Clarifying by Discretizing

Jordan Martel, Edward D Van Wesep, and Robert A Van Wesep*

June 27, 2017

Abstract

Cheap talk often occurs in the form of discrete messages, such as Yelp ratings, film reviews, student grades, and debt ratings. In the classic Crawford and Sobel (1982) model, coarse messaging is always possible in equilibrium, but is not maximally informative if the sender's and receiver's interests coincide. In the situations listed above, however, the sender's and receiver's interests likely *do* coincide and the message functions appear to be deliberately chosen to optimally serve the sender's and receiver's interests. We provide a simple model in which signals are received with some exogenous noise and show that the optimal message function is often discrete: *while discrete messages are less precise, they are easier to interpret.* The noise structure determines whether messaging should be discrete and, if so, how coarse the optimal message function should be.

JEL: D83, G00, M00

Keywords: Cheap Talk, Discrete, Coarse, Yelp, Credit Ratings, Stock Analysts

*Jordan Martel: University of Colorado Boulder, Leeds School of Business. email: jordan.martel@colorado.edu. Edward Van Wesep: University of Colorado Boulder, Leeds School of Business. email: edward.vanwesep@colorado.edu. Robert Van Wesep: email: ravanwesep@gmail.com. The authors would like to thank Nageeb Ali, Tony Cookson, Brendan Daley, Brett Green, Chong Huang, Marina Niessner, and Joel Sobel for helpful comments.

Cheap talk often occurs in the form of discrete messages, such as Yelp ratings (1-5 stars), film reviews (1-4 stars; "thumbs up or down"), corporate credit ratings (AAA to D), and student grades (A+ to F). In the classic model of Crawford and Sobel (CS, 1982), optimal communication devices are discrete only when the interests of the sender and receiver are not fully aligned.[1] An extensive literature has consequently developed to analyze the differences in interests in each of the above settings (and many others), presumably in part because the existence of discrete messaging has been seen as an indication that such differences must be critical.[2]

We show that optimal ratings can be discrete *even with a perfect coincidence of interests* so long as messages are received with some exogenous error. CS assume that whatever message the receiver sends is interpreted perfectly by the receiver. In practice, this is not likely to be true if the message space is large.[3] For example, suppose that the message space permits any message in the interval $[0, 1]$. Presumably, when we take the model to the real world, we do not believe that any sender actually sends, for example, an irrational number as a message. Instead, real-world examples of continuous messages are 400 word restaurant reviews, 600 word movie reviews, or 10,000 word stock analyst reports. If this is how a continuum of messages is implemented in practice, then there may be some difficulty in perfectly interpreting the message. Even a simple message like "I give this student a B+" may be difficult for a potential employer or graduate admissions officer to interpret.

As it is critically important for this paper, we stress that both a film critic and a reader should agree on the words found in a film review. Both the professor and the employer should be able to correctly identify if the grade "B+" appears on a transcript. The difficulty is in interpreting what message the critic or the professor intended to send with those words and grade. Our assumption is that language is perfectly readable, but not perfectly interpretable.

We show that, if messages are interpreted with some error, then discrete messages come with a cost and a benefit.[4] On the one hand, restricting the sender to a finite number of discrete messages when the state space is infinite means discarding information. Discrete messages restrict the sender's ability to precisely "say what she wants." On the other hand, discrete messages allow for accurate interpretation: there is no mistaking the intention of a four star movie review for that of a three star review. Put simply: *discrete messages are less precise, but easier to interpret.*

[1] CS studies the complete set of equilibria in a cheap talk game, not simply the optimal (maximally informative) equilibrium. Coarse equilibria can occur even when the sender's and receiver's interests align: babbling equilibria, for example, are everpresent in CS models. These equilibria, however, are not optimal unless they are unique. To the extent that we see discrete message functions in practice that seem to be actively chosen by the sender and receiver, it is unlikely that they reflect a babbling equilibrium.

[2] See Sobel (2013) for an extensive review.

[3] Indeed, many a comedy has been based on the premise of misunderstanding!

[4] The errors that we study are exogenous errors in interpretation, as in Shannon (1948). More recent work has tended to focus on endogenous errors. For example, there is work concerning strategic errors, in which noise intentionally inserted into messaging improves outcomes. Myerson (1991), for example, studies a simple game in which the sender can either send or not send a message, and if the message is sent, it may or may not be received. Non-receipt therefore is associated both with the message not being sent, and with the message being sent and lost. Since Myerson (1991), a substantial literature on endogenous noise has developed. See Blume, Board, and Kawamura (2007) and references therein.

In this short paper, we establish four theoretical points. First, when errors of interpretation exist, discrete messaging can be optimal even though it discards information: there need not be any misaligned interests in order to drive stock analysts to use a simple rating system, like "strong buy" to "strong sell."

Second, when the maximum error of interpretation is larger, then the optimal number of messages is smaller. The fact that only five Yelp ratings are used suggests that it is difficult for a customer deciding on a restaurant to understand just what a reviewer's experience at the restaurant was. The fact that there are 21 corporate credit ratings means that it is easier to convey a firm's credit-worthiness.

Third, the advantage of discrete messages over continuous messages shrinks as the maximum error of interpretation shrinks.

Fourth, for any given maximum error of interpretation, if the variance in the error within that domain is small enough, discrete messaging is not optimal. Discrete messages are ideal if the interpretation error has fat tails – i.e., if the likelihood of large errors is not remote relative to the likelihood of small errors.

