Brevity*

Péter Eső† and Dezső Szalay‡

October 2, 2020

Abstract

We study incentives for information acquisition in a game of strategic information transmission. We show that information that is sufficienty costly to acquire can only be transmitted coarsely. We provide comparatives statics results showing that a more nuanced categorization provides weaker incentives for information acquisition than a coarser one.

JEL: D82, D83

Keywords: strategic information transmission; cheap talk; moral hazard; information acquisition

*We are very grateful for excellent comments to Andreas Blume, Inga Deimen, Ian Jewitt, Meg Meyer, Elliot Lipnowski, and seminar participants at Universities of Arizona, Bonn, Oxford, and Warwick, EUI and at ESSET (Gerzensee). Szalay greatfully acknowledges financial support from the German Research Foundation (DFG) through the Cluster of Excellence Hausdorff Center for Mathematics (HCM). All errors are our own.

†Department of Economics, Oxford University; peter.eso@economics.ox.ac.uk.

‡Department of Economics, University of Bonn, and CEPR; szalay@uni-bonn.de.

"Brevity is the soul of wit."

William Shakespeare, *Hamlet* (1602)

1 Introduction

Succinct advice is necessarily simplified yet often perceived insightful and valuable. We provide an explanation: lack of detail in communication provides stronger incentives for the sender to become informed. In our model there is no conflict of interest nor complexity induced communication cost. Brevity, required or expected from the sender by the receiver, resolves moral hazard in information acquisition by imposing a loss on the sender who wanders off-equilibrium by not acquiring information.

Imagine a referee who shares the editor's objective in that they both want to publish good papers and reject bad ones. Assessing a paper takes the referee unobservable time and effort. If the referee is expected to offer their evaluation on a fine scale they might be tempted to shirk and submit a non-committal report. In contrast, when forced to recommend either acceptance or rejection, the referee may find it optimal to study the paper carefully, fearing the embarrassment of being exactly wrong. Unrefined, and hence lossy communication provides stronger incentives for information acquisition and may end up socially beneficial overall.

Formally, in the familiar model of cheap talk between a sender and a receiver, with an arbitrary log-concave state distribution and identical quadratic losses, we show that binary communication provides stronger incentives for the sender to acquire information than communication in four or even infinitely many categories. Moreover, we show that information acquisition followed by binary communication is equilibrium exactly when it is jointly beneficial for the two parties as compared to uninformed babbling.

A local monotonicity result holds for all logconcave densities: starting from binary communication, dividing good and bad in subcategories always decreases incentives for information acquisition. Extending this result to global monotonicity requires more structure. For a class of densities that includes the uniform and the Laplace at its extremes, we show that adding categories decreases incentives for information acquisition. More categories improve decision-making based on informed advice but also make it more tempting to shirk on information acquisition, and the latter effect dominates. The cardinality of the welfare optimal and incentive compatible categorization is non-increasing in the cost of acquiring information. Information that is truly hard to uncover can only be communicated in binary categories.

Binary categories are used, among other examples, in the criminal justice system. The requirement of returning an either guilty or innocent verdict arguably makes judges and juries think very hard about the case. Professional reference letters (often either over-the-top or just lukewarm) and grades also appear to convey less information than they could.[1] More than two but still few categories are used in investment advice (strong buy, buy, etc.), illustrating the trade-off between the desire of communicating more detail and providing the advisor with incentives for acquiring costly information.

We assume that the sender's information acquisition is all-or-nothing and that the categorization forms an equilibrium in the overall game as well as in the continuation game — there is no commitment to the transmission of information. We explore the robustness of our results by allowing for ex ante optimal categorizations, where the sender's covert activity is to pay attention to a signal that is designed by the receiver, and draw the connection from that analysis to recent developments at the frontier of the Bayesian persuasion literature. In all model variations incentives for information acquisition preclude that the sender can recommend actions close to the prior mean, which would allow them to shirk and disguise their ignorance. The inability to distinguish states that differ only marginally from the prior mean creates an endogenous commitment to nonmarginal revisions of actions.

We introduce the model in Section 2 and explain the object of analysis in Section 3. Sections 4 through 6 contain our main results, Section 7 explores an important extension. Section 8 discusses connections in the literature.

2 Model

The payoff-relevant state of nature is ω, drawn from $[-\overline{\omega},\overline{\omega}] \subseteq \mathbb{R}$ according to a log-concave density f that is symmetric about 0 and has finite variance σ^2. The support of the distribution may be equal to a strict subset of $\mathbb{R}$ or may be equal to $\mathbb{R}$; symmetry is assumed mainly for ease of exposition.

There are two players: the Sender (S, he) and the Receiver (R, she). The game is as follows. First, the Receiver chooses (the cardinality of) a message space M. Then, the Sender decides whether to covertly learn the state of nature, at cost $c > \sigma^2$. Then he sends message $m \in M$ to the Receiver. Finally the Receiver picks an action, $y \in \mathbb{R}$. The Sender's payoff is $U_S = -(y-\omega)^2 - ec$, where $e \in \{0,1\}$ is a pure choice that indicates whether or

[1] "The most frequently awarded grade in Harvard College is actually a straight A" said Harvard's Dean of Undergraduate Education in the Harvard Crimson on December 3, 2013.

not he has learned the state; the Receiver's payoff is $U_R = -(y - \omega)^2$.

This is an information transmission game without conflict of interest (bias) but with moral hazard in information acquisition. There is no direct cost of sending messages, so talk is cheap. The assumption that the Sender's cost of learning the state, c, exceeds the variance of the state distribution, σ^2, is important. As we show in the next section, this implies that voluntary information acquisition followed by truthful, full revelation cannot be part of any equilibrium of the game.

In the rest of the paper, we use the term "equilibrium" in reference to a perfect Bayesian equilibrium.[2] Since S and R have the same objective after the information acquisition stage, we select equilibria that are best for both parties in the communication continuation game. Such equilibria transmit as much information as possible with the given message space, and the cardinality of induced actions taken by R is at most equal to the cardinality of M. By allowing the Receiver to pick the message space ex ante we select receiver-optimal equilibria in the overall game. Therefore the assumption that R chooses the cardinality of M could be replaced by appropriately refining equilibria in an unrestricted message space.

Logconcavity of the density of the state distribution is a useful technical assumption that is common in information economics.[3] In our model it ensures that the partition of states conveyed by information transmission from the Sender to the Receiver is unique for a given number of partition elements. This is explored in the next section.

3 Precoded communication and informative equilibria

In an equilibrium where the Sender is expected to acquire information, if he indeed learns ω then he reports the message for which the Receiver's response is closest to it. Off the equilibrium path, if he were not informed, then he would report the message for which the response is closest to the prior mean. The Receiver, upon receiving the Sender's message, forms beliefs over the states of nature that are consistent according to Bayes rule with the prior and the Sender's equilibrium strategy. In the first period the Sender chooses whether or not to acquire information about ω depending on which action yields the greatest expected payoff to him.

In order to characterize all equilibria with information transmission, first, consider the

[2] See Fudenberg and Tirole (1991), Definition 8.1, pp. 325-326.

