 = 1$ and $x_{j+1} = x_j + 4b - 2$ for $j \in \mathbb{N}$. Next, find the index $i^*$ that

$$\sum_{i=1}^{i^*} x_i \leq n < \sum_{i=1}^{i^*+1} x_i$$

and denote the residual by

$$y = n - \sum_{i=1}^{i^*} x_i.$$

Let $\{y_i\}_{i=1}^{i^*}$ be the maximal (in a lexicographic sense) weakly increasing sequence where $y_i \in \{0\} \cup \mathbb{N}$ and $\sum_{i=1}^{i^*} y_i = y$.

The optimal CS equilibrium partitions the states $\{1, 2, ..., n\}$ into “intervals” of magnitude $x_i + y_i$.

For example, let $\Theta = \{\phi, 1, 2, ..., 50\}$ and $b = 2$. Then, in the optimal equilibrium $\phi$ is isolated. The first elements of the $x$-sequence are $x_1 = 1$, $x_2 = 7$, $x_3 = 13$ $x_4 = 19$, $x_5 = 25$, accordingly $i^* = 4$ and $y = 10$. The lexicographically maximal weakly increasing sequence $\{y_i\}_{i=1}^{4}$ is then given by $\{2, 2, 3, 3\}$.

| $\phi$ | 1,...,3 | 4,...,12 | 13,...,28 | 29,...,50 |

A standard feature of this equilibrium is that higher “intervals” contain more elements. Namely, when $b$ is fixed, as $n$ increases the receiver obtains less precise information regarding the high states of the world. This is, of course, not the case in Propositions 2 and 3. Recall that under the assumptions of Proposition 2 (which are clearly satisfied here) a fully informative equilibrium exists. If $\phi$ is not part of the state space, then, according to Proposition 3, a complete separation of states $\theta \geq 20$ is attainable in equilibrium.

$^5$Details are given in Frug (2014b).
6 Bias towards Extreme Actions

This section illustrates by way of example how intuitions of Proposition 1 (exploration direction and strategic pressure) can be applied to a situation in which the sender’s bias is not of constant direction.

Let $\Theta = \{-2, -1, 0, 1, 2\}$ with $p(\theta) = \frac{1}{5}$ for each $\theta \in \Theta$. Let $u^R(a, \theta) = -(\theta - a)^2$ and $u^S(a, \theta) = -(\theta + \text{sign}(\theta) - a)^2$ be the players’ utility functions. These functions capture the idea that, relative to the receiver, the sender is biased towards the extreme actions. In particular, within each region (positive/negative), the sender has a constant bias of 1.

It is easy to see that if the sender is completely informed, under the optimal equilibrium the receiver learns whether the state is negative, positive, or 0 (the unique point of agreement).

Now we incorporate gradual learning. Assume that for each $k \in \Theta$, there is an experiment $\epsilon_k$ which reveals to the sender whether $\theta = k$. As before, each experiment takes one unit of time, the sender can choose when to submit a report, and at most one report is allowed.

Observe that, similar to the logic in Proposition 1, if there exists a separating equilibrium, it must be the case that $\epsilon_2$ is performed before $\epsilon_1$ and that $\epsilon_{-2}$ is performed before $\epsilon_{-1}$. I now show that such an equilibrium exists.

Claim 1 A fully informative equilibrium exists.

Proof Consider the following sender’s strategy $s^* = (f, m^f)$, where

$$f(t) = (-1)^t \left\lfloor 3 - \frac{t}{2} \right\rfloor = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ -2 & 2 & -1 & 1 & 0 \end{pmatrix}$$

and $m^f_t \neq \sigma$ if and only if $\epsilon_{f(t)} = 1$. The observable part of this strategy has full support. Thus, the receiver’s beliefs are derived from the Bayes’ rule and the receiver’s best reply is

$$a^*(m) = \begin{cases} (-1)^t \left\lfloor 3 - \frac{t}{2} \right\rfloor & m_t \neq \sigma, t \leq 4 \\ 0 & m_t = \sigma, \forall t \leq 4. \end{cases}$$

We now show that $s^*$ is the sender’s best response to $a^*(m)$. Denote by $d_s(\theta)$ the distance between the sender’s most-desirable action and the action induced by the strategy $s$ at state $\theta$. The sender’s expected utility if she plays according to $s$ is simply $E[u^S|s] = -\frac{1}{5} \sum_{\theta \in \Theta} (d_s(\theta))^2$. Under $s^*$, whenever $\theta \neq 0$ the sender experiences a loss of 1, and thus $E[u^S|s^*] = -\frac{4}{5}$.

Clearly, $m^f$ is optimal given $f$. Note that if $s' = (g, m^g)$ is a profitable deviation, it must be the case that the sender is better off at $\theta = 1$ or at $\theta = -1$. We now show that such an $s'$ does not exist.