We believe that this explanation for discrete reviews has empirical relevance. The consumer rating website Yelp collects user ratings of businesses on a one to five star rating scale. Yelp then aggregates those ratings into a single weighted average, applying weights which may account for its perception of each review's quality.[5] Those weighted averages are precise (for example, 3.8662 stars). Yelp then rounds to the nearest half-star before reporting a grade to consumers who use the website. This rounding is excellent for economists, who use it as the basis for estimating the effect of Yelp reviews on business performance, but it is not clear why Yelp would want to throw away this information by rounding.[6] Journalists at Vice.com asked a Yelp representative why they round, and received the following response:[7]

> "Well, we wanted to convey something that was easy for our audience to understand, and so if you're thinking of star ratings for a business on a scale of one to five, it's pretty clear-cut and dry."

This is precisely the point of our model. It is true that displaying the information in a discrete way with only nine possible ratings is less precise, but it is easier for readers to interpret. Sometimes, the benefit of the latter exceeds the cost of the former.

[5] For example, if a review is perceived to be associated with a competing business, it may be discounted.

[6] Luca (2016) and Anderson and Magruder (2012) use the fact that Yelp rounds to the nearest half star to compare restaurants with similar average reviews but different reviews as seen by customers. For example, a restaurant with an average of 3.24 stars will show up as a "3 star" restaurant on Yelp, whereas one with an average of 3.26 stars will show up as a "3.5 star" restaurant. They find that Yelp reviews have a substantial effect on performance.

[7] The Yelp Factor (2017). Vice Money. https://news.vice.com/story/heres-how-yelp-reviews-affect-restaurant-success

There is a budding literature on discrete messages that are not based upon the classic CS assumption of differing interests between the sender and receiver. Notably, Sobel (2012), Cremer, Garicano, and Prat (2007), and Garicano and Prat (2011) offer models in which messaging is by assumption discrete, and more precise messaging is directly costly (perhaps infinitely so). In our model, there is no direct cost to creating more or less precise messages. Talk is cheap, and errors are independent of the choice of message function.

Agranov and Schotter (2012) and Crawford, Gneezy and Rottenstreich (2008) analyze a model in which the sender is benevolent, but there are multiple receivers who do not share preferences. In these models, discrete messages can be optimal. While the precise reason for discretization is different than in CS, the assumption of different preferences for various actors is still present.

Our goal is to hew as closely to the original CS motivation as possible, without assigning any direct costs to more precise messaging. In our model, continuous and discrete message functions are equally costly for the sender and receiver. The message is received with error, but that error is independent of the message function. There is no obvious reason to believe that it would be worth "throwing away" information by discretizing. And yet, there is.

For the sake of clarity, we develop the model with the simplest possible assumptions and keep the paper concise. The model should be seen as an "existence" result: we show that a particular discrete message function can be optimal under certain natural assumptions, but there are many other choices of assumptions that one could make. We show that, under some other choices, discrete messaging cannot be optimal, but we do not attempt to fully characterize the set of optimal message functions under any arbitrary set of assumptions. We leave a discussion of interpretations to the conclusion, but note here that this model can be considered a model of optimal experimental design or optimal design of statistical tests as well as a model of communication between two people.

1 The model

Let there be a state of nature $q \sim U[0,1]$ which is known to the sender. The sender chooses an *intended* message $m \in [0,1]$. The receiver observes a noisy version of the intended message, which we call the *received* message, $\widetilde{m} = m + \varepsilon$, where ε is distributed according to a symmetric and continuous distribution f_e over $[-\overline{\varepsilon}, \overline{\varepsilon}]$, with $\overline{\varepsilon} = \frac{1}{2N} > 0$ for some integer N.[8] The receiver then takes an action $A \in [0,1]$. Both sender and receiver have utility $U(q, A) = -(A-q)^2$. The timeline of the game is shown in Figure 1.

Prior to the start of the game, the sender can specify an intended message function $m(q)$ that she will use. The receiver chooses an action based upon the function $m(q)$ and the received message

[8] Note that values of $\overline{\varepsilon}$ which do not divide $N/2$ for some integer N may not yield discrete messages as optimal. As discussed above, our model establishes that discrete messages can be optimal, but does not establish that they are always optimal.

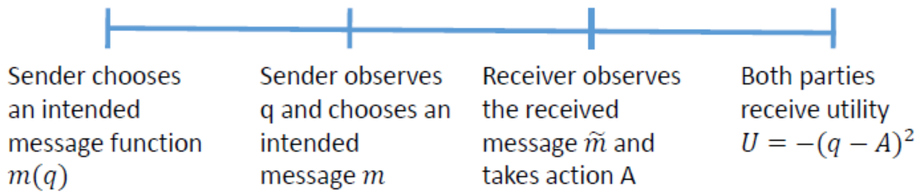

Figure 1: Timeline.

$\widetilde{m}$, and we denote this function $A(\widetilde{m})$. As we will see, commitment will not be an issue in this game: once a message function and action function are specified, neither party has any incentive to deviate.