[3] Among others Laplace, normal and uniform distributions belong to this class. See An (1998) for a characterization and Bagnoli and Bergstrom (2005) for many useful results and applications.

continuation game in case information about the state has been acquired by the Sender. If R has chosen a message space that allows to communicate the state perfectly, then that is obviously best for both parties. If R has chosen a message space with a finite cardinality N, then complete information transmission is technically rendered infeasible. For every positive integer N there exists a unique interval partition $\mathcal{P}$ of the states, which is symmetric about the prior mean 0, such that in an equilibrium of the continuation game following $e = 1$ (information acquisition) the Sender reveals which partition element $P_i^N = \left[a_{i-1}^N, a_i^N\right]$ the state falls into. If N is even, then the marginal types on the upper (non-negative) half of the support of the distribution are

$$0 = a_0^N < a_1^N < \ldots < a_n^N = \overline{\omega},$$

where $n = N/2$. There is another set of cutoffs symmetrically on the lower half of the support as well. If N is odd, then let $n = (N+1)/2$ and delete a_0^N, and the remainder is notationally the same as above.

The Receiver's best reply to any message sent by types in P_i^N coincides with the mean state in partition element P_i^N, that is (again, by symmetry, restricting attention to the upper half of the support),

$$y_i^N = \mu_i^N \equiv \mathbb{E}\left[\omega \middle| \, a_{i-1}^N \leq \omega \leq a_i^N\right]. \tag{1}$$

The marginal Sender types (those on the boundaries of the partition elements) must be indifferent between inducing adjacent best replies, that is,

$$a_i^N - \mu_i^N = \mu_{i+1}^N - a_i^N. \tag{2}$$

Equations (1) and (2) define the optimal quantization (or precoding) of the state space for a given number of messages, that is, the optimal way of coding and decoding the continuous source into the discrete set of actions. This problem is studied in information theory; early references are Lloyd (1957) published as Lloyd (1982) and Max (1960); Kieffer (1983) shows that the solution is unique a for a state distribution with a logconcave pdf. In contrast to the Crawford and Sobel (1982) model of strategic information transmission, there is no bias.[4]

If precoded language of cardinality N is used in communicating about the state, then the Receiver's best reply will match the true state of nature with some residual variance. Denote a typical residual variance of the state, conditional on it falling into partition ele-

[4]For more on the role of logconcavity in cheap talk with sender bias, see Szalay (2012).

ment P_i^N, by

$$\sigma_i^2 \equiv \mathbb{E}_\omega \left[\left(\omega - \mu_i^N \right)^2 \middle| \omega \in P_i^N \right].$$

The key incentive condition for the Sender to acquire information, when anticipating information transmission inducing the unique N-partition in the continuation, is whether the cost of acquiring information is "worth" the reduction in the expected residual variance of the state around the Receiver's induced actions. If the Sender decides not to acquire information (saving cost c) then he will forego this reduction, and induce the Receiver's action that is closest to the prior mean, μ_1^N.

The following lemma establishes when an equilibrium exists with information acquisition followed by communication via the optimal partitional language of cardinality N.

Lemma 1: There exists an equilibrium in which the Sender exerts effort ($e = 1$) and induces N different actions if, and only if,

$$\mathbb{E}\left[\left(\mu_i^N \right)^2 \right] + \left(\mu_1^N \right)^2 \geq c. \tag{3}$$

Proof: The Sender's expected utility on the equilibrium path, having acquired information about ω and anticipating to induce N different actions, is $-c - \mathbb{E}\left[\sigma_i^2\right]$.

If the Sender secretely deviated to $e = 0$ then he prefers to induce an action as close to the prior mean of ω as possible. In the equilibrium where the Receiver expects him to be informed and communicating in N different categories he is constrained to induce some action y_i^N. The constrained-optimal induced action conditional on not having learned the state is $y_1^N = \mu_1^N$, i.e., the mean of a partition element that is nearest to 0 (the prior mean). Therefore the Sender's maximal expected utility off the equilibrium path is

$$-\mathbb{E}_\omega \left[\left(\omega - \mu_1^N \right)^2 \right] = -\sigma^2 - \left(\mu_1^N \right)^2.$$

Comparing this expression with $-c - \mathbb{E}\left[\sigma_i^2\right]$ we find that information acquisition is indeed weakly preferred by the Sender in an equilibrium with N optimally-induced actions if, and only if,

$$-c - \mathbb{E}\left[\sigma_i^2\right] \geq -\sigma^2 - \left(\mu_1^N \right)^2. \tag{4}$$

By a variance decomposition identity,

$$\mathbb{E}\left[\sigma_i^2\right] + \mathbb{E}\left[\left(\mu_i^N\right)^2\right] = \sigma^2.$$

Combining this with inequality (4), we find that the condition that the Sender prefers to acquire information in an equilibrium with N different actions induced via the optimal partitional language of size N, is equivalent to condition (3). □

Evidently, an odd number of categories destroys the incentive to acquire information, because $\mu_1^N = 0$. In fact, N must be not only even but also finite in our problem:

Lemma 2: $\mathbb{E}\left[\left(\mu_i^N\right)^2\right] + \left(\mu_1^N\right)^2 \geq c$ only if N is even and finite.

The proof, relegated to the Appendix, establishes that logconcavity of the density implies $\lim_{N\to\infty}\left\{\mu_1^N\right\} = 0$, and hence

$$\lim_{N\to\infty}\left\{\mathbb{E}\left[\left(\mu_i^N\right)^2\right] + \left(\mu_1^N\right)^2\right\} \leq \sigma^2 < c.$$

The reason is that categories get wider towards the tails. If the support is a compact interval, then it is easy to see that the first of very many categories must be very narrow, to make the partition fit into the support. If the support is $\mathbb{R}$, then the argument is quite a bit more subtle.[5]

It follows that $c > \sigma^2$ makes it impossible to have an equilibrium with information acquisition where ω is fully revealed or an equilibrium where ω is transmitted in arbitrarily fine categories. In contrast, in the next two sections we show that if c is greater than but close to σ^2, then there are equilibria with information acquisition followed by precoded information transmission using a small and even number of categories, provided the distribution of the state has tails that are strictly thinner than the exponential.

4 Incentives with Binary Categories

In this section we show that communication in binary categories always provides stronger incentives than fully revealing communication, and that when communication in binary

[5] Logconcavity implies nonincreasing mean residual life – which refers to the distance of the tail conditional expectation to the truncation point – and the proof shows and exploits that the difference between the highest induced action and the highest marginal type must be strictly positive.

terms induces the Sender to learn the state, this is indeed preferred by the players jointly, ex ante, to uninformative communication.

Suppose that communication is anticipated in binary categories, that is, language partition $\mathcal{P}$ is defined by $-a_1 = -\overline{\omega}$, $a_0 = 0$, $a_1 = \overline{\omega}$. The natural interpretation for this is that a message reveals either a "high" or a "low" ω relative to cutoff $a_0 = 0$.

The Receiver's induced actions are $y \in \{-\mu_+, \mu_+\}$, where $\mu_+ = \mathbb{E}\left[\omega | \omega \geq 0\right]$. These induced actions match the state in expectation conditional on the information revealed by the binary language. The residual variance, that is, the imprecision with which the induced action matches the state is $\sigma_+^2 = \sigma_-^2 = \mathbb{E}[(\omega - \mu_+)^2 | \omega \geq 0]$.

The condition for the existence of an equilibrium with information acquisition followed by information transmission in binary categories, inequality (4) for $N = 2$, is simply

$$-c - \sigma_+^2 \geq -\sigma^2 - \mu_+^2.$$

This holds for some $c > \sigma^2$ (the condition for infinitely many categories not being compatible with information acquisition in equilibrium, maintained throughout) precisely when

$$\mu_+ \geq \sigma_+. \tag{5}$$

Equivalently, the coefficient of variation (standard deviation divided by the mean) of the distribution of ω conditional on $\omega \geq 0$ must not exceed 1. This condition is always satisfied: The symmetric, logconcave distribution of ω truncated to the upper half of the support remains logconcave, and it is a known result from reliability theory (Barlow and Proschan, 1981) that every logconcave distribution exhibits a coefficient of variation that is less than or equal to 1.