First, assume that the sender is better off at $\theta = -1$. In this case, under $s'$, the sender induces the action $a = -2$ at the state $\theta = -1$. If $g(1) = -1$, and $m_1 \neq \sigma$ if and only if $\epsilon_{-1} = 1$, the action $a = -2$
is not induced at the state $\theta = -2$. In this case $d_\theta(-2) \geq 2$, and $E[u^S|s^\prime] \leq -\frac{1}{5}((d_\theta(-2))^2 + (d_\theta(2))^2) \leq -1 < E[u^S|s^*]$. Otherwise, it must be the case that according to the sender’s strategy, $\epsilon_{g(1)} = 0 \Rightarrow m_1 \neq \sigma$. In this case, there are at least four states in which the action $a = -2$ is induced. Thus, $d_\theta(\eta) \geq 2$ for some $\eta > 0$ and $E[u^S|s^\prime] \leq -\frac{1}{5}(d_\theta(\eta))^2 < -1 < E[u^S|s^*]$. 

Now, assume that the sender is better off at $\theta = 1$. In this case, under $s^\prime$, the sender induces the action $a = 2$ at the state $\theta = 1$. If $1 \in \{g(1), g(2)\}$, and $m_2 \neq \sigma$ if and only if $\epsilon_1 = 1$, then either $2 \notin \{g(1), g(2)\}$ or $-2 \notin \{g(1), g(2)\}$. Consequently, $d_\theta(\eta) \geq 2$ for some $\eta \in \{-2, 2\}$, and so $E[u^S|s^\prime] \leq -\frac{1}{5}((d_\theta(-2))^2 + (d_\theta(2))^2) \leq -1 < E[u^S|s^*]$. Otherwise, it must be the case that according to the sender’s strategy, $\epsilon_{g(1)} = 0 \land \epsilon_{g(2)} = 0 \Rightarrow m_2 \neq \sigma$. In this case, the action $a = 2$ is induced in at least three different states, and so there is $\eta \leq 0$ with $d_\theta(\eta) \geq 2$, and therefore $E[u^S|s^\prime] \leq -\frac{1}{5}((d_\theta(\eta))^2 + (d_\theta(2))^2) \leq -1 < E[u^S|s^*]$. QED

Again, compared to the completely informed sender benchmark, the players manage to attain a mutually preferred equilibrium in which the sender is put under strategic pressure. Certain actions can be induced only during the early stages of the game.

In this section, the logic of Proposition 1 is applied to positive and negative regions separately. While it may seem tempting to consider another equilibrium candidate in which, roughly speaking, the whole problem is separated into two sub-problems where the sender explores, say, the positive region first and afterwards turns to the negative region, it turns out that such an intuition is inconsistent with equilibrium. This is because, contrary to the example in the Introduction, now the sender can directly explore any of the states ($\epsilon_0$ is part of the sender’s arsenal). To achieve complete separation in equilibrium, it is important that states in a given region (negative/positive) not be completely ruled out as long as there is still more than one state to be examined within the opposite region.

7 Alternative Learning Technologies

To demonstrate how the gradualness of the sender’s learning can enhance informativeness in communication we assumed a particular learning technology. Refer to the sender’s learning in the previous sections as learning by identification because at each stage only one element can be inspected. Clearly, the result of having a fully revealing equilibrium in Proposition 2 is an extreme case, which relies on this particular learning technology. In this section we address other learning technologies and elaborate on the intuition of strategic pressure.

We begin with introducing a slightly more general model which accounts for a more general sender’s experimentation. This model will serve as a unified framework for what has been presented above and all the following examples.

An experiment $Q = \{Q(\theta) : \theta \in \Theta\}$ is a partition of $\Theta$ with the standard interpretation that upon performing $Q$, the sender learns that $\theta \in Q(\theta)$. A learning technology is characterized by the set of

---

6A more general definition could be that an experiment $Q = \{q(\theta)|\theta \in \Theta\}$ is a family of distributions over $\Theta$ in the sense
available experiments. This set, denoted by $EX = \{Q_i | i \in I\}$, where $I$ is a (finite) set of indices, is commonly known. Each experiment takes one unit of time and it is impossible to conduct more than one experiment at a time. The experiments performed and their outcomes are unobservable and nonverifiable.

At the end of each period, the sender submits a cheap-talk report regarding the progress in exploration. As in previous sections, often it is convenient and without a significant loss of generality to impose certain restrictions on the reporting protocol, e.g., to allow for only one report. Action $a \in \mathbb{R}$ must be selected by the receiver at time $T > |\Theta|$. This captures the idea that, in principle, the time available for exploration places no restrictions on the expert’s ability to explore the state.

Clearly, the completely uninformative equilibrium always exists. If we assume that the set $EX$ is rich in the sense that upon performing all of the experiments in $EX$ the sender becomes completely informed, it is immediate that we can always mimic the benchmark of the informed sender. We now turn to analyze two examples of alternative learning technologies that may be natural under some circumstances, and then we conclude this section with several general observations.

### 7.1 Criteria Learning

The total value may be determined by several different criteria of varying importance. Exploring these criteria separately one after another is a natural way of learning in this environment. Regardless of the order in which the criteria are explored, the information is acquired gradually. Consider the following example in which state $\theta$ is determined by the values of two different variables.