The objective is to find the intended message function $m(q)$ and action function $A(\widetilde{m})$ to maximize the expected payoff of the sender and receiver. Let $\langle\widetilde{m}(q)\rangle$ be the set of possible received messages given the message $m(q)$ and let R be the set of values of $\langle q, \tilde{m}(q)\rangle$. R is the region in $\mathbb{R}^2$ bounded on the left by $q = 0$, on the right by $q = 1$, above by the graph of $q \mapsto m(q) + \bar{\varepsilon}$ and below by the graph of $q \mapsto m(q) - \bar{\varepsilon}$. Note that

$$\text{the area of } R = 2\bar{\varepsilon}. \tag{1}$$

Suppose $q \mapsto m(q)$ is monotone increasing.[9] Then for any $\tilde{m} \in [-\bar{\varepsilon}, 1 + \bar{\varepsilon}]$, the set of values of q that might have given rise to $\tilde{m}$ is an interval $[q_-(\tilde{m}), q_+(\tilde{m})]$. Let $w = w(\tilde{m}) \equiv q_+(\tilde{m}) - q_-(\tilde{m})$ be the width of this interval. Note that

$$\int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} w(\tilde{m})\, d\tilde{m} = \text{the area of } R = 2\bar{\varepsilon}, \tag{2}$$

so the average value of $w(\tilde{m})$ (over the interval $[-\bar{\varepsilon}, 1 + \bar{\varepsilon}]$) is

$$\frac{2\bar{\varepsilon}}{1 + 2\bar{\varepsilon}}.$$

The task of the receiver is to choose $A(\tilde{m})$ to maximize her expected utility. See Figure 2 for a visual reference.

[9] The optimal message function can be easily shown to be monotone, and the requirement that it is increasing is without loss of generality.

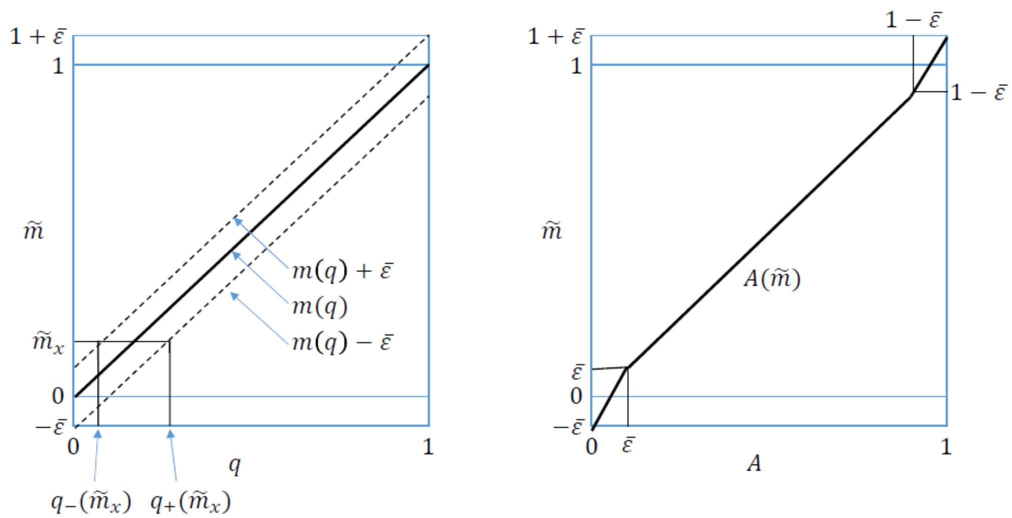

Figure 2: In the left panel, the solid line $m(q)$ plots the identity intended message function $m(q) = q$. The dashed lines show the upper and lower bounds of the received messages. The parallelogram between the dashed lines is the support of $\langle q, \tilde{m}(q) \rangle$ and has area $2\bar{\varepsilon}$. For a given received message $\widetilde{m}_x$, the state associated with that message must lie in the interval $[q_-(\tilde{m}), q_+(\tilde{m})]$. In the right panel, we plot the optimal action $A(\tilde{m})$ as a response to the receive message $\tilde{m}$, for the identity intended message function plotted in the right panel, assuming uniform errors. The optimal action lies in the middle of the interval $[q_-(\tilde{m}), q_+(\tilde{m})]$. Between $\bar{\varepsilon}$ and $1-\bar{\varepsilon}$, this action is simply $A(\tilde{m}) = \tilde{m}$, because the distribution of the state is symmetric about $\tilde{m}$. Because messages less than zero or greater than one are never sent, the action function kinks, and is twice the slope for received messages $\tilde{m} < \bar{\varepsilon}$ or $\tilde{m} > 1-\bar{\varepsilon}$.