The conclusion of the preceding analysis is summarized in the following proposition.

Proposition 1: A binary categorization sets (strictly) stronger incentives than fully revealing communication or communication with an arbitrarily fine categorization, $N \to \infty$, for all distributions with a (strictly) logconcave density.

With a binary language, as opposed to full revelation of the state, the informed Sender's utility is lower on equilibrium path. However, the Sender faces an even greater loss off the equilibrium path, because in case he does not acquire information, he is still forced to choose between one of the two possible induced Receiver actions. It is crucial that when a binary language is used, the Sender cannot deviate to not acquiring the information and then conveying this fact to the Receiver. Messages have no intrinsic meaning; in a binary-categorization equilibrium all elements of M are interpreted either as "$\omega \geq 0$" or "$\omega \leq 0$".[6] We note that fully revealing communication is not the same as communication in arbitrarily many categories. However, in light of Lemma 2, neither is helpful for incentives.

A binary-categorization equilibrium makes the Sender acquire information before communication, and this imposes a cost on him that exceeds his own benefit from inducing a Receiver action that better matches the state than the action that equals the prior mean, because $c > \sigma^2$. This begs the question whether the equilibrium in which binary categorization induces information acquisition and binary action is socially preferable to the no-effort, no-communication equilibrium. It always is:

Proposition 2: Whenever information acquisition followed by a binary decision is an equilibrium, the ex-ante expected joint surplus is weakly greater in this equilibrium than in the no-effort, babbling equilibrium.

Proof. As seen in the main text, information acquisition followed by binary decisions is an equilibrium if, and only if,

$$-c - \sigma_+^2 \geq -\sigma^2 - \mu_+^2.$$

Subtracting σ_+^2 on both sides and using the law of total variance,

$$-c - 2\sigma_+^2 \geq -\sigma^2 - \left(\mu_+^2 + \sigma_+^2\right) = -2\sigma^2.$$

This is precisely the condition that the expected joint surplus with information acquisition followed by a binary decision weakly exceeds the joint surplus generated by a babbling equilibrium. □

5 Monotonicity

In the previous section we have established that communication in a binary language provides stronger incentives for the Sender's covert information acquisition than communication in a language that would allow to describe the state perfectly. The next natural

[6]Note that the binary categorization still admits a babbling equilibrium, where S mixes uniformly between high and low and R learns nothing. So, there is an equilibrium that is as if ignorance were expressed. However, reaching this different equilibrium requires a change in the Receiver's strategy and belief, which are outside the control of the Sender.

question to ask is whether these incentives are indeed monotonic in the number of categories.

Formally, monotonicity of incentives in the number of categories is equivalent to

$$\mathbb{E}\left[\left(\mu_i^N\right)^2\right] + \left(\mu_1^N\right)^2 \tag{6}$$

being strictly decreasing in $N \equiv 2n$ for $n = 1, 2, \ldots$.[7] This formula is the left-hand side of condition 3 in Lemma 1. We establish two results. First, we establish local monotonicity at $n = 1$ for any distribution with a logconcave density. Second, we demonstrate global monotonocity in a subset of logconcave densities.

Proposition 3: For any (strictly) logconcave density, a binary categorization provides (strictly) stronger incentives for information acquisition than communication in four categories.

The proof is relegated to the Appendix. Adding more categories increases the value of communicating on equilibrium path. The value from decision-making when learning which of four partition elements the state belongs to is higher compared to just learning whether the state is high or low. This suggests that the result should actually go the other way. However, there is a second effect which counteracts the first. With more categories to choose from, the action that is closest to zero — the action that an uninformed sender would like to induce — necessarily moves closer to zero. The second effect tends to make it less unpleasant for the sender to be uninformed, because the error in decision-making is smaller. In fact, the second effect necessarily dominates the first for any logconcave density.

Two forces are behind this result. First, the marginal value of going from 1 to 2 categories must necessarily be larger than the marginal value from going from 2 to 4 categories. By (5), knowing whether the state is high or low reduces uncertainty by at least 50%, leaving at most 50% to making the categories finer. Second, the length of the intervals in a finer categorization must increase towards the extremes of the support.[8] This implies that μ_1^N moves in sufficiently fast.

It can be shown that Proposition 3 generalizes to a comparison of the binary categorization to categorizations with an arbitrary, even cardinality for any logconcave density:

[7] Recall that a language with an odd number of categories can never provide enough incentives for information acquisition.

[8] See the proof of Lemma 2.

binary categorization provides indeed the strongest incentives for information acquisition. However, we may want to establish a stronger notion of monotonicity: when is it the case that fewer categories provide more incentives? This question is substantially more difficult to address than the comparison with binary categories, because the equilibrium partitions for both cardinalities cannot be computed in closed form. We now turn to a class of distributions that allow us to overcome this obstacle. Recall that $\mathbb{E}[\omega] = 0$. We impose in what follows:

Assumption: the distribution of ω satisfies

$$\mathbb{E}[\omega | \omega \geq z] = \mu_+ + \alpha \cdot z \text{ for } z \geq 0 \text{ and } \alpha \in \left[\frac{1}{2}, 1\right]. \tag{7}$$

This class of distributions featuring linear tail conditional expectations is introduced in Deimen and Szalay (2019) where a number of results are shown. First, the associated density[9] is derived; it is logconcave for $\alpha \in \left[\frac{1}{2}, 1\right]$. Here $\alpha = \frac{1}{2}$ corresponds to the uniform distribution while $\alpha = 1$ corresponds to the Laplace distribution. The parameter α captures the weight in the tails of the distribution. Second, it is shown that the equilibrium variation of the induced actions is[10]

$$\mathbb{E}\left[\left(\mu_i^N\right)^2\right] = \frac{2}{2-\alpha}\mu_+^2 - \frac{\alpha}{2-\alpha}\left(\mu_1^N\right)^2. \tag{8}$$

This expression is derived from a dynamic programming procedure. To get the intuition, consider the case of $N = 4$, with natural categories "high" and "very high" (and analogously for the negative realizations). The indifference condition of the marginal type dividing the "very high" from the "high" category, the law of iterated expectations linking μ_1^4, μ_2^4 and μ_+ and the linearity of the very high mean in the marginal type allow us to eliminate the very high mean from $\mathbb{E}\left[\left(\mu_i^N\right)^2\right]$ to arrive at exactly expression (8). Dynamic programming is invoked when there are more categories and we apply this logic repeatedly.

Note that μ_1^N depends on the distribution and the number of categories, so the expression is not "closed form". However, the form is closed enough to reveal the effect of changing the number of categories:

Proposition 4. Suppose the state ω follows a distribution in the class satisfying (7) for

[9] We reproduce the density in the proof of Proposition 6 in the appendix.

[10] See the proof of Proposition 3 in the online appendix of Deimen and Szalay (2019).

$\alpha \in \left[\frac{1}{2}, 1\right]$. Then, (6) is decreasing in $N = 2n$ for all N. Finer categories provide less incentives for information acquisition for $\alpha < 1$. Incentives for information acquisition are independent of the number of categories for the Laplace distribution.

