

Coarse Thinking and Persuasion
Sendhil Mullainathan and Andrei Shleifer¹
Revised, September 2006

1. Introduction

In most societies, a tremendous amount of resources is devoted to persuasion (McCloskey and Klamer 1995). Selling, advertising, political campaigns, organized religion, law, much of the media, and good parts of education are devoted to changing beliefs in a way advantageous to the persuader. Moreover, persuasion is not simply about expending resources: the exact content of the message crucially shapes its effectiveness. But what constitutes persuasive content?

Economists typically assume that only one type of persuasive content matters: objectively useful information. Stigler (1987, p. 243), for example, writes that “advertising may be defined as the provision of information about the availability and quality of a commodity.” Economists have typically modeled persuasion, including advertising (Stigler 1961), political campaigns (Downs 1957), or legal argument (Milgrom and Roberts 1986, Dewatripont and Tirole 1999) as provision of information. In some frameworks, such as Grossman (1981), Grossman and Hart (1980), Milgrom (1981), and Okuno-Fujiwara et al. (1990), the persuader uses information strategically, but he conveys information nonetheless².

Psychologists understand persuasion quite differently. They argue that people evaluate various propositions or objects using representativeness, categorization,

¹ Department of Economics, Harvard University. We are grateful to Nicholas Barberis, Daniel Benjamin, Olivier Blanchard, Matthew Gentzkow, Simon Gervais, Richard Holden, David Laibson, Jesse Shapiro, and especially Nicola Gennaioli and Giacomo Ponzetto for helpful comments, and to Josh Schwartzstein for excellent research assistance. The errors remain our own.

² Recent years saw a revival of research on communication and persuasion from the information-theoretic perspective. See, e.g., Glazer and Rubinstein (2001, 2004).

metaphors, analogies, and other “boundedly rational” strategies (see, e.g., Tversky and Kahneman 1982, Lakoff 1987, Edelman 1992, Zaltman 1997, 2003). Much of this research has proved difficult to incorporate into economics. In this paper, we try. We present an economic model of inference and persuasion suggested by some of the psychological work. Our model clarifies how non-informative persuasion operates and describes its limits. We also apply the model to advertisement of mutual funds.

To motivate our analysis, consider four persuasive messages:

Alberto Culver Natural Silk Shampoo was advertised with a slogan “We put silk in the bottle.” The shampoo actually contained some silk. Nonetheless, during the campaign, the company spokesman conceded that “silk doesn’t really do anything for hair” (Carpenter, Glazer and Nakamoto 1994).

A successful television and print advertising campaign for American Express Travelers Checks featured Karl Malden, an actor who played a detective on the television show “The Streets of San Francisco.” The ads advised the audience that, for safety, they should carry American Express Travelers Checks rather than cash.

After the crash of the internet bubble, the brokerage firm Charles Schwab launched an advertising campaign that featured judges, ship captains, crossing guards, doctors, grandmothers, and other steady and reliable individuals, each standing next to a Schwab investment specialist. The headline of one ad, depicting a Schwab professional standing next to a pediatrician, stated that “Both are seen as pillars of trust.”

As an example from political persuasion, consider Arnold Schwarzenegger’s memorable speech at the 2004 Republican National Convention. In the best remembered part of his speech, Schwarzenegger defended free trade: “To those critics who are so

pessimistic about our economy, I say: Don't be economic girlie men!... Now they say India and China are overtaking us. Don't you believe it. We may hit a few bumps – but America always moves ahead. That's what Americans do.”

What do these persuasive messages have in common? They all appear to have the feature that a piece of information that has some objective value in one context gets transferred by the persuader into another context, where it has limited or no value. A piece of data that might carry objective weight in one situation is brought to bear on another, where it is uninformative.

In the Alberto Culver example, the advertiser and his audience know that silky hair is attractive. And although silk does nothing for hair, this ad rests on the audience's perception that it will make hair attractive. The audience is expected to apply the information (silkeness) useful in one context (hair) to another context (shampoo), where it is useless.

In the American Express example, the audience, many of whose members have watched “The Streets of San Francisco,” know that Karl Malden keeps people safe. The ads seek to convince them that the travelers' checks recommended by Karl Malden will do likewise. The ad is based on the idea that the streets of San Francisco and travelers checks are both things Malden makes safe. The information (Karl Malden is involved) that is useful in one context (an episode of a television show) for predicting safety is being transferred to another context (travelers checks), where it is objectively useless. For how can it matter objectively which actor endorses travelers' checks?