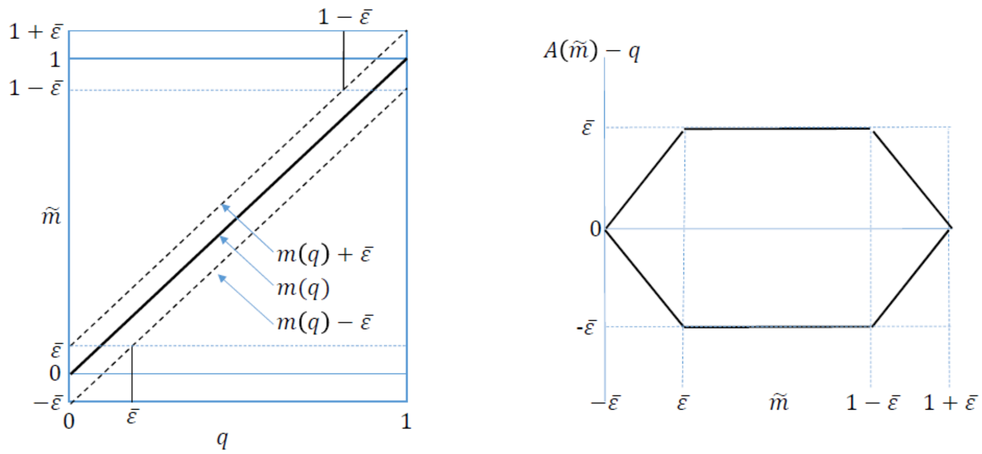

Figure 3: The left panel largely reproduces the left panel of Figure 2. The solid line plots the identity intended message function. The dashed lines plot the associated maximum-error received message functions. The right panel moves $\tilde{m}$ to the horizontal axis, and plots the width of the band between the dashed lines on the vertical axis. This width is plotted symmetrically about the horizontal axis. The resulting hexagon is the region R, with a change of variables. The received message and the error in the resulting action $(A(\tilde{m})-q)$ are uniformly distributed over the hexagon.

1.1 The case of uniform errors

Consider first the case that ε is uniformly distributed. Then the distribution of $\langle q, \tilde{m}(q)\rangle$ is uniform in R.

Lemma 1 *If f_e is uniform, then the optimal action is $A(\tilde{m}) = \bar{q}(\tilde{m}) \equiv (q_-(\tilde{m}) + q_+(\tilde{m}))/2$, which is the expected value of q given $\tilde{m}$.*

Proof. For a given received message $\tilde{m}$, the expected utility from an action a is

$$-\int_{q_-(\tilde{m})}^{q_+(\tilde{m})} (q-a)^2\, dq = \frac{1}{3}(q_-(\tilde{m})-a)^3 - \frac{1}{3}(q_+(\tilde{m})-a)^3,$$

which is evidently concave, and whose derivative with respect to a is $(q_+(\tilde{m})-a)^2 - (q_-(\tilde{m})-a)^2$ which equals zero at $a = (q_-(\tilde{m}) + q_+(\tilde{m}))/2$. ■

The optimal action function for the identity intended message function, $m(q) = q$, is plotted in the right panel of Figure 2 as an example.

Given the optimal action derived in Lemma 1, we can perform a change of variables in order to plot the region R in a way that will provide some intuition for the results to come. The left panel of Figure 3 largely reproduces the left panel of Figure 2. The solid line plots the identity intended message function and the dashed lines plot the maximum and minimum received messages for each state q. We have added $\bar{\varepsilon}$ and $1-\bar{\varepsilon}$ to the horizontal axis to note that the width of $[q_-(\widetilde{m}), q_+(\widetilde{m})]$ is constant between those points, and linearly decreasing to zero above and below.

The right panel moves $\widetilde{m}$ to the horizontal axis and plots the width of $[q_-(\widetilde{m}), q_+(\widetilde{m})]$ on the vertical axis. This width is plotted symmetrically about the horizontal axis so that the deviation from zero is the error in the action, i.e., $A(\widetilde{m}) - q$. The resulting hexagon is the region R, and the received message and the error in the action are uniformly distributed over the hexagon. As should be clear, the fact that the loss function is quadratic in $A(\widetilde{m}) - q$ means that an alternative intended message function which moves the mass of R toward the boundaries of $\widetilde{m}$ and away from the extremes at the middle would be superior to the identity.

Definition 1 *Define $m_d(q) \equiv [(N+1)q]/N$, where $[x]$ is the integer part of x.*

$m_d(q)$ is a discrete step function, with $N+1$ steps, each of size $2\bar{\varepsilon}$, with a first step at $m=0$ and an $N+1st$ step at $m=1$.

For clarity, it will often be helpful to refer to the *cost* of an intended message function. We define the cost to be the expected value over q and ε of the negative of the utility, given an intended message function $m(q)$ and the optimal response $A(\widetilde{m})$. The goal, then, is to choose cost-minimizing intended message functions.

Proposition 1 *If f_e is uniform, then the optimal intended message function is $m_d(q)$ for any integer $N > 0$.*

Proof. The cost of an arbitrary intended message function, given the response function $A(\tilde{m}) = \bar{q}(\tilde{m})$, is

$$\int_{q_-(\tilde{m})}^{q_+(\tilde{m})} (q - \bar{q}(\tilde{m}))^2 \, dq = \int_{-w(\tilde{m})/2}^{w(\tilde{m})/2} u^2 \, du = \frac{1}{12} w(\tilde{m})^3.$$

The cost as a function of m is

$$C[m] = \frac{1}{12} \int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} w(\tilde{m})^3 \, d\tilde{m} > \left[\frac{1}{12} \int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} w(\tilde{m})\right]^3 = \frac{2\bar{\varepsilon}}{12},$$

which is attained by the discrete intended message function $m_d(q)$. ■

The intuition is as follows. Since $\int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} w(\tilde{m})\, d\tilde{m} = 2\bar{\varepsilon}$ is the same for every function m, and $w \mapsto w^3$ is a convex function on $[0, \infty)$, Jensen's Inequality establishes that if there exists an intended message function $q \mapsto m(q)$ for which $\tilde{m} \mapsto w(\tilde{m})$ is constant, then this m minimizes C. There is such a function and it is $m_d(q) \equiv [(N+1)q]/N$.