Let $\zeta, \eta \sim U\{0, \ldots, k\}$ and let $\theta = (k + 1)\zeta + \eta$. To verify that $\theta \sim U\{0, k^2 + 2k\}$, write every $x \in \{0, \ldots, k^2 + 2k\}$ in terms of its unique representation $x = (k + 1)x_1 + x_2$, where $x_i \in \{0, \ldots, k\}$. Then, $Pr(\theta \leq x) = Pr(\zeta < x_1) + Pr(\zeta = x_1) \cdot Pr(\eta \leq x_2) = \frac{x_1}{k+1} + \frac{1}{k+1} \cdot \frac{x_2+1}{k+1} = \frac{1}{(k+1)^2} \cdot (x + 1)$. Again, assume that the players minimize the (expected) quadratic loss and the sender’s bias is denoted by $b \in \mathbb{N}$.

**Claim 2**  Let $b = 1$. The optimal equilibrium under the completely informed sender benchmark splits $\Theta$ into $k + 1$ “intervals” of different lengths.

**Proof**  To see this, observe that for each $I$, we have $\sum_{i=1}^{I} (1 + 2(i - 1)) = I^2$. According to procedure 1, the length of the $i$-th interval is exactly $1 + 2(i - 1)$. The claim follows as $|\Theta| = (k + 1)^2$. QED

**Comment**  Clearly, for every $b > 1$, every equilibrium splits $\Theta$ into a smaller number of intervals.

We now introduce strategic gradual learning. Assume that two experiments are available,

$$EX = \{\text{check}_{\zeta}, \text{check}_{\eta}\},$$

such that each experiment returns the realized value of the corresponding random variable.

that, if the true state is $\theta$, then upon performing $Q$ the sender believes the the state of the world is $\eta$ with probability $q(\theta)(\eta)$. 

14
If a decision maker has to explore the state himself and he has only limited time to explore, it is natural that he would focus on the most important criterion. On the other hand, if a decision maker has sufficient time to check all the attributes before taking an action, the ordering of the experiments would not matter. In our setup, the decision maker gathers information via reports that are submitted by a biased expert. In this case, even in the absence of direct time constraints, optimal equilibrium again involves indirect pressure that forces the expert to submit an early report. Contrary to *learning by identification*, if the bias is sufficiently low, additional reporting can be useful.

**Claim 3** Let $b \leq \lceil \frac{k}{2} \rceil$. There exists an equilibrium that is strictly preferred to the optimal equilibrium under the completely informed sender benchmark.

**Proof** We construct an equilibrium. Consider the following strategy profile: the sender performs $\text{check}_\zeta$ (explores the most important criterion) and submits a “separating” report at $t = 1$, denoted by $m^*(\zeta)$. Regardless of the report at $t = 1$, no valuable information is transmitted at $t = 2$ (the sender “babbles”). The receiver plays his best response (denoted by $a^*(m)$). The choice of $b \leq \lceil \frac{k}{2} \rceil$ ensures that, provided that the sender’s information structure corresponds to knowing only the value of $\zeta$, she is willing to reveal the truth. To show that this is indeed an equilibrium we now show that the sender cannot benefit from performing $\text{check}_\eta$ instead of $\text{check}_\zeta$. Let $a^S(\eta)$ denote the most-desirable action, from the sender’s perspective, given a realization of $\eta$ (while $\zeta$ is unknown). Notice that

$$a^S(\eta) = \arg\max_a E[u^S(a, \theta) | \theta \in \{(k + 1)\zeta + \eta\}] = \frac{\eta + \lceil k(k + 1) + \eta \rceil}{2} + b = \frac{k(k + 1)}{2} + \eta + b.$$

Substitute and rearrange terms to get

$$E[u^S(a^*(m(\eta)), \theta)] \leq -\frac{1}{(k + 1)^2} \sum_{\eta=0}^{k} \sum_{\zeta=0}^{k} (\zeta \cdot (k + 1) + \eta + b - a^S(\eta))^2 = -(k + 1) \sum_{\zeta=0}^{k} (\zeta - \frac{k}{2})^2 =$$

$$= -\frac{k(k + 2)(k + 1)^2}{12} \leq -(\frac{(k + 1)^2}{12} + 1) < -\frac{k^2 + 2k + 12}{12} = E[u^S(a^*(m^*(\zeta)), \theta)].$$

The RHS is the sender’s expected utility if she follows the equilibrium strategy. The LHS is her utility if she deviates and privately explores $\eta$ instead of $\zeta$. The inequality denoted by (*) holds for every integer $k \geq 2$. It is straightforward to verify that for $k = 1$, the suggested profile of strategies also constitutes an equilibrium. In the latter case, however, it is a weak equilibrium.