The Schwab ads take advantage of the belief that people in some positions – judges, pediatricians, grandmothers – are both trustworthy and wise. The ads hope that

this information (trustworthiness), which from the general experience of the readers is a good description of some types of professionals (judges and pediatricians), rubs off on its investment professionals as well, where the proposition is not as self-evident.

Schwarzenegger taps into a common conceptualization of foreign trade as competition. His audience undoubtedly finds the information (America always moves ahead) as plausible in the sports context, especially coming from a former champion. Accepting that trade is a sport, Schwarzenegger invites the audience to transfer the information that Americans typically win into that context. Of course, both the sports analogy and the winning message are largely irrelevant for the assessment of whether free trade, outsourcing, or globalization is beneficial to the public.

Schwarzenegger's speech was not accidental. In his advice to Congressional candidates, the Republican political strategist Frank Luntz (2006) writes: "Never, never, never begin a response to outsourcing by saying it is beneficial to the U.S. economy. Never... Don't talk like economists. Words like 'protectionist,' 'capitalist,' and 'isolationist' turn the average voter off. The key word is winning. It is essential that you capture the theme of winning and insert it into all your communications efforts."

The transfer of data that is informative in one context into another, where it might not be, is obviously related to associational, analogical, or metaphoric thinking, all of which have been discussed by psychologists (e.g., Lakoff 1987, Zaltman 1997). Importantly, such thinking is both common and extremely useful in everyday life because it reduces the evaluation of new situations to comparison with familiar ones. Some writers think that, for this reason, our brains have evolved so as to make metaphor and analogy standard hard-wired forms of reasoning (Edelman 1992). Of course, the patterns

of thought that are *usually* extremely helpful are not *always* so. We suggest that persuaders take advantage of people utilizing a strategy which, though useful in “similar” contexts, is not useful in the situation of interest to the persuader.

To this end, we present a model of persuasion that deviates from Bayesian rationality in one specific way. We assume that people put situations into categories, and failing to differentiate between co-categorized situations use one model for all the situations in the same category. We refer to such inference as coarse thinking. Coarse thinking implies transference: the impact of a message in one situation depends on how that message is interpreted in other co-categorized situations. Persuasion takes advantage of coarse thinking.

The model yields several implications. First, it shows how messages that are non-informative in a given situation nonetheless affect beliefs, and delineates the contexts in which they do so more effectively. Second, it demonstrates the effects of increasing the sophistication of the audience on the efficacy of non-informative persuasion. The recognition by the audience that the persuader is acting strategically reduces, but does not eliminate, the scope for non-informative persuasion. In fact, with coarse thinking, the persuader can get away with concealing payoff-relevant information without his audience assuming the worst. This result gets us away from the full revelation conclusions of, for example, Grossman-Hart (1980) or Milgrom (1981). On the other hand, if the audience is fully Bayesian and treats each situation differentially – does not engage on coarse thinking – the possibility of uninformative persuasion disappears, at least in our model.

Economists have struggled with the realization that much of persuasion is not informative for many years. A perceptive early contribution is Tullock (1967). Posner

(1995) draws a sharp distinction between persuasion (rhetoric) and provision of information, and discusses Aristotle's and Plato's ideas on the subject. More recent research includes studies of hatred (Glaeser 2005), media (Mullainathan and Shleifer 2002, 2005, Gentzkow and Shapiro 2006), and political persuasion (Becker 2001, Murphy and Shleifer 2004, Glaeser, Ponzetto, and Shapiro 2005). Stigler and Becker (1977) and Becker and Murphy (1993) examine models in which advertising enters the utility function directly. As far as we know, however, our paper is the first to study persuasion in a model of categorical thinking that seems so central to psychological work.

We develop our analysis in three steps. In Section 2, we present a model of inference and persuasion for both Bayesian and coarse thinkers, and show how it can account for the examples we have described, as well as make a range of additional predictions. In that section, we assume that the audience takes information at face value and does not consider that the persuader is acting strategically. In Section 3, we show that the results, with some caveats, extend to the case of a sophisticated audience that recognizes that the persuader is acting strategically. In Section 4, we present an extended application of the basic model to the advertisement of mutual funds, and show that the stylized facts about such advertisement are consistent with the predictions of the model. Section 5 concludes.