Proof. Substituting into (6), we obtain

$$\mathbb{E}\left[\left(\mu_i^N\right)^2\right] + \left(\mu_1^N\right)^2 = \frac{2}{2-\alpha}\mu_+^2 + \frac{2(1-\alpha)}{2-\alpha}\left(\mu_1^N\right)^2,$$

which depends on N only through its effect on μ_1^N. With more categories, μ_1^N must move closer to 0. Suppose not and thus $a_1^{N+2} \geq a_1^N$. Take a_1 as given and compute a_2 and so on by a forward equation (the arbitrage condition of a_1 and so on). As an implication of log-concave densities, the truncated means move slowly in the thresholds and the solutions of all forward equations are increasing in the initial condition a_1. Since $a_1 = a_1^N$ is set so as to have exactly enough space for $n-1$ thresholds in the positive part, there is too little space for n thresholds and a_1 needs to be lower to make more space.[11] □

No dynamic programming needs to be invoked for the uniform distribution, since everything can be computed explicitly. Let ω be uniform on $[-1, 1]$ with variance $\sigma^2 = 1/3$. For N even, the length of any interval in the partition is $\frac{2}{N}$ implying residual variances $\sigma_i^2 = 1/(3N^2)$ and $\mu_1^N = 1/N$. Using the variance decomposition identity, $\sigma^2 = \mathbb{E}\left[\sigma_i^2\right] + \mathbb{E}\left[\left(\mu_i^N\right)^2\right]$ we obtain $\mathbb{E}\left[\left(\mu_i^N\right)^2\right] = \frac{1}{3} - \frac{1}{3N^2}$, representing the on-path value of communication, an increasing function of N. Adding the off-path hassle, $\left(\mu_1^N\right)^2 = \frac{1}{N^2}$, we obtain

$$\mathbb{E}\left[\left(\mu_i^N\right)^2\right] + \left(\mu_1^N\right)^2 = \frac{1}{3} + \frac{2}{3N^2},$$

in accordance with the general procedure. On net, this value is decreasing in N, so incentives become weaker as N increases.

We can now illustrate the decreasing returns logic. In the limit as $N \to \infty$, we obtain $\lim_{N\to\infty} \mu_1^N = 0$. For the distributions in the class, it is easy to show[12] that the moments are related as $\mu_+^2 = \frac{2-\alpha}{2}\sigma^2$, so indeed the limit features perfectly informative communication. This implies that the value from communicating just high or low, $\frac{1}{2}\mu_-^2 + \frac{1}{2}\mu_+^2 = \mu_+^2$ (by symmetry) is equal to 75% of perfect communication $(\frac{2-\frac{1}{2}}{2})$ for the uniform distribution,

[11] A more detailed proof of this fact, applying to biased and as well as non-biased communication, can be found in Deimen and Szalay (2020).

[12] See the proof of Proposition 6, equation (12), for the density of the distribution which is used to compute the variance.

while the value for the Laplace distribution is 50%. Since the tail of a logconcave density is at most exponential, the Laplace is an extreme case, implying that for any distribution with a logconcave density, the first two messages contain most of the information that Sender and Receiver care about: there are decreasing returns to adding categories going from two to four categories. Moreover, for any N, the effect through reduced unpleasantness of being unable to induce the desired action if the Sender shirks always impacts the incentive constraint stronger than the effect of increased categories on on-path communication, except for the borderline case of the Laplace, where the two effects wash out exactly.

6 Properties of an Optimal Categorization

A Receiver-optimal categorization solves the problem

$$\max_{N} \mathbb{E}\left[\left(\mu_i^N\right)^2\right] - \sigma^2$$
$$s.t. \mathbb{E}\left[\left(\mu_i^N\right)^2\right] + \left(\mu_1^N\right)^2 \geq c.$$

Proposition 5. Suppose the state ω follows a distribution in the class satisfying (7) for $\alpha \in \left[\frac{1}{2}, 1\right]$. Then, the optimal language N^* is the largest N that meets the constraint. N^* is nonincreasing in c for $c \leq (2-\alpha)\,\sigma^2$. For $c > (2-\alpha)\,\sigma^2$, no information is acquired in equilibrium.

The proof follows directly from the monotonicity established in Proposition 4 combined with the fact that the maximal incentive is reached for $N = 2$ and equal to $2\mu_+^2 = (2-\alpha)\,\sigma^2$.

Fewer categories provide stronger incentives for information acquisition. An optimal categorization provides just enough incentives to guarantee that the Sender is informed. Given the structured environment described by (7), there is an inverse relationship between the cost of acquiring information and the number of categories the Sender can use to communicate, whenever meaningful communication is possible at all. Things that are more difficult to find out are described in coarser terms.

7 Extensions: Alternative Information Acquisition Technologies

We have maintained two restrictive assumptions so far: information acquisition is all-or-nothing and there is no commitment to the categorization. We now drop these assumptions. As a natural extension, we let the Receiver choose the partitions that the Sender gets to observe. A further, natural question that we address is how restrictive this new class of technologies is.

In this variant of our model, c reflects more of an attention cost than an information acquisition cost. Without spending c, the Sender does not observe any signal. As before, we let the Sender report one out of the $N = 2n$ distinct signals that are privately observed on equilibrium path. The Receiver's problem is now

$$\max_{0 \leq a_1 \leq a_2, \ldots, a_{n-1} \leq a_n = \overline{\omega}} \sum_{i=1}^{n} p_i \mu_i^2 - \sigma^2 \tag{9}$$

$$s.t. \sum_{i=1}^{n} p_i \mu_i^2 + \mu_1^2 \geq c$$

where $p_i = 2 \int\limits_{a_{i-1}}^{a_i} f(\omega)\, d\omega$.

The information transmission incentive constraints on equilibrium path — when the Sender pays the attention cost — are slack, since the Sender and Receiver have the same objectives and the Sender only knows the interval that the state belongs to. Off equilibrium path, after shirking on paying attention, the Sender would report that the state falls in category 1. The attention incentive constraint is deliberately identical to the previous analysis to facilitate comparison. Consequently, the only difference to the previous problem for any given N is the location of the marginal types dividing the categories.

The solution need not be different from the all-or-nothing information acquisition case. Recall that the incentive to pay attention – the left-hand side of the incentive constraint – is monotonic in n for the considered class of densities. Hence, for small N – and c not excessively high – the solution of the unconstrained problem meets the constraint and hence is optimal. Thus, the new perspective adds anything new only if N is relatively high so that the unconstrained solution violates the constraint. In that case, the Receiver can increase a_1 to control incentives and move $a_2, \ldots, a_{n-1}$ closer together to fit them into $[a_1, \overline{\omega}]$. To emphasize the difference to the previous case, we investigate the case where N

gets very large.

Proposition 6. In the limit as $N \to \infty$, for distributions with a linear tail conditional expectation, the optimal partitional learning achieves the same expected payoff as a policy that reveals the state perfectly for $\omega > a_1$ and pools states $\omega \in [0, a_1]$ together (and symmetrically on the negative half). If $c \in \left(\sigma^2, (2-\alpha)\,\sigma^2\right)$, it is optimal to have $a_1 > 0$: the first category has positive interval length. Moreover, the simple pooling policy is optimal in the class of bi-pooling mechanisms.

For realizations of the state in the tail of the distribution, it does not matter – in terms of expected utility – whether the sender learns and can reveal the state perfectly or whether he observes the optimal partition that is arbitrarily fine. States that are close to the prior mean are pooled together in two intervals of strictly positive length. As a result, the Sender cannot express that the state is just marginally positive or negative. Rather, a substantial up- or downward revision takes place to $\pm\mathbb{E}\left[\omega|\, \omega \in [0, a_1]\right]$. Since from a_1 onwards, the decision-schedule is the identity function, there is a further upwards jump right after a_1.