We established a possible equilibrium in the environment of strategic gradual learning. To complete the proof we note that the suggested equilibrium splits $\Theta$ into $k + 1$ intervals of the same length. Thus, the variance of this partition is necessarily lower than the variance of the unbalanced partition that is obtained under the optimal equilibrium in the completely informed sender benchmark (see claim 2). QED
Note that claim 3 does not characterize the optimal equilibrium under strategic learning. Contrary to previous sections, here the players can benefit from allowing for more than one round of meaningful communication. It is easy to see that if $4b \leq k + 1$ after $\zeta$ is revealed at $t = 1$, the players can have another report regarding the residual state space in which they play the standard CS equilibrium. For example, if $k = 3$ and $b = 1$, the resulting receiver’s information partition under the optimal equilibrium is \{\{0\},\{1..3\},\{4\},\{5..7\},\{8\},\{9..11\},\{12\},\{13..15\}\}.

Comment In effect, I assumed two criteria such that $\zeta$ is lexicographically more important than $\eta$. It was convenient in the proof that the resulting distribution of $\theta$ be uniform so that we could directly apply procedure 1. This is obtained by setting the $\zeta$’s coefficient to be equal to $k + 1$. It is easy to see that for any coefficient greater than $k + 1$, the suggested play remains an equilibrium and whenever the coefficient is not too large, gradualness of learning (strictly) helps.

7.2 Cutoff Learning

The current learning technology, contrary to learning by identification, admits the extreme opposite result. It is shown below that if the sender can at each period select an element $x \in \Theta$ and privately learn whether $\theta \geq x$, gradual in learning does not increase the amount of information that can be transmitted in equilibrium, relative to the informed sender benchmark.

A “cutoff $k$” experiment, $c_k = 1_{\theta \geq k}$, reveals to the sender whether $\theta \geq k$. The set of available experiments in the cutoff learning regime is given by the collection

$$EX = \{c_k : k \in \Theta\}.$$

In each period the sender privately chooses a cutoff $k_t$, which may depend on the results of previous experiments. Thus, a sender’s strategic learning is a plan $(k_t(h_t))_{t \in T}$, where $h_t = (c_{k_s})_{s < t}$. Given any history, the experiment $c_{\min \Theta}$ is uninformative from the sender’s perspective and so it can be interpreted as “doing nothing.” In this sense, our definition of a learning plan allows the sender to avoid acquiring information.

The following result is given under fairly general assumptions. Let $\Theta$ be a linearly ordered discrete state space (for a state space like that of Proposition 2, consider $\phi < \theta$ for every $\theta \neq \phi$). Assume that for each $\theta$, $u^R(\cdot, \theta)$ is concave and that for each $\theta$, $u^R_1(\cdot, \theta) = 0$ for some $a$. This ensures that, given any information, there is a unique optimal action for the receiver. Also assume that $u^i$ is supermodular.

Proposition 4 Under cutoff learning regime, every equilibrium is equivalent (in terms of the transmitted information) to some CS equilibrium.

Proof See Appendix.
We conclude this section with several general observations about the relation between the learning technology and equilibrium informativeness. Observation 3 shows that even in the enriched environment of Proposition 2, full separation is indeed rare. It is shown that learning by identification is “almost” the unique set of experiments that allows for complete separation. One particular implication of this observation is that adding an experiment to a rich collection of experiments can make the players worse off. On the other hand, in Observation 4 it is shown that adding experiments to rich collections of experiments can also make the players better off. Observation 5 shows that giving up experiments that are unused in the optimal equilibrium can also make the players better off.

**Observation 3** Consider the environment of Proposition 2 ($\Theta = \{\phi, 1, \ldots, n\}$ and $EX = \{\epsilon_k | k \in \Theta - \{\phi\}\}$) and an experiment $Q \notin EX$. If there exist $A \subset Q$ such that $\phi \notin \cup A$ and $\min A + 1 \in \cup A$, adding the experiment $Q$ to the set $EX$ will make the players worse off.

**Proof** Similarly to Proposition 1, for each $\theta \in \Theta$ denote by $t(\theta)$ the last period in which the information that the state of the world equals $\theta$ is transmitted to the receiver in a separating equilibrium $e$. It must be the case that, according to the sender’s equilibrium learning plan, by the period $t(k)$ the sender cannot find out whether $\theta = k - 1$ for any $k > 1$. Thus, it follows that $t(n) < t(n - 1) < \ldots < t(2) < t(1)$. The sender is better off performing $Q$ at $t(\min A + 1)$ and inducing $\arg\max_a u_R(a, \min A + 1)$ at $\theta = \min A$. QED

This implies that adding an experiment $Q$ is quite likely to make the players worse off. For example, if $Q$ contains a nontrivial convex element, enriching $EX$ with $Q$ makes the players worse off in equilibrium. An example of an experiment $Q$ such that the set of experiments $\{Q, \epsilon_k | k \in \Theta - \{\phi\}\}$ allows for a separating equilibrium under the uniform quadratic case with $b = 1$ is $Q = \{\{\phi, 2\}, \{\phi, 2\}^c\}$.