Figure 4 provides graphical intuition. We reproduce the pair of figures in Figure 3, replacing the identity with the step function. For clarity, we assume that $\bar{\varepsilon} = 1/10$, so that there are six steps. The solid line in the left panel plots $m_d(q)$ and the dashed lines represent the upper and lower bounds of the received message, given $m_d(q)$. Note that this function has steps that use the

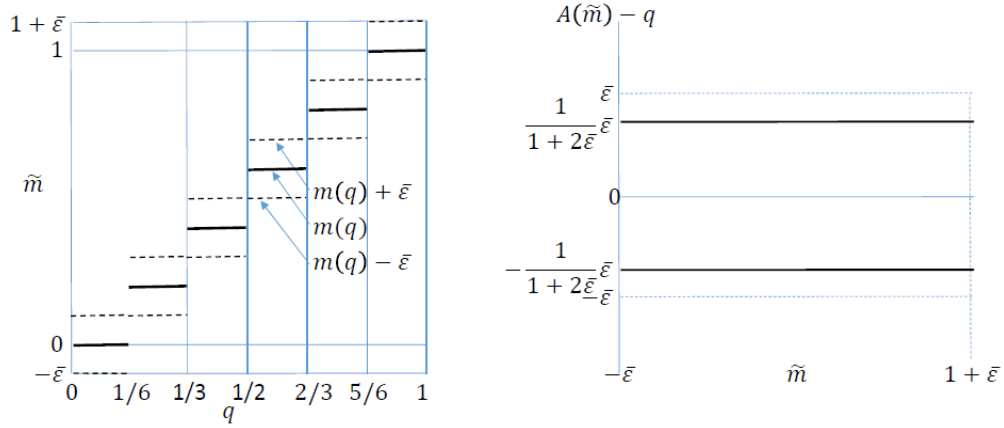

Figure 4: The solid line in the left panel plots the step function $m_d(q)$. The dashed lines represent the upper and lower bounds of the received message, given $m_d(q)$. Note that this function has steps that use the entire message space, from 0 to 1, but the received messages perfectly identify the intended message. They do not overlap, and they do not leave gaps. The right panel uses the same change of variables as Figure 3, and the band between the solid lines delineates the region R, over which the received message and the error in the action are jointly uniformly distributed.

entire message space, from 0 to 1 and the received messages perfectly identify the intended message: they do not overlap, and they do not leave gaps. The right panel uses the same change of variables as Figure 3, and the band between the solid lines delineates the area R, over which the received message and the error in the action are jointly uniformly distributed.

It should be clear that there is no way to redistribute errors over $\widetilde{m}$ without creating errors that are costlier than those created by the step function. The area of $\langle q, \tilde{m}(q)\rangle$ is fixed at $2\bar{\varepsilon}$ and the width is fixed at $1+2\bar{\varepsilon}$. Therefore, the average height is fixed as well, at $\frac{2\bar{\varepsilon}}{1+2\bar{\varepsilon}}$. Because the loss for any given $\tilde{m}$ is quadratically increasing in that height, Jensen's Inequality establishes that the loss-minimizing distribution of $\langle q, \tilde{m}(q)\rangle$ has a fixed height, so long as such a distribution exists. As it turns out, such a distribution does exist, and it is the step function $m_d(q)$. If $\bar{\varepsilon} = 1/10$, as in Figure 4, then the width of each step is $\frac{2\times 1/10}{1+2\times 1/10} = \frac{1}{6}$.

Corollary 1 *If f_e is uniform, then the number of discrete messages in the optimal intended message function is $N+1 = \frac{1}{2\bar{\varepsilon}} + 1$.*

Proof. This follows immediately from the definition of m_d. ■

This establishes that the optimal number of messages from which the sender chooses is driven by the maximum error in interpreting a given message.

For m_d, $w(\tilde{m}) = 1/(N+1)$ for all $\tilde{m}$; and

$$C[m_d] = \frac{1+2\bar{\varepsilon}}{12(N+1)^3} = \frac{1}{12N(N+1)^2}.$$

This may be compared to the cost for the identity intended message function: $m_i(q) = q$, which is a natural alternative intended message function.

Corollary 2 *If f_e is uniform, then the ratio of the cost of the identity intended message function to the cost of the discrete intended message function m_d is $\frac{(N+1)^2(2N-1)}{2N^3}$, which equals 2 when $N = 1$ and is decreasing with N, with limit 1 as $N \to \infty$.*

Proof. In this case (assuming $\bar{\varepsilon} \leq 1/2$)

$$w(\tilde{m}) = \begin{cases} \tilde{m} + \bar{\varepsilon} & \text{if } \tilde{m} \in [-\bar{\varepsilon}, \bar{\varepsilon}] \\ 2\bar{\varepsilon} & \text{if } \tilde{m} \in [\bar{\varepsilon}, 1-\bar{\varepsilon}] \\ 1 + \bar{\varepsilon} - \tilde{m} & \text{if } \tilde{m} \in [1-\bar{\varepsilon}, 1+\bar{\varepsilon}], \end{cases}$$

and

$$C[m_i] = \frac{2}{3}\bar{\varepsilon}^3(1-\bar{\varepsilon}). \tag{3}$$

Letting $\bar{\varepsilon} = 1/2N$,

$$C[m_i] = \frac{2N-1}{24N^4}. \tag{4}$$

The ratio of the respective costs is

$$\frac{C[m_i]}{C[m_d]} = \frac{(N+1)^2(2N-1)}{2N^3}.$$

■

This establishes that the discrete intended message function is most advantageous – relative to a natural alternative, the identity function – when potential errors of interpretation are large. Not surprisingly, as the potential error shrinks, the advantage of mitigating these errors with discrete intended messages shrinks as well.