We conclude that some form of – at least partially – categorical communication remains robust in various specifications. Moreover, the optimal form of information provision does not allow the Sender to induce marginal upward or downward revisions of the Receiver's action but rather requires a discrete jump.

8 Discussion and Literature

Our point of departure is the Crawford and Sobel (1982) model of strategic information transmission, which explains communication in finitely-many categories in the face of differences in preferences over decisions. More recently, Sobel (2012) analyzes constraints on the number of messages that can be sent as a rationale for categorization, where a sender and a receiver invest simultaneously into abilities to understand and to transmit. Constraints on the quantity of information transmission have a long tradition in the literature on computer science and electrical engineering with early contributions by Lloyd (1982) (original manuscript from 1957) and Max (1960). The connection between the literatures — that strategic information transmission equals quantization with added incentives — is not widely known among economists to date. We analyze the quantization problem with a view to providing incentives for information acquisition, and complement the existing explanation for categorization by one based on moral hazard in information acquisition. Few contributions to date have studied information acquisition followed by cheap talk; exceptions are Pei (2015), Argenziano et al (2016) and Deimen and Szalay (2019). In contrast to our analysis, the number of messages is not a choice variable in these models. Deimen and Szalay (2019) develop the class of distributions that lend tractability to strategic information transmission beyond the uniform case and introduce the dynamic programming technique to characterize equilibrium utilities.

Most of our analysis assumes all-or-nothing information acquisition followed by cheap talk communication. Relative to the recent literature on Bayesian persuasion (Kamenica and Gentzkow (2011)), all-or-nothing information acquisition is restrictive. On the positive side, we drop the commitment in information transmission; we have the sender observe the outcome of information acquisition and choose to transmit whatever he wants. Building on Gentzkow and Kamenica (2016), Arieli et al (2019) and Kleiner et al (2020), we allow for more general ways to learn about the state. The receiver designs an experiment and the sender needs to pay attention to obtain a meaningful signal. For the quadratic objectives in our problem, an experiment is described by a distribution of conditional means. Any distribution that satisfies the law of iterated expectations and is at most as risky as the distribution of the state is a feasible experiment. Arieli et al (2019) and Kleiner et al (2020) show that the solution to problems of maximizing a function of conditional means subject to a majorization constraint is in the class of bi-pooling mechanisms, where either of two things happens, state by state: either the state is perfectly revealed or states in at most two intervals are pooled together and the sender observes the same signal for states in these two intervals. The information acquisition constraint places additional restrictions on the distribution of the posterior mean, so it is not obvious that the fully optimal form of information provision is still a bi-pooling mechanism.[13]

Our interest in generalizations lies in richer ways of quantization, where we allow the receiver to choose intervals that the receiver gets to observe, i.e., simple pooling policies. When the receiver can create arbitrarily many of such pools, then an optimal way to place them – an ex ante optimal form of quantization – is outcome equivalent to having two equally sized pools on each side of the prior mean and to reveal the state perfectly for realizations in the tails of the distribution. There is no gain from allowing for bi-pooling relative to this solution.

The form of the experiment reminds of Szalay (2005) where incentives for information

[13]More specifically, the distribution must have no mass around its mean. Clearly, this does not disprove the conjecture that the fully optimal mechanisms remains in the class of bi-pooling mechanisms.

acquisition in a delegation problem are set by eliminating the compromising options. In common with Szalay (2005), the induced choice schedule displays some commitment to choices in the tails of the distribution as opposed to choices close to the prior mean. The essential difference is a discontinuity of the choice schedule around the bound of the first category, a_1 in the present problem. The discontinuity arises because the actions taken by the Receiver are optimal against the information provided by the Sender as opposed to excessively extreme by commitment. This difference is essential: the current approach generates an *endogenous* commitment to extreme options, because the sender cannot distinguish states in the two sets adjacent to the prior mean. This endogenous commitment to the extreme choices is also present in Che and Kartik (2009) in a model of disclosure, and, more recently, in Lipnowski et al (2020) in a model of attention management. In common with the present approach, the inability to distinguish states around the prior mean generates better incentives to pay attention.

Appendix

Proof of Lemma 2. Consider any $a_1 \in [0, \overline{\omega})$. We show that logconcavity implies that the optimal partition of $[a_1, \overline{\omega}]$ features $\lim_{n\to\infty} (\mu_2 - a_1) = 0$. The case $a_1 = 0$ is Lemma 2, $a_1 > 0$ is needed for Proposition 6.

Interval length is nondecreasing towards the tails of the distribution. Since any logconcave density is strongly unimodal (see Dharmadhikari and Joag-Dev (1988) and references therein), the symmetric density is nonincresing over the positive half. For a strictly decreasing density, intervals increase strictly. For a nonincreasing density, the mean over any interval is at most equal (smaller for the strict case) to the midpoint of the truncation points; to satisfy the arbitrage condition, intervals must get longer.

If $\overline{\omega}$ is finite, then $\overline{\omega} - a_1$ is finite. Suppose there is an ε such that If $a_2 - a_1 \geq \varepsilon$ for any n. But then the interval partition does not fit into $[a_1, \overline{\omega}]$. Hence, $\lim_{n\to\infty} (\mu_2 - a_1) = 0$.

If $\overline{\omega} = \infty$, then note that logconcavity implies nonthick tails (relative to the exponential), strict logconcavity implies thin tails. We show that a_{n-1} must be finite. Consider first the case of strict logconcavity. Since intervals are nondecreasing, the one below a_{n-1} is at least as long as the ones below. By the arbitrage condition of a_{n-1}, $a_{n-1} - \mu_{n-1} = \mu_n - a_{n-1}$.

Now suppose that $\lim_{n\to\infty} a_{n-1} = \infty$. Due to thin tails we would need to have

$\lim_{n\to\infty}(\mu_n - a_{n-1}) = 0$. However, this would imply that all intervals must have zero length. But the support cannot be covered with countably infinitely many isolated points. Now, since $\lim_{n\to\infty} a_{n-1}$ is finite, on the finite support $[a_1, \lim_{n\to\infty} a_{n-1}]$, the argument above implies that there cannot be an $\varepsilon > 0$ such that $\lim_{n\to\infty} a_2 - a_1 \geq \varepsilon$.

Consider now the case of the Laplace distribution (exponential distribution on the positive part) which implies neither thin nor thick tails in the sense of constant mean residual life. For the Laplace case, $\mu_n - a_{n-1} = \mu_+$, a constant, independent of the location of a_{n-1}. Hence, the location of a a_{n-1} cannot be used so directly to argue it must be bounded.

Consider a_1 as given and construct an optimal quantization on $[a_1, \infty)$. Suppose there is $\varepsilon > 0$ such that $\lim_{n\to\infty} a_2 - a_1 \geq \varepsilon$. Since intervals are increasing and all intervals have positive length, $\varepsilon > 0$ implies that we can find a number $r < 1$ such that $a_i - a_{i-1} \leq r \cdot (a_{i+1} - a_i)$ for all $i = 2, \ldots, n-2$. Note that $a_{n-1} = a_1 + \sum_{i=2}^{n-1} (a_i - a_{i-1})$. Hence, the series is convergent and thus $\lim_{n\to\infty} a_{n-1} = a^* < \infty$. However, it is not possible to to have infinitely many intervals of length ε and longer on the interval $[a_1, a^*]$. Hence, the initial hypothesis that there is a ratio $r < 1$ is wrong. Since intervals are increasing, the ones closest to the low end must tend to zero. □

Proof of Proposition 3. For $N = 4$, denote the partition thresholds inserted on the upper and lower half of the support by $\pm a$, and the new conditional means on the upper (positive) half by μ_1 and μ_2. For $N = 2$, (6) is simply $2\mu_+^2$. Therefore the claim is $\mu_1^2 + \mathbb{E}[\mu_i^2] \leq 2\mu_+^2$ with strict inequality if the density of ω is logconcave.