**Observation 4** Adding experiments to rich collections of experiments can make the players better off.

**Example** Consider the uniform quadratic case with $\Theta = \{1, \ldots, 10\}$ and $b = 1$. By Proposition 4 and procedure 1, the optimal equilibrium under the cutoff learning regime induces the partition $\{\{1\}, \{2, 3, 4\}, \{5, 6\}, \{7, 8\}, \{9, 10\}\}$ for the receiver. Let $Q = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8\}, \{9, 10\}\}$. If $Q$ is added to the set of available experiments the following constitutes an equilibrium. The sender performs $Q$ at $t = 1$ and immediately reports truthfully the outcome of this experiment (all subsequent reports are completely uninformative). Provided that the sender’s information structure is given by $Q$, it is immediate that truth-telling is incentive compatible and a direct calculation shows that the sender’s expected payoff in this equilibrium is $-\frac{5}{4}$. To see that the sender does not have an incentive to deviate and perform another experiment, note that every cutoff experiment partitions $\Theta$ in two. This implies that performing such an
experiment at \( t = 1 \) will induce at most two receiver’s actions. Thus, there are at least four states with \( d_{s'}(\theta) \geq 2 \) where \( s' \) denotes a sender’s strategy by which she performs any experiment other than \( Q \) at \( t = 1 \).\(^7\) Therefore, \( E[u^S(a^*, \theta)|c_j] < -\frac{1}{10} \cdot 4 \cdot 2^2 < -\frac{5}{4} \).

**Observation 5** Dispensing with experiments that are unused in the optimal equilibrium can make the players better off.

**Example** Consider the uniform quadratic case with \( \Theta = \{1, \ldots, 5\} \) and \( b = 1 \). By Proposition 4 and procedure 1, the optimal equilibrium under the cutoff learning regime induces the partition \( \{\{1\}, \{2, \ldots, 5\}\} \) for the receiver. Moreover, this can be attained by a report at \( t = 1 \) after the sender performs only \( c_2 \). The set of experiments \( EX' = \{c_2, c_5\} \) is obtained by dispensing with some unused experiments. The following sender’s strategy is consistent with equilibrium. Perform \( c_5 \) at \( t = 1 \) and report truthfully. If \( \theta < 5 \), perform \( c_2 \) at \( t = 2 \) and report truthfully. The partition obtained in this equilibrium refines the receiver’s information structure under the optimal equilibrium with \( EX = \{c_k : k \in \Theta\} \). Thus, both players are now better off.

8 Conclusion

In this paper, I showed how the players can benefit from the gradualness of an expert’s learning when it is strategic and unobservable. The main intuition is that of “strategic pressure.” A common feature of desirable equilibria is that certain actions can be induced only during the early stages of the game. Since learning takes time, the expert’s information structure in the early stages is still not refined, regardless of her particular learning strategy. Knowing that, the expert strategically chooses which experiments to perform first. This provides the decision maker with an idea about the expert’s information structure and how it develops over time. The set of inducible actions shrinks sufficiently fast with time, in sync with the expert’s physical ability to acquire information.

The main ingredients of the environment in this paper were (1) costless experimentation and (2) the time available for learning is sufficiently long so that it is possible to perform all of the available experiments. If the expert is expected to submit a report at the end of her learning (as she becomes completely informed), the problem of experiment scheduling is completely vacuous. On the other hand, the artificial pressure, presented in most of the results in this paper, raised an important issue of strategic scheduling of experiments for the expert. The only motivation for experiment scheduling in our environment is strategic. This may suggest a broader question of strategic scheduling. One may ask, given a general collection of available experiments, how should the (biased) expert schedule these experiments? This question can be addressed both in the environment of observable experiments (unobservable outcomes) and unobservable experimentation.

\(^7\)Recall that \( d_{s'}(\theta) \) denotes the distance between the sender’s most-desirable action and the action induced by the strategy \( s' \) at state \( \theta \).
9 Appendix

Proposition 3 There exists an equilibrium that attains complete separation of every state \( \theta \geq 10b \).

Proof Consider the following profile of strategies: sender’s strategy \( s^* = (f, m^f) \): \( f(t) = n + 1 - t \) for all \( t \in \{1,..,n\} \), \( m^f \neq \phi \) iff \( \epsilon_{f(t)} = 1 \) and \( \tau \in \{1,..,n + 1 - 10b\} \); receiver’s strategy: \( a^*(\{m_t\}) = n + 1 - \tau \) if \( m_\tau \neq \sigma \) for some \( \tau \in \{1,..,n + 1 - 10b\} \) and \( a(\{m_t\}) = 5b \) if \( m_t = \sigma \) for all \( t \in \{1,..,n + 1 - 10b\} \).

First, it is clear that the receiver’s strategy is a best response to the sender’s strategy. We now prove, in three steps, that \( s^* \) is a sender’s best response to the receiver’s strategy, and thus the suggested profile constitutes an equilibrium.