Proposition 1 establishes that injecting some exogenous garbling into the CS world eliminates the classic CS result that when the sender's and receiver's interests are aligned, any one-to-one function of the state to the message is optimal. We have shown that, regardless of how minute the garbling – i.e., how small is $\bar{\varepsilon}$ – discrete messages can be superior to continuous messages.

1.2 The case of non-uniform errors

It is evident that the preceding analysis depends critically on the fact that q and ε are both uniformly distributed, by virtue of which the distribution of q given $\tilde{m}$ is uniform for any message function m and depends only on the width of the region R at $\tilde{m}$. We will now consider the general situation presented at outset; i.e., we continue to assume that q is uniformly distributed in $[0, 1]$, but we will only suppose that ε has a symmetric continuous distribution in the interval $[-\bar{\varepsilon}, \bar{\varepsilon}]$ with density f_ε.[10] We will see that, when the variance of the error is small, the optimal intended message function is no longer discrete. We do not derive the optimal intended message function, but show that it cannot be discrete.

Proposition 2 *If f_e is not uniform, then the optimal intended message function is not discrete for sufficiently low variance of ε.*

Proof. The distribution of q given $\tilde{m}$ depends in general on f_ε, m, and $\tilde{m}$. If m is a discrete intended message function m_d, however, the distribution of q given $\tilde{m}$ is the same uniform distribution previously analyzed. We can therefore calculate the cost of the discrete intended message function using the same method as above.

We can also calculate the cost of a specific alternative to the discrete function, which is the identity $m(q) = q$. For the identity, the density of the 2-dimensional random variable $\langle q, \tilde{m}(q)\rangle$ is $f_\varepsilon(q - \tilde{m})$, and the cost is

$$C[m_i] = \int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} \int_{q_-(\tilde{m})}^{q_+(\tilde{m})} (q - \bar{q}(\tilde{m}))^2 f_\varepsilon(q - \tilde{m})\, dq\, d\tilde{m}.$$

Let $\delta = q - \tilde{m}$, $\bar{\delta}(\tilde{m}) = \bar{q}(\tilde{m}) - \tilde{m}$, and $\delta_\pm(\tilde{m}) = q_\pm(\tilde{m}) - \tilde{m}$. By definition,

$$\bar{\delta}(\tilde{m}) = \left[\int_{\delta_-(\tilde{m})}^{\delta_+(\tilde{m})} \delta f_\varepsilon(\delta)\, d\delta\right] \times \left[\int_{\delta_-(\tilde{m})}^{\delta_+(\tilde{m})} f_\varepsilon(\delta)\, d\delta\right]^{-1},$$

so[11]

$$C[m_i] = \int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} \int_{\delta_-(\tilde{m})}^{\delta_+(\tilde{m})} \delta^2 f_\varepsilon(\delta)\, d\delta\, d\tilde{m} - \int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} \bar{\delta}(\tilde{m})^2 \int_{\delta_-(\tilde{m})}^{\delta_+(\tilde{m})} f_\varepsilon(\delta)\, d\delta\, d\tilde{m}. \tag{5}$$

Note that

$$[\delta_-(\tilde{m}), \delta_+(\tilde{m})] = \begin{cases} [-\tilde{m}, \bar{\varepsilon}] & \text{if } \tilde{m} \in [-\bar{\varepsilon}, \bar{\varepsilon}] \\ [-\bar{\varepsilon}, \bar{\varepsilon}] & \text{if } \tilde{m} \in [\bar{\varepsilon}, 1 - \bar{\varepsilon}] \\ [-\bar{\varepsilon}, 1 - \tilde{m}] & \text{if } \tilde{m} \in [1 - \bar{\varepsilon}, 1 + \bar{\varepsilon}]. \end{cases}$$

[10] The assumption of continuity is a convenience. The general conclusions apply also to singular distributions, which may be regarded as limits of continuous distributions.

[11] This is essentially the familiar identity $E\left((X - \bar{X})^2\right) = E(X^2) - \bar{X}^2$, adjusted for the mass of the horizontal section of R at $\tilde{m}$.

For $\tilde{m} \in [-\bar{\varepsilon}, \bar{\varepsilon}]$ let $\tilde{m}' = 1 + \tilde{m}$. Then $\{[\delta_-(\tilde{m}), \delta_+(\tilde{m})], [\delta_-(\tilde{m}'), \delta_+(\tilde{m}')]\}$ is a partition of $[-\bar{\varepsilon}, \bar{\varepsilon}]$ (for measure-theoretic purposes, the intersection being a singleton). We may therefore combine the corresponding contributions to the first integral in (5) to obtain

$$\begin{aligned} C[m_i] &= \int_{-\bar{\varepsilon}}^{1-\bar{\varepsilon}} \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \delta^2 f_\varepsilon(\delta)\, d\delta\, d\tilde{m} - \int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} \bar{\delta}(\tilde{m})^2 \int_{\delta_-(\tilde{m})}^{\delta_+(\tilde{m})} f_\varepsilon(\delta)\, d\delta\, d\tilde{m} \\ &= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \delta^2 f_\varepsilon(\delta)\, d\delta - \int_{-\bar{\varepsilon}}^{1+\bar{\varepsilon}} \bar{\delta}(\tilde{m})^2 \int_{\delta_-(\tilde{m})}^{\delta_+(\tilde{m})} f_\varepsilon(\delta)\, d\delta\, d\tilde{m} \\ &\leq \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \delta^2 f_\varepsilon(\delta)\, d\delta. \end{aligned}$$

As this inequality indicates, we are not particularly interested in the second integral, but it is worth noting that by virtue of the symmetry of f_ε, $\delta(\tilde{m}) = 0$ for $\tilde{m} \in [\bar{\varepsilon}, 1 - \bar{\varepsilon}]$, so the contribution from this interval of $\tilde{m}$ is 0.