Let $p = \Pr[\omega \in [0, a] \,|\, \omega \geq 0]$. By the law of iterated expectations, $\mu_+ = p\mu_1 + (1-p)\mu_2$, hence

$$(1-p)(\mu_2 - \mu_1) = \mu_+ - \mu_1, \tag{10}$$

where both sides are positive.

For any distribution with a (strictly) logconcave density $\frac{\partial}{\partial t}\mathbb{E}[\omega|\, \omega \geq t] \leq 1$ (< 1) for all t. (See Prékopa, 1973). Therefore, for any cutoff $a > 0$,

$$\mu_2 - \mu_+ \equiv \int_0^a \frac{\partial}{\partial t}\mathbb{E}[\omega|\omega \geq t]dt \leq a \tag{11}$$

with a strict inequality if the pdf of ω is strictly logconcave. In the four-partition, by the indifference of S at $\omega = a$, we have $\mu_2 - a = a - \mu_1$. Combined with (11) this yields

$$\mu_1 + \mu_2 \leq 2(\mu_+ + \mu_1),$$

with strict inequality if the pdf of ω is strictly logconcave. Note that both sides are positive. Multiply this and (10) together to obtain

$$(1-p)(\mu_2^2 - \mu_1^2) \leq 2(\mu_+^2 - \mu_1^2).$$

After rearranging, this is equivalent to

$$\mu_1^2 + p\mu_1^2 + (1-p)\mu_2^2 \leq 2\mu_+^2,$$

with strict inequality if the pdf of ω is strictly logconcave, which proves the claim. □

Online Appendix

Proof of Proposition 6:

We first derive first-order conditions. Transform the objective, using the law of total variance, back into its original form. Letting $\lambda > 0$ denote the shadow cost of increasing c, the Lagrangian takes the form

$$\mathcal{L} = -\sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} (\omega - \mu_i)^2 \hat{f}(\omega)\, d\omega + \lambda \left(-\sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} (\omega - \mu_i)^2 \hat{f}(\omega)\, d\omega - c + \sigma^2 + \mu_1^2 \right).$$

where $\lambda > 0$ measures the shadow cost of increasing c and $\hat{f} = 2f$ is the density conditional on $\omega \geq 0$. Setting $a_0 = 0$ and $a_n = \overline{\omega}$, and neglecting monotonicity constraints on a_i (because they are satisfied automatically at the solution), the first-order conditions for a_i, $i = 2, \ldots, n-1$ are

$$\left(-(a_i - \mu_i)^2 \hat{f}(a_i) + (a_i - \mu_{i+1})^2 \hat{f}(a_i) \right)(1 + \lambda) = 0,$$

while the first-order condition for a_1 is

$$\left(-(a_1 - \mu_1)^2 \hat{f}(a_1) + (a_1 - \mu_2)^2 \hat{f}(a_1) \right)(1 + \lambda) + \lambda 2 \mu_1 \mu_1' = 0.$$

The former condition is the same as the one we obtain without commitment, except for the fact that a_1 is different: rearranging the condition for a_1, we obtain

$$\left(-(a_1 - \mu_1)^2 \hat{f}(a_1) + (a_1 - \mu_2)^2 \hat{f}(a_1) \right) = -\frac{\lambda}{(1 + \lambda)} 2 \mu_1 \mu_1'.$$

Since the right side of the inequality is negative, the left side must be as well

$$(a_1 - \mu_2)^2 < (a_1 - \mu_1)^2,$$

implying that

$$a_1 > \frac{\mu_2 + \mu_1}{2},$$

so that the first threshold is set larger than in an equilibrium without commitment. The remaining thresholds are set at the optimal quantizers of the interval $[a_1, \overline{\omega}]$.

The first-order conditions for a_i, $i = 2, \ldots, n-1$ are necessary and sufficient for an optimum for given a_1 by the uniqueness of optimal quantizers. The first-order condition for a_1 is necessary for an optimum.

The proposition is proved by a series of Lemmata.

Lemma A1. Consider any given $a_1 \in [0, \overline{\omega})$. For any distribution with a logconcave density, the optimal partition of the interval $[a_1, \overline{\omega}]$ has the feature that $\lim_{n\to\infty} (\mu_2 - a_1) = 0$.

Proof: See the proof of Lemma 2 in the Appendix. □

Lemma A2. For the class of distributions with a linear tail conditional expectation, in the limit as $n \to \infty$, the expected value from communicating in the ex ante optimally quantized way converges to the expected value from revealing ω perfectly for $\omega \geq a_1$ and from pooling states together to the same signal for $\omega \in [0, a_1)$. Moreover, a_1 is bounded away from zero to meet the moral hazard constraint.

Proof: Deimen and Szalay (2019) show that the value from optimal partitioning above a_1 is

$$\mathbb{E}\left[\left(\alpha\mu_i - \alpha\mu_+\right)^2 \middle| \omega \geq a_1 \right] = \frac{\alpha}{2-\alpha}\left((\alpha\mu_+)^2 - (\alpha\mu_2)^2 \right) + 2\alpha^2 a_1 \left(\frac{\alpha}{2-\alpha}(\mu_+ + \mu_2) - \alpha\mu_+ \right).$$

By the law of iterated expectations, considering μ_i as the realizations of a discrete random variable

$$\mathbb{E}\left[\mu_i \middle| \omega \geq a_1 \right] = \mu_+ + \alpha a_1.$$

Using this fact as well as the fact that the conditional distribution, conditional on $\omega \geq a_1$, satisfies $\sum_{i=2}^{n} p_i = 1$, we find that

$$\sum_{i=2}^{n} p_i (\mu_i - \mu_+)^2 = \sum_{i=2}^{n} p_i (\mu_i)^2 - \mu_+^2 - 2\mu_+\alpha a_1.$$

Thus, dividing by α^2 and decentering again by adding $\mu_+^2 + 2\mu_+\alpha a_1$, we get

$$\mathbb{E}\left[(\mu_i)^2 \middle| \omega \geq a_1 \right] = \frac{\alpha}{2-\alpha}\left((\mu_+)^2 - (\mu_2)^2 \right) + 2a_1 \left(\frac{\alpha}{2-\alpha}(\mu_+ + \mu_2) - \alpha\mu_+ \right) + \mu_+^2 + 2\mu_+\alpha a_1.$$

Letting $a_1 = \mu_2 = a$ (which obtains due to Lemma A1)

$$\begin{aligned} & \frac{\alpha}{2-\alpha}\left(\mu_+^2 - a^2\right) + 2\alpha a\left(\frac{1}{2-\alpha}\left(\mu_+ + a\right) - \mu_+\right) + \mu_+^2 + 2\mu_+\alpha a \\ = \; & a^2 + 2a\left(\mu_+ - \left(1-\alpha\right)a\right) + \frac{2}{2-\alpha}\left(\mu_+ - a\left(1-\alpha\right)\right)^2. \end{aligned}$$

Consider now the value from the optimal pooling policy which reveals ω perfectly beyond some value a^* and reveals only that $\omega \in [0, a_1)$.