2. A Simple Model

Basic setup

An individual must assess the quality of a given object. He could be assessing the quality of a shampoo, the expected net return on a mutual fund, or the

appeal of a political candidate. We denote this underlying quality of the object by q . The individual receives a potentially informative message \mathbf{m} about q . He then uses this message to form expectations about underlying quality. We assume that this quality is the sum of n components $q = q_1 + q_2 + \dots + q_n$. The individual components are real numbers $q_i \in (-\infty, \infty)$, so the aggregate is also a real number $q \in (-\infty, \infty)$. Denote the vector of individual components by \mathbf{q} , so $\mathbf{q} = (q_1, q_2, \dots, q_n)$. The message \mathbf{m} is a vector of dimension n , (m_1, \dots, m_n) . Each m_i conveys information about the sub-component q_i of quality. Each m_i is drawn from a compact set $M_i \subset (-\infty, \infty)$ or equals “empty” (\emptyset), indicating that there is no data on that dimension.

A key feature of our model is that the individual faces many similar, but not identical, situations s drawn from a finite set S . Messages are informative in some situations in S but not others. Absent persuasion, messages are distributed according to the density function $p(\mathbf{m}, \mathbf{q}, s)$. Crucially, we assume that distributions are independent across dimensions given each situation. In this case,

Assumption 1:

$$p(\mathbf{m}, \mathbf{q} \mid s) = p_1(m_1, q_1 \mid s) p_2(m_2, q_2 \mid s) \cdots p_n(m_n, q_n \mid s) \quad (1)$$

where $p_i(m_i, q_i \mid s)$ denotes the marginal distribution on dimension i given situation s . This marginal distribution is the primitive that defines the joint probability of seeing a given message on some dimension for a given quality on that dimension. We assume that the distributions are continuous and atomless for $m_i \neq \emptyset$ and that

$$\int_{m_i \in M_i} p_i(m_i, q_i \mid s) dm_i = 1 - p_i(m_i = \emptyset, q_i \mid s) \quad (2)$$

for all s . The distribution of \mathbf{m} , $p(\mathbf{m}, \mathbf{q}/s)$, depends on the underlying quality because, for messages to be informative, certain data must be more likely for certain quality levels. The distribution also depends on the situation because different situations require different models for interpreting messages. For simplicity, we write $p(\mathbf{q})$, $p(\mathbf{m})$ and $p(s)$ to denote the marginal distributions:

$$p(s) = \int \int_{\mathbf{m} \ \mathbf{q}} p(\mathbf{m}, \mathbf{q}, s) d\mathbf{q} d\mathbf{m} + \int_{\mathbf{q}} p(\emptyset, \mathbf{q}, s) d\mathbf{q} \quad (3)$$

$$p(\mathbf{q}) = \sum_{s \in S} \int_{\mathbf{m}} p(\mathbf{m}, \mathbf{q}, s) d\mathbf{m} + \sum_{s \in S} p(\emptyset, \mathbf{q}, s) \quad (4)$$

$$p(\mathbf{m}) = \sum_{s \in S} \int_{\mathbf{q}} p(\mathbf{m}, \mathbf{q}, s) d\mathbf{q} \quad (5)$$

We assume that the expected quality on each dimension given each situation is zero:

Assumption 2:

$$\begin{aligned} E(q_i | s) &= \int_{q_i} \left[\int_{m_i} q_i p_i(m_i, q_i | s) dm_i + q_i p_i(\emptyset, q_i | s) \right] dq_i \\ &= \int_{q_i} q_i p_i(q_i | s) dq_i \\ &= 0 \end{aligned} \quad (6)$$

for all s . We further assume that the empty message \emptyset is in general uninformative, so that

Assumption 3:

$$p_i(q_i | \emptyset, s) = \frac{p_i(\emptyset, q_i | s)}{\int_{q_i} p_i(\emptyset, q_i | s) dq_i} = p_i(q_i | s) \quad (7)$$

for all q_i, s .

A Bayesian receiving a message m in situation s updates according to the informativeness of that message in that situation. Specifically, a Bayesian's expectation of quality on a particular dimension given a message m is:

$$E^B(q_i | \mathbf{m}, s) = E^B(q_i | m_i, s) = \int_{q_i} q_i p_i(q_i | m_i, s) dq_i \equiv f_i(m_i, s) \quad (8)$$

The assumption of independence of the distribution of messages across dimensions allows us to define this expectation as a function solely of m_i . Moreover, we can write the Bayesian expectation of overall quality q as:

$$E^B(q | \mathbf{m}, s) = \sum_i E^B(q_i | m_i, s) = \sum_i f_i(m_i, s) \equiv f(\mathbf{m}, s) \quad (9)$$

From Assumption 3, it follows that $f_i(\emptyset, s) = 0$, so that an *exogenous* absence of a signal on a particular dimension contains no information about quality. We define a dimension i to be *uninformative* in situation s if $f_i(m_i, s) = 0$ for all m_i . Otherwise, we refer to a dimension as *informative*. From Assumption 2, it follows that for all that informative dimensions there exist m_i' and m_i'' such that $f_i(m_i', s) > 0$ and $f_i(m_i'', s) < 0$.