The cost $C[m_i]$ is therefore bounded above by the variance of ε. For the uniform distribution this variance is $\frac{2}{3}\bar{\varepsilon}^3$, which may be compared with the exact computation of $C[m_i]$ given above in Equation(3). If the variance of ε is sufficiently small compared to that of the uniform distribution on $[-\bar{\varepsilon}, \bar{\varepsilon}]$ a discrete intended message function cannot be more efficient than m_i. ■

Some intuition for this result is provided in Figure 5. We reproduce the right panels of Figures 3 and 4 in order to show which errors are more and less likely for each intended message function. Both the rectangle and the hexagon have area $2\bar{\varepsilon}$. Larger errors in the action occur in the thatched trapezoids in the center when the identity function is used rather than the discrete function. These are large and costly errors. Larger errors in the action occur in the dotted triangles at the boundaries when the discrete function is used rather than the identity. These are smaller and less costly errors. This is why, when the distribution of ε is uniform, the discrete function is less costly.

When the distribution of ε is not uniform, the supports of the received messages and errors in action do not change, but the distributions are no longer uniform over these supports. As the variance of ε falls, the mass of errors in action that lies in the trapezoids falls relative to the mass of errors in action that lies in the triangles, leading the identity to eventually be superior to the discrete function.

2 Conclusion

It is well known that misaligned interests of the sender and receiver lead to discrete messaging. This has led many to assume that, when we observe discrete messages, it is likely that the sender's and receiver's preferences are misaligned. At least two strands of literature have developed alternative rationales for discrete messages: more coarseness may be less costly, or there may be divergent preferences among multiple receivers. We have shown that an intuitive alternative assumption

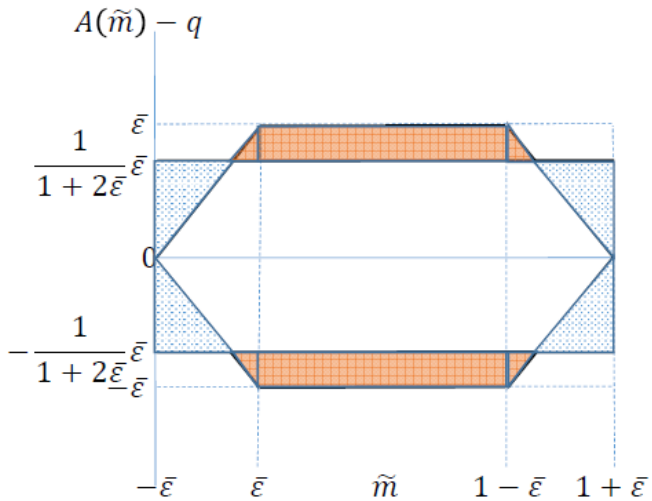

Figure 5: We reproduce the right panels of Figures 3 and 4 in order to show which errors are more and less likely for each intended message function. Both the rectangle and the hexagon have area $2\bar{\varepsilon}$. Larger errors in the action occur in the thatched trapezoids in the center when the identity intended message function is used rather than the discrete function. These are large and costly errors. Larger errors in the action occur in the dotted triangles at the boundaries when the discrete function is used rather than the identity. These are smaller and less costly errors. When the distribution of ε is not uniform, the bounds of the received messages and errors in action do not change, but the distribution is no longer uniform over these areas. As the variance of ε falls, the mass of errors in action that falls in the trapezoids falls relative to the mass of errors in action that fall in the triangles, leading the identity to eventually be superior to the discrete function.

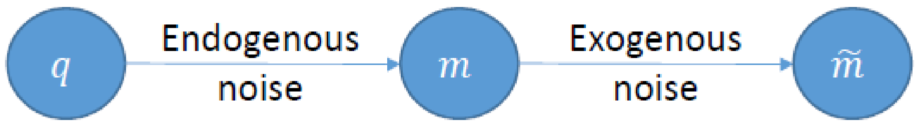

Figure 6: There are two sources of noise. Endogenous noise has been well-studied since Crawford and Sobel (1982), and is created by senders with continuous information sending discrete messages. Exogenous noise is introduced in this manuscript as a difficulty for receivers in interpreting messages. Sufficiently coarse messages create endogenous noise but eliminate exogenous noise. This can be advantageous if the variance of the exogenous noise is high enough.

– that messages are often misinterpreted – can also generate discrete messaging. What is the intuition? There are two types of error in this model. First is the endogenous error of CS: how easily can the receiver infer the true state, q, assuming that the message, m, is known? In models based off of the CS framework, this endogenous error is omnipresent. Second is the exogenous error of our model: how easily can the receiver infer the intent of the message, m, from the message received, $\widetilde{m}$? These types of error are illustrated in Figure 6.