Step 1: Optimality of \( m^f \) given the learning plan \( f \).  The sender cannot benefit from postponing a report once \( \theta \) is revealed because this will induce a lower receiver’s action, which is worse for the sender. Also, the sender cannot benefit from a “false positive” report \( m_{t(\bar{a})} \neq \sigma \) given that \( \epsilon_{f(\tau)} = 0 \) for all \( \tau \leq t(\bar{a}) \), for some \( \bar{a} \geq 10b \). It is shown below that the sender’s expected utility from a “false positive” report at \( t(\bar{a}) \) (LHS) is lower than her expected utility from continuing to follow the strategy \( s^* \) (RHS):

\[
-\frac{1}{\bar{a} - 1} \sum_{\theta=1}^{\bar{a}-1} (\theta + b - \bar{a})^2 < -\frac{1}{\bar{a} - 1} \left( \sum_{\theta=1}^{10b - 1} (b)^2 + \sum_{\theta=\bar{a} - 10b + 1}^{\bar{a}-1} (\theta + b - \bar{a})^2 \right) = \\
= -\frac{1}{\bar{a} - 1} \left( \sum_{\theta=1}^{10b-1} (\theta + b - 10b)^2 + (\bar{a} - 10b)b^2 \right) < -\frac{1}{\bar{a} - 1} \left( \sum_{\theta=1}^{10b-1} (\theta + b - 5b)^2 + (\bar{a} - 10b)b^2 \right).
\]

It is left to show that there does not exist an alternative learning plan \( g : \{1,..n\} \rightarrow \Theta \) and a \( g \)-measurable reporting policy \( m^g \) such that \( s = (g, m^g) \) is a profitable deviation for the sender.

Let \( K = \{(g, m^g) : \forall a \geq 10b, m_{t(a)} \neq \sigma \implies \exists t \leq t(a), \epsilon_{g(t)} = 1 \} \). Also, let us denote by \( a(\theta|s) \) the action induced at state \( \theta \) if the sender plays according to strategy \( s \).

Step 2: The strategy \( s^* = (f, m^f) \) is optimal in \( K \). Let \( s = (g, m^g) \in K \) such that \( g \neq f \). We show that \( s \) is suboptimal. Let \( k \) be the minimal integer with \( g(k) \neq f(k) \). Therefore, \( g(k) < f(k) \).

Let \( g' = (g(k), f(k)) \circ g \) be the learning plan in which the experiments \( \epsilon_{f(k)} \) and \( \epsilon_{g(k)} \) are swapped. If \( a(g(k)|s) \leq a(f(k)|s) \), \( g' \) is clearly preferred to \( g \), as the sender is better off in state \( f(k) \) (because a higher action can be induced) and her payoffs remain the same everywhere else. If \( a(g(k)|s) > a(f(k)|s) \), \( g' \) allows the sender to induce the higher action \( a(g(k)|s) \) in the higher state \( f(k) \) and the lower action \( a(f(k)|s) \) in the lower state \( g(k) \) without changing the state-to-action mapping at other states. Supermodularity of the sender’s preferences ensures that \( g' \) is preferred in this case as well. There exists an optimal strategy in \( K \) with a pure learning plan, and thus \( f \) is optimal in \( K \).

Step 3. To conclude we show that \( f \) is better than every strategy outside \( K \). These strategies have the property that there is \( a \geq 10b \), such that after performing \( t(a) \) “unsuccessful” experiments, the sender

\(^8\) \( a(\theta|s) \) is well defined whenever the randomness in the sender’s strategy is confined to the selection among the nonempty messages.
induces the action \( a (m_{t(a)} \neq \sigma) \). We now add some formal notation.

Let \( N_t = \{(g, m^g) : \forall \theta < \tilde{t} \leq t(10b), \epsilon_{g(t)} = 0 \implies m_t \neq \sigma \} \) and \( N = \bigcup_{t \leq t(10b)} N_t \). Also, in the following steps, for a strategy \( s = (g, m^g) \) to be considered, we denote \( \zeta = \max_\Theta - \{g(t)|t \leq \bar{t}\} \) and \( \eta = \min_\Theta - \{g(t)|t \leq \bar{t}\} \).

Step 3.1: An optimal strategy in \( N_t \) satisfies \( \eta \leq \theta < \zeta \implies \theta \notin \Theta - \{g(t) : t \leq \bar{t}\} \). Let \( s = (g, m^g) \in N_t \). Assume that there exists \( \theta \in \{g(t)|t \leq \bar{t}\} \) such that \( \eta < \theta < \zeta \). Recall that \( \theta \notin \Theta - \{g(t)|t \leq \bar{t}\} \) \( \implies a(\omega|s) = a(\bar{t}) = n - \bar{t} + 1 \).

If \( a(\theta|s) > n - \bar{t} + 1 \), by supermodularity, the sender is better off performing \( \epsilon_\zeta \) instead of \( \epsilon_\theta \) and inducing the action \( a(\theta|s) \) in state \( \zeta \) (and action \( n - \bar{t} + 1 \) at \( \theta \)).

If \( a(\theta|s) < n - \bar{t} + 1 \), by supermodularity, the sender is better off performing \( \epsilon_\eta \) instead of \( \epsilon_\theta \) and inducing the action \( a(\theta|s) \) in state \( \eta \) (and action \( n - \bar{t} + 1 \) at \( \theta \)).

If \( a(\theta|s) = n - \bar{t} + 1 \), the sender gains nothing by performing \( \epsilon_\theta \). If, at \( \eta \), the sender prefers one of the inducible actions that are below \( n - \bar{t} + 1 \), she can be better off if instead of \( \epsilon_\theta \) she performs \( \epsilon_\eta \). Alternatively, \( u^S(n - \bar{t} + 1, \zeta) < u^S(n - \bar{t} + 2, \zeta) \) and \( \eta > 1 \). Moreover, it must be the case that either \( g^{-1}(\eta - 1) < \bar{t} \) or \( g^{-1}(\theta) < \bar{t} \). Then, the sender can be better off if \( \epsilon_\zeta \) is performed instead of \( \epsilon_{\eta - 1} \) or \( \epsilon_\eta \).

Step 3.2: Let \( s = (g, m^g) \) be an optimal strategy in \( N_t \); then \( g^{-1}(\theta_0) > g^{-1}(\theta_1) > g^{-1}(\theta_2) \) for \( \theta_0 < \eta \) and \( \zeta < \theta_1 < \theta_2 \). Again, this is a direct consequence of the supermodularity of the sender’s preferences. Let \( \hat{\theta} \) be the highest state that is examined after some \( \theta < \hat{\theta} \). The sender is better off changing the order of these two experiments.

Step 3.3: Let \( s = (g, m^g) \) be an optimal strategy in \( N_{t(\hat{a})} \); then \( \eta \leq \hat{a} - 3b \). Assume by contradiction that \( \hat{a} - 3b \in \{g(t) : t \leq t(\hat{a})\} \). From 3.1 and the fact that \( |\Theta - \{g(t) : t \leq t(\hat{a})\}| \geq \hat{a} - 1 \), it follows that \( \zeta > \hat{a} + 6b \). Hence, \( d_s(\zeta) \geq 7b \). If the sender performs \( \epsilon_\zeta \) before any experiment \( \epsilon_\omega \) s.t. \( \omega < \eta \), she will be able to induce the action \( \zeta \) at state \( \zeta \). Thus, at this state a loss of \( (d_s(\zeta))^2 \) is replaced with a loss of \( b^2 \).

On the other hand, giving up \( \epsilon_{\hat{a} - 3b} \) increases the loss in this state by at most \( (2b)^2 \).

Step 3.4: The optimal sender’s strategy in \( N \) is an element of \( N_{t(10b)} \). Let \( \bar{t} < t(10b) \). Consider an optimal strategy in \( N_t \), and let \( a > 10b \) be such that \( t(a) = \bar{t} \). We show that there is a better strategy (from the sender’s perspective) in \( N_{t+1} \). From 3.2, first the sender explores all the states above \( \zeta \) (from top to bottom) and then turns to the exploration of states below \( \eta \). Action \( a \) is induced at each of the states \( \eta, \eta + 1, .., \zeta - 1, \zeta \). Consider the strategy \( s' = (g', m^g') \in N_{t+1} \), where \( s \) and \( s' \) differ only in that before exploring states below \( \eta \), the sender performs \( \epsilon_\zeta \). \( m_{t(\zeta)} \neq \sigma \) iff \( \epsilon_\zeta = 1 \), and the action \( a - 1 \) is induced at period \( \bar{t} + 1 \). To see that the sender prefers \( s' \) to \( s \), it is sufficient to note that the sender’s payoff at \( \zeta \) under \( s' \) is higher than her payoff at \( \eta \) under \( s \). This is illustrated in the figure below.

It is immediate that the sender obtains identical payoffs at all states above \( \zeta \) and below \( \eta \) under both strategies \( s \) and \( s' \). Also, it is easy to see that the sender’s payoff at \( \theta \in \{\eta + 1, .., \zeta\} \) under \( s \) is equal to her payoff at \( \theta - 1 \) under \( s' \). To conclude, note that the sender’s payoff at \( \zeta \), under \( s' \), is \(-b^2 \). From 3.3, \( \eta \leq a - 3b \), which means that the sender’s payoff at \( \eta \), under \( s \), is at most \(-(2b)^2 \).

Let \( s = (g, m^g) \) denote the optimal strategy in \( N_{t(10b)} \). We now conclude the proof by showing that the
sender prefers \( s^* \) to \( s \). Under the strategy \( s \), at least at \( 10b - 1 \) states, \( \theta_1 < \theta_2 < \ldots < \theta_{10b-1} \), the action \( a = 10b \) is induced. First, assume that \( \theta_1 \geq 3b \). The vector of smallest possible distances \( (d_s(\theta_j))_{j=1}^{10b-1} \) is given by

\[
(5b-1, 5b-2, \ldots, 1, 0, 1, \ldots, 5b-2, 5b-1).
\]

On the other hand, the distances \( (d_s^*(\theta_j))_{j=1}^{10b-1} \) are at most

\[
(b + 1, \ldots, 6b - 1, b, \ldots, b).
\]

Notice that for any \( \theta \in \{g(t) : t < t(10b)\} \) we have \( a(\theta|s) = a(\theta|s^*) \). Since we assume the uniform quadratic case, we can compare \( s \) and \( s^* \) by summing up the square distances on \( \{\theta_1, \ldots, \theta_{10b-1}\} \) and comparing the costs (negative utility). The following inequality holds\(^9\) for any \( b \in \mathbb{N} \):

\[
\sum_{j=1}^{10b-1} d_s(\theta_j) \geq 2 \sum_{k=1}^{5b-1} k^2 > (5b)b^2 + \sum_{k=1+b}^{6b-1} k^2 \geq \sum_{j=1}^{10b-1} d_s^*(\theta_j);
\]

therefore, \( s^* \) is better than \( s \). By replacing one of the states in \( \{\theta_1, \ldots, \theta_{10b-1}\} \) with a state \( \theta < 3b \) the LHS of the strict inequality increases by, at least, \( (\theta + b - 10b)^2 - (5b-1)^2 \) and the RHS increases by, at most, \( (\theta + b - 5b)^2 - (6b-1)^2 \). For any \( b \in \mathbb{N} \) and \( \theta < 3b \), we have \( (\theta + b - 10b)^2 - (5b-1)^2 > (\theta + b - 5b)^2 - (6b-1)^2 \).

Thus, the sender is better off under \( s^* \) rather than under \( s \) even if we dispense with the assumption that \( \theta_1 \geq 3b \). Therefore, \( s^* \) is better than \( s \), and the proposition follows. QED

**Proposition 4** Under the cutoff learning regime, every (nonredundant) equilibrium induces a receiver’s information partition that is consistent with some CS equilibrium.

**Proof** First, observe that the sender’s information structure at each time \( t \in T \) is monotonic: for \( J_t \subset \Theta \), which is an element of the sender’s information partition at time \( t \), if \( k_1, k_2 \in J_t \), then every \( k_1 \leq k \leq k_2 \) satisfies \( k \in J_t \). Otherwise there exist \( i < j \) such that \( c_i = 0 \) and \( c_j = 1 \), a contradiction.

\(^9\)By using the formula for the sum of squares of the first \( K \) naturals: \( \sum_{k=1}^{K} k^2 = \frac{K^3}{3} + \frac{K^2}{2} + \frac{K}{6} \)
We now show that every equilibrium is monotonic. Assume by contradiction that \( e \) is a nonmonotonic equilibrium. By the properties of \( u^R \) and as \( e \) is nonredundant, every element \( \mu \in \bar{\mu}_e \) induces a distinct, unique action. There exists a maximal element \( \hat{\theta} \in \Theta \) such that, under \( e \), \( \bar{a} \) is induced with positive probability at \( \hat{\theta} \), \( a \) is induced with positive probability at \( \hat{\theta} + 1 \), and \( \bar{a} < \bar{a} \). In particular, it follows that \( c_{\hat{\theta}+1} \) is part of the sender’s exploration program and that it is performed (weakly) before the reports that induce the actions \( \bar{a} \) and \( a \) are expected at \( e \).

By supermodularity of \( u^S \), if \( u^S(a, \hat{\theta} + 1) \geq u^S(\bar{a}, \hat{\theta} + 1) \) then \( u^S(a, \hat{\theta}) > u^S(\bar{a}, \hat{\theta}) \). Thus, the sender is better off replacing \( c_{\min supp(\mu_2)} \) with \( c_{\hat{\theta}+1} \) because this allows her to join the state \( \hat{\theta} \) to the reporting program that induces \( a \) without affecting the state-to-action mapping anywhere else. On the other hand, assume that \( u^S(a, \hat{\theta} + 1) < u^S(\bar{a}, \hat{\theta} + 1) \). If \( \hat{\theta} + 1 = n \), the sender is better off simply changing the reporting strategy as the element \( \hat{\theta} + 1 \) is isolated by the time \( \bar{a} \) is induced. In the case \( \hat{\theta} + 1 < n \), the sender is better off replacing \( c_{\hat{\theta}+1} \) with \( c_{\hat{\theta}+2} \). Again, after each of these changes she can join the state \( \hat{\theta} + 1 \) to the reporting program that induces \( a \) without affecting the state-to-action mapping anywhere else. Thus, in the cutoff learning regime nonmonotonic equilibria are impossible.

Let \( e \) be an equilibrium. For \( \mu \in \bar{\mu}_e \) denote by \( a(\mu) \) the action selected by the receiver if he holds the belief \( \mu \). Assume by contradiction that the induced belief structure is inconsistent with a CS equilibrium. Then, there are two adjacent elements \( \mu_1, \mu_2 \in \bar{\mu}_e \) such that either \( u^S(a(\mu_1), \min supp(\mu_2)) > u^S(a(\mu_2), \min supp(\mu_2)) \) or \( u^S(a(\mu_2), \max supp(\mu_1)) > u^S(a(\mu_1), \max supp(\mu_1)) \). Both actions \( a(\mu_1) \) and \( a(\mu_2) \) are still inducible at the time in which \( c_{\min supp(\mu_2)} \) is performed. The inequalities above imply that the sender can benefit from replacing \( c_{\min supp(\mu_2)} \) with either \( c_{\min supp(\mu_2) - 1} \) or \( c_{\min supp(\mu_2) + 1} \), contradicting the assumption that \( e \) is an equilibrium. QED

References