The “silk in the bottle” example from the introduction can be understood in this setup. One situation would be a consumer looking for shampoo, where the message would be “We put silk in the bottle.” In this situation, $m = \text{“silk”}$ and $s = \text{“shampoo.”}$ In another situation, this consumer might encounter a message “Her hair is silky.” In that situation, $m = \text{“silk”}$ and $s = \text{“hair.”}$ In this example, the same message is uninformative in the first situation but informative in the second.

A Bayesian thinker in this model is able to distinguish among the different situations s he faces, and therefore to update his beliefs correctly for each situation. In the silk example, a Bayesian uses a different model for each situation, and therefore correctly infers that silk raises the probability that hair is of high quality, but does not change that probability for a shampoo. Such a Bayesian simply ignores the shampoo ad.

Coarse Beliefs

The essential assumption in our model is that individuals do not have separate mental representations for every situation. Instead, they have only one representation for all the situations in a category and are effectively unable to differentiate between these situations. For example, they might interpret silkiness in a similar way for shampoo as for hair, or think of winning in the same way in sports as in foreign trade. Categorical thinking has been previously modeled by Mullainathan (2000), Mullainathan and Shleifer (2002), Fryer and Jackson (2004), and Peski (2006); our framework relies most closely on the ideas in Mullainathan (2000).

Specifically, we assume that individuals group situations together into categories. Define $c(s)$ to be a map from the set of situations S to the set of categories C . Define $C(s)$ to be the set of situations that are grouped together with the situation s , i.e.

$$C(s) = \{s' \mid c(s) = c(s')\} \quad (10)$$

Define $p(C)$ to be the sum of the probabilities of all situations in category C :

$$p(C) = \sum_{s \in C} p(s) \quad (11)$$

We assume that, upon receiving a message on a given dimension, the coarse thinker updates in the following fashion:

Assumption 4:

$$p_i(q_i | m_i, C) = \sum_{s' \in C} p_i(q_i | m_i, s') \frac{p(s')}{p(C)} \quad (12)$$

Coarse thinking is characterized by a different update function than that of a Bayesian. Individuals respond not to the actual informativeness of a message for the current situation s , but rather to its average informativeness across all situations categorized together with s . We are further assuming that a coarse thinker, by our definition, is not simply a Bayesian who observes C rather than the specific situation s . He would derive the joint distribution $p_i(m_i, q_i, C(s)) = \sum_{s' \in C(s)} p_i(m_i, q_i, s')$. From this, he would generate the conditional distribution:

$$p_i(q_i | m_i, C(s)) = \sum_{s' \in C(s)} p_i(q_i | m_i, s') \frac{p_i(m_i | s') p(s')}{p_i(m_i | C(s)) p(C(s))}$$

Thus this type of Bayesian, upon observing a message m_i on dimension i , still responds to a weighted average of the informativeness of this message across co-categorized situations. However, instead of just weighing the informativeness of the message in situation s' by the likelihood of that situation in the category $\left(\frac{p(s')}{p(C(s))} \right)$, this Bayesian

would take into account how likely it is that the message would appear in that situation $\left(\frac{p(m | s')}{p(m | C(s))} \frac{p(s')}{p(C(s))} \right)$. We have chosen the coarse updating rule we have because we

feel it better captures the notion of the coarse thinker not having a distinct model for each

underlying situations. This alternate Bayesian rule requires an understanding of the marginal distributions of each message in each situation.