A discrete intended message function can eliminate the second type of error at the expense of creating the first. *Discrete messages are less precise, but they are easier to interpret.* This intuition is consistent with the quote from a Yelp spokesman presented in the introduction: "We wanted to convey something that was easy for our audience to understand, and so if you're thinking of star ratings for a business on a scale of one to five, it's pretty clear-cut and dry."

We allow the size of the exogenous noise to vary in two ways. First, we vary the maximum error, $\bar{\varepsilon}$. As this error decreases, the level of exogenous noise decreases. This does not, however, change the fact that the optimal intended message function can be discrete. A lower maximum error both reduces the incentive to introduce endogenous noise, which is the point of discretization, and increases the maximum number of useful messages in a discrete message function, which increases the value of discretization. These offset.

Second, we allow for alternative distributions of the error within the support $[-\bar{\varepsilon}, \bar{\varepsilon}]$, and show that, when the variance of the error is small enough, discrete intended messages cannot be optimal. When the variance of the error is small, discrete intended message functions "waste" too much of the message space on infrequent events. It is better to be less easily interpreted in order to have more precision in the message.

This model does not require any actors to have differing preferences and does not place any explicit cost on finer messaging. It is equally costly to use the intended message function $m(q) = q$ as to use the intended message function $m(q) = 1$. Messages are received with error, but that error is independent of the message sent. It is therefore surprising that it can be optimal to discard information at the messaging stage.

We note that some readers might find the assumption of exogenous errors unnatural. In order to keep the paper short and to-the-point, we do not micro-found these errors. We do not believe that it would be difficult to do so. For example, we could suppose that the sender and receiver could exert costly effort to shrink the magnitude of the error: by working hard, the sender could make herself more clear and the receiver could better understand the sender's point. We might assume that the maximum error $\bar{\varepsilon}(e_S + e_R) = \bar{\varepsilon}(0)/(e_S + e_R)$, where $e_S, e_R > 0$ are the sender's and receiver's efforts, and $\bar{\varepsilon}(0)$ is the maximum error when no effort is undertaken. We might also assume that the cost of effort for $i \in \{S, R\}$ is the identity $c(e_i) = e_i$. Under these assumptions, there will be an optimal level of effort for both sender and receiver and, consequently, an optimal level of $\bar{\varepsilon}$. This value of $\bar{\varepsilon}$ would be associated with an optimal intended message function which could be discrete or continuous. This specific microfoundation is not, to our minds, particularly interesting, but it may help readers who are concerned that exogenous errors are not well founded.

We conclude with a brief discussion of applications where this idea may apply. Most obvious are those applications on which we have already focused: customer reviews, professional reviews, corporate credit ratings, student grades, etc. The idea also clearly applies to settings like within-firm communication, in which discrete (i) commands to subordinates, (ii) reports to superiors, or (iii) delivery of information between divisions could be optimal.

As the sender and receiver share preferences, this intuition should also apply to a single-agent setting. When we perform statistical tests, we convert the underlying (and unknowably complex) data into simple summary information, whose interpretation is useful but almost certainly not perfect. The statistical test can be thought of as the sender's intended message $m(q)$. The econometrician is the receiver and chooses the test $m(q)$ in order to have the most precise knowledge possible, perhaps to undertake an action (like make a policy) which she presumably wants to best reflect the state. Does the econometrician want continuous information, like regression coefficients, or does she want simple discrete information? In practice, the answer is often discrete. For example, pollsters design polls with few responses. Even after issuing polls with, say, four responses (strongly agree, agree, disagree, strongly disagree), results are often presented in only two "buckets": agree and disagree. This discretizing is clearly less precise, but perhaps easier to interpret.

3 References

Agranov, Marina, and Andrew Schotter (2012). Ignorance is Bliss: An Experimental Study of the Use of Ambiguity and Vagueness in the Coordination Games with Asymmetric Payoffs. *American Economic Journal: Microeconomics* 4 (2), pp 77-103.

Anderson, Michael, and Jeremy Magruder (2012). Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database. *The Economic Journal* 122 (563), pp 957-989.

Blume, Andreas, Oliver Board, and Kohei Kawamura (2007). Noisy Talk. *Theoretical Economics* 2, pp 395-440.

Crawford, Vincent, Uri Gneezy, and Yuval Rottenstreich (2008). The Power of Focal Points is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures. *American Economic Review* 98 (4), pp 1443-1458.

Crawford, Vincent, and Joel Sobel (1982). Strategic Information Transmission. *Econometrica* 50 (6), pp 1431-1451.

Cremer, Jacques, Luis Garicano, and Andrea Prat (2007). Language and the Theory of the Firm. *Quarterly Journal of Economics* 122 (1), pp 373-407.

Garicano, Luis, and Andrea Prat (2011). Organizational Economics with Cognitive Costs. Working paper.

Luca, Michael (2016). Reviews, Reputation, and Revenue: The Case of Yelp.com. *Harvard Business School Working Paper*, No. 12-016, September 2011. Revised March 2016.

Myerson, Roger B (1991). Game Theory: Analysis of Conflict. Harvard University Press, Cambridge, Massachusetts.

Sobel, Joel (2012). Complexity versus Conflict in Communication. Information Sciences and Systems (CISS), 46th Annual Conference.

Sobel, Joel (2013). Giving and Receiving Advice. *Advances in Economics and Econometrics: Tenth World Congress*, Volume 3, Chapter 10.