The density of distributions with a linear tail conditional expectation, conditional on $\omega \geq 0$, obtained in Deimen and Szalay (2019), is

$$\hat{f}\left(\omega\right) = \alpha\mu_+^{-\frac{\alpha}{1-\alpha}}\left(\mu_+ - \omega\left(1-\alpha\right)\right)^{\frac{2\alpha-1}{1-\alpha}} \tag{12}$$

with associated survival function conditional on $\omega \geq 0$

$$\left(1 - \hat{F}\left(\omega\right)\right) = \mu_+^{-\frac{\alpha}{1-\alpha}}\left(\mu_+ - \omega\left(1-\alpha\right)\right)^{\frac{\alpha}{1-\alpha}}$$

Let $\tilde{f}\left(\omega\right) = \frac{\hat{f}(\omega)}{1-\hat{F}(a)}$. Integrating by parts twice, we find

$$\int_a^{\overline{\omega}} \omega^2 \tilde{f}\left(\omega\right) d\omega = a^2 + 2a\left(\mu_+ - \left(1-\alpha\right)a\right) + \frac{2}{2-\alpha}\left(\mu_+ - a\left(1-\alpha\right)\right)^2.$$

The same value is feasible in both problems. Hence, the solutions satisfy $a_1 = a^*$.

Consider the level of the left-side of the incentive constraint as a function of a_1 :

$$\left(1+\hat{F}\left(a_1\right)\right)\left(\mu_1\left(a_1\right)\right)^2 + \left(1-\hat{F}\left(a_1\right)\right)\left(a_1^2 + 2a_1\left(\mu_+ - \left(1-\alpha\right)a_1\right) + \frac{2}{2-\alpha}\left(\mu_+ - a_1\left(1-\alpha\right)\right)^2\right),$$

where $\mu_1\left(a_1\right) = \mathbb{E}\left[\omega \middle| \omega \in [0, a_1]\right]$.

We must have $a_1 > 0$. If $a_1 = 0$, then the left side of the constraint attains value $\frac{2}{2-\alpha}\left(\mu_+\right)^2 = \sigma^2 < c$, so that the constraint is violated. On the other hand, for $a_1 = \overline{\omega}$, the left side of the constraint takes value $2\left(\mu_+\right)^2 = \left(2-\alpha\right)\sigma^2 > c$.

By continuity, for $\alpha < 1$ and $c \in \left(\sigma^2, \left(2-\alpha\right)\sigma^2\right)$, there exists a_1 that meets the constraint. □

Lemma A3. The optimal bi-pooling policy is a simple pooling policy that reveals ω perfectly for $\omega \geq a_1$ and from pooling states together to the same signal for $\omega \in [0, a_1)$.

Proof: The optimal bi-pooling policy solves

$$\max_{0\leq a_1\leq a_2\leq a_3\leq\overline{\omega}} p\mu_1^2+\int_{a_1}^{a_2}\omega^2\hat{f}(\omega)\,d\omega+\int_{a_3}^{\overline{\theta}}\omega^2\hat{f}(\omega)\,d\omega \tag{13}$$

$$s.t. p\mu_1^2+\int_{a_1}^{a_2}\omega^2\hat{f}(\omega)\,d\omega+\int_{a_3}^{\overline{\theta}}\omega^2\hat{f}(\omega)\,d\omega+\mu_1^2=c$$

$$p=\int_{0}^{a_1}\hat{f}(\omega)\,d\omega+\int_{a_2}^{a_3}\hat{f}(\omega)\,d\omega$$

$$\mu_1=\frac{\int_0^{a_1}\omega\hat{f}(\omega)\,d\omega+\int_{a_2}^{a_3}\omega\hat{f}(\omega)\,d\omega}{p}$$

$$\mu_1\leq a_1.$$

The last constraint ensures that the agent who shirks wants to claim he believes the conditional mean is μ_1.

Step 1: a policy is optimal only if it is either a simple pooling policy that reveals ω perfectly beyond some a_1 and pools states together for $\omega\in[0,a_1)$ or it is a bi-pooling policy with one pool at the low end and one pool at the top end; formally $a_3=\overline{\omega}$.

To prove this, suppose contrary to the claim that $a_1<a_2<a_2<\overline{\omega}$. It suffices to consider policies such that

$$\frac{\int_0^{a_1}\omega\hat{f}(\omega)\,d\omega+\int_{a_2}^{a_3}\omega\hat{f}(\omega)\,d\omega}{\int_0^{a_1}\hat{f}(\omega)\,d\omega+\int_{a_2}^{a_3}\hat{f}(\omega)\,d\omega}=\mu_1=\text{constant},$$

Differentiating totally wrt a_2 and a_3, keeping a_1 and μ_1 constant, we obtain

$$\left(\frac{\hat{f}(a_3)\,a_3}{p}-\frac{\hat{f}(a_3)\,\mu_1}{p}\right)da_3-\left(\frac{\hat{f}(a_2)\,a_2}{p}-\frac{\hat{f}(a_2)\,\mu_1}{p}\right)da_2=0.$$

Provided that $a_1<a_2$ and $a_3<\overline{\omega}$, we can adjust a_3 after a change of a_2 such that

$$\frac{da_3}{da_2}=\frac{\hat{f}(a_2)}{\hat{f}(a_3)}\frac{a_2-\mu_1}{a_3-\mu_1}.$$

Taking account of this relationship, the objective and constraint change according to

$$\begin{aligned} & \left(\hat{f}\left(a_3\right) da_3 - \hat{f}\left(a_2\right) da_2\right) \left(\mu_1\right)^2 + b^2 \hat{f}\left(a_2\right) da_2 - \left(a_3\right)^2 \hat{f}\left(a_3\right) da_3 \\ = \; & \hat{f}\left(a_2\right) da_2 \left(\left(a_3 - a_2\right)\left(\mu_1 - a_2\right)\right) \end{aligned}$$

Hence, by raising a_2 (and adjusting a_3 upwards accordingly), we can raise the payoff and the left side of the constraint, contradicting the supposed optimality.

Noting that $a_1 = a_2 < a_3 < \overline{\omega}$ is in fact a simple pooling policy with only one pool at the low end and $a_1 < a_2 < a_3 = \overline{\omega}$ is a bi-pooling policy with a pool at the low end and a pool at the high end, completes the proof of the step.

We can relabel thresholds, since at most two of them are needed.

Step 2: a bi-pooling policy with a pool at the low end and a pool at the top end is optimal only if $a_1 = \mu_1$ and $a_2 < \overline{\omega}$ or $a_1 > \mu_1$ and $a_2 = \overline{\omega}$.

Suppose contrary to the claim that $\mu_1 < a_1 < a_2 < \overline{\omega}$. Given that $c < (2-\alpha)\sigma^2$, $a_1 = a_2$ is not optimal and hence need not be considered.

Considering again policies such that μ_1 is constant, differentiating totally wrt a_1 and a_2, we obtain

$$\left(\frac{\hat{f}\left(a_1\right) a_1}{p} - \frac{\hat{f}\left(a_1\right) \mu_1}{p}\right) da_1 - \left(\frac{\hat{f}\left(a_2\right) a_2}{p} - \frac{\hat{f}\left(a_2\right) \mu_1}{p}\right) da_2 = 0.$$

Provided that $\mu_1 > a_1$ and $a_2 < \overline{\omega}$

$$\frac{da_2}{da_1} = \frac{\hat{f}\left(a_1\right)}{\hat{f}\left(a_2\right)} \frac{\left(a_1 - \mu_1\right)}{\left(a_2 - \mu_1\right)}.$$

Taking account of this, objective and incentive constraint change according to

$$\begin{aligned} & \left(\hat{f}\left(a_1\right) da_1 - \hat{f}\left(a_2\right) da_2\right) \left(\mu_1\right)^2 - \left(a_1\right)^2 \hat{f}\left(a_1\right) da_1 + \left(a_2\right)^2 \hat{f}\left(a_2\right) da_2 \\ = \; & \hat{f}\left(a_1\right) da_1 \left(\mu_1 - a_1\right)\left(a_1 - a_2\right). \end{aligned}$$

Hence, if $\mu_1 > a_1$ and $a_2 < \overline{\omega}$, then we can raise the objective by increasing a_1.