As a result of the coarse updating, upon receiving a message, the coarse thinker's updated expectation is:

$$\begin{aligned}
E^C(q | \mathbf{m}, s) &= \sum_{i=1}^n E^C(q_i | m_i, s) & (13)^3 \\
&= \sum_{i=1}^n \int_{q_i} q_i p_i(q_i | m_i, C(s)) dq_i \\
&= \sum_{i=1}^n \int_{q_i} \sum_{s' \in C(s)} q_i p_i(q_i | m_i, s') \frac{p(s')}{p(C(s))} dq_i \\
&= \sum_{i=1}^n \sum_{s' \in C(s)} f_i(m_i, s') \frac{p(s')}{p(C(s))}
\end{aligned}$$

From this we can define:

$$k_i(m_i, s) \equiv \sum_{s' \in C(s)} f_i(m_i, s') \frac{p(s')}{p(C(s))} \quad (14)$$

so that

$$E^C(q | \mathbf{m}, s) = \sum_{i=1}^n E(q_i | m_i, s) = \sum_{i=1}^n k_i(m_i, s) \equiv k(\mathbf{m}, s) \quad (15)$$

We can also rewrite $k(\mathbf{m}, s)$ as an intuitive recombination of $f(\mathbf{m}, s)$:

$$k(\mathbf{m}, s) = \sum_{i=1}^n \sum_{s' \in C(s)} f_i(m_i, s') \frac{p(s')}{p(C(s))} = \sum_{s' \in C(s)} \sum_{i=1}^n f_i(m_i, s') \frac{p(s')}{p(C(s))} = \sum_{s' \in C(s)} f(\mathbf{m}, s') \frac{p(s')}{p(C(s))} \quad (16)$$

Coarseness implies that updating happens as if the person cannot distinguish between the various situations in a particular category. When individuals form categories that include only one situation, their behavior is the same as that of a Bayesian. Note that categorical thinking is **not** equivalent to Bayesian thinking when a Bayesian is given information about the category C rather than the situation s . A Bayesian with limited

³ In contrast, a Bayesian who observes categories but not specific situations updates according to

$$E^{BC}(q | \mathbf{m}, s) = \sum_{i=1}^n \sum_{s' \in C(s)} f_i(m_i, s') \frac{p_i(m_i | s') p(s')}{p_i(m_i | C(s)) p(C(s))}.$$

information about the situation would instead attempt to use the specific message m received to also make an inference about which situation s he is in. In contrast, the categorical thinker confuses the situations s and acts as if the different situations are not distinct. He simply does not have a different model for different situations.

As shown by Mullainathan (2000), this model produces empirically plausible patterns of under- and over-reaction. By comparing the updating rules, we can see which messages are under- or over-responded to by a coarse thinker relative to the Bayesian one. To do this, it is useful to decompose $k(\mathbf{m},s)$ as follows:

$$k(\mathbf{m},s) = f(\mathbf{m},s) \frac{p(s)}{p(C)} + \sum_{s' \in C(s), s' \neq s} f(\mathbf{m},s') \frac{p(s')}{p(C)} \quad (17)$$

Comparing this to the Bayesian's response $f(\mathbf{m},s)$ suggests two distortions. First, since $p(s) < p(C(s))$, the response of the coarse thinker only mutedly depends on how informative that message m is for situation s itself, as measured by $f(\mathbf{m},s)$. This effect can lead to an under-reaction to news. Take, for example, the case where $f(\mathbf{m},s) > 0$ and $f(\mathbf{m},s')=0$ for all s' other than s . We see here that $k(\mathbf{m},s) = f(\mathbf{m},s)(p(s)/p(C)) < f(\mathbf{m},s)$. Coarse thinking leads to an under-response to news relative to the Bayesian. The uninformativeness of the message in other situations dilutes its impact in situation s , precisely because $p(s) < p(C(s))$.

Second, the response of the coarse thinker also depends on a term that the Bayesian's response does not depend on: the informativeness of the message in other situations s' in the same category. This implies that the coarse thinker could react to non-information or over-react to information. Take the case, for example, where $f(\mathbf{m},s)=0$, so the message is uninformative in situation s , but is informative in other situations in the

same category ($f(\mathbf{m}, s') > 0$ for all other s' in $C(s)$). Then, $k(\mathbf{m}, s) =$

$$\sum_{s' \in C(s), s' \neq s} f(\mathbf{m}, s') \frac{p(s')}{p(C)} > 0 = f(\mathbf{m}, s).$$

The coarse thinker is now reacting to a piece of data uninformative in situation s because it is informative in other situations in the same category. His use of the same model to interpret messages in all situations leads him to over-react to non-informative data. This, we submit, is exactly what a person responding to silk in shampoo, or to Karl Malden’s travelers’ checks, is doing. This transference of information across situations drives our results below. Indeed, the strategy of persuasion is to trigger such transference: successful persuasion in our model takes advantage of over-reaction.

The perspective on over-reaction