Hence, either the policy is a simple interval pooling policy with pooling at the low end only, in which case we are done, or it is a bi-pooling policy with $\mu_1 = a_1$ and $a_2 < \overline{\omega}$.

Step 3: A bi-pooling policy with $\mu_1 = a_1$ and $a_2 < \overline{\omega}$ is suboptimal.

The Lagrangian of problem (13), using the structure provided by the first steps, is

$$L = p(a,b)(\mu_1(a,b))^2 + \int_a^b \omega^2 \hat{f}(\omega)\,d\omega \qquad +\lambda\left((p(a,b)+1)(\mu_1(a,b))^2 + \int_a^b \omega^2 \hat{f}(\omega)\,d\omega - c\right)$$
$$+\delta_a\left((\mu_1(a,b)) - a_1\right) + \delta_b\left(a_2 - \overline{\omega}\right)$$

where $\delta_a \geq 0$ and $\delta_b \geq 0$ are the Kuhn-Tucker constraints on the constraints $\mu_1(a,b) - a_1 \leq 0$ and $a_2 - \overline{\omega} \leq 0$. We note that the constraints $a_1 \geq 0$ and $a_2 \geq a_1$ are automatically satisfied. To see this, note that $a_1 = 0$ implies $\mu_1 = 0$ and thus $a_2 = \overline{\omega}$, which does not meet the constraint. Likewise, $a_1 = a_2$ provides too much incentives for information acquisition given that $c < (2-\alpha)\sigma^2$.

Taking derivatives, necessary conditions for a solution are

$$(1+\lambda)\hat{f}(a_1)\left((\mu_1(a_1,a_2))^2 - (a_1)^2\right) + \lambda 2\mu_1(a_1,a_2)\frac{\partial \mu_1(a_1,a_2)}{\partial a_1}$$
$$+(1+\lambda)2p(a_1,a_2)\mu_1(a_1,a_2)\frac{\partial \mu_1(a_1,a_2)}{\partial a_1} + \delta_a\left(\left(\frac{\partial \mu_1(a_1,a_2)}{\partial a_1}\right) - 1\right) = 0$$

and

$$-(1+\lambda)\hat{f}(a_2)\left((\mu_1(a_1,a_2))^2 - (a_2)^2\right) + \lambda 2(\mu_1(a_1,a_2))\frac{\partial \mu_1(a_1,a_2)}{\partial b}$$
$$+(1+\lambda)2p(a_1,a_2)\mu_1(a_1,a_2)\frac{\partial \mu_1(a_1,a_2)}{\partial b} + \delta_a\left(\left(\frac{\partial \mu_1(a_1,a_2)}{\partial b}\right)\right) + \delta_b = 0$$

We note that

$$\frac{\partial \mu_1(a_1,a_2)}{\partial a_1} = \left(\frac{\hat{f}(a_1)a_1}{p(a_1,a_2)} - \frac{\hat{f}(a_1)\mu_1(a_1,a_2)}{p(a_1,a_2)}\right)$$

Substituting back into the first-order condition for a_1 we obtain,

$$-(1+\lambda)\hat{f}(a_1)\left((\mu_1(a_1,a_2)) - a_1\right)^2 + \lambda 2\mu_1(a_1,a_2)\frac{\hat{f}(a_1)}{p(a_1,a_2)}(a_1 - \mu_1(a_1,a_2))$$
$$+\delta_a\left(\frac{\hat{f}(a_1)}{p(a_1,a_2)}(a_1 - \mu_1(a_1,a_2)) - 1\right) = 0.$$

$a_1 - \mu_1(a_1,a_2)$ is a stationary point only if $\delta_a = 0$. Hence, we must be able to find the stationary point from the unconstrained problem which neglects the constraint a priori.

However, the derivative of the unconstrained problem

$$-\left(1+\lambda\right)\hat{f}\left(a_1\right)\left(\left(\mu_1\left(a,b\right)\right)-a_1\right)^2+\lambda 2\mu_1\left(a_1,a_2\right)\frac{\hat{f}\left(a_1\right)}{p\left(a_1,a_2\right)}\left(a_1-\mu_1\left(a,b\right)\right)$$

is strictly negative for $a < \mu_1\left(a,b\right)$, and either negative or positive for $a > \mu_1\left(a,b\right)$, depending on the value of λ. It follows that $a = \mu_1\left(a,b\right)$ is either a saddle point or a minimum of the unconstrained problem, but never a maximum, as would be required for the solution to take this form.

It follows from steps 1-3 that the solution in the class of bi-pooling policies is a simple pooling policy. □

References

An, M. Y. (1998), "Logconcavity versus Logconvexity: A Complete Characterization", *Journal of Economic Theory*, 80, 350-369.

Argenziano, R., S. Severinov, and F. Squintani (2016), "Strategic Information Acquisition and Transmission", *American Economic Journal: Microeconomics*, 8(3), 119-155.

Arieli, I. Babichenko, Y., Smorodinsky, R. and Yamashita, T. (2019), "Optimal Persuasion via Bi-Pooling", *mimeo*

Barlow, R. E., and Proschan, F. (1981), *Statistical theory of reliability and life testing: probability models*, To Begin With: Silver Springs, Maryland.

Bagnoli, M. and Bergstrom, T. (2005), "Logconcave Probability and its Applications", *Economic Theory*, 26(2), 445-469.

Crawford, V. and Sobel, J. (1982), "Strategic Information Transmission", *Econometrica*, 50, 1431–1451.

Deimen, I., and D. Szalay (2019), "Delegated Expertise, Authority, and Communication", *American Economic Review.*Vol 109(4), 1349-74

Deimen, I., and D. Szalay (2020), "Authority in a Theory of the Firm", *mimeo.*

Dharmadhikari, S., and Joag-dev, K. (1988) "Unimodality, Convexity, and Applications", *Academic Press.*

Kleiner, A., Moldovanu, B. and Strack, P. (2020) "Extreme Points and Majorization: Economic Applications", *mimeo*

Kieffer, J. (1983), "Uniqueness of locally optimal quantizer for log-concave density and convex error weighting function", *IEEE Transactions on Information Theory*, 29, 42–47.

Lipnowski, E, Mathevet, L., and Wei, D. (2020), "Attention Management", *American Economic Review: Insights*, 2(1), 17–32.

Lloyd, S.P (1982), "Least Squares Quantization in PCM", *IEEE Transactions on Information Theory*, 28, 129–1377.

Max, J (1960), "Quantizing for Minimum Distortion", *IRE Transactions on Information Theory*, 6, 7–12.

Pei, H. D. (2015), "Communication with Endogenous Information Acquisition", *Journal of Economic Theory*, 160, 132–149.

Sobel, J. (2012), "Complexity versus Conflict in Communication," in *Proceedings of 46th Annual CISS (Conference on Information)*, Princeton, NJ, 2012, pp. 1-6

Szalay, D. (2005), "The Economics of Extreme Options and Clear Advice", *Review of Economic Studies*, 72, 1173-1198.

Szalay, D. (2012), "Strategic information transmission and stochastic orders", *mimeo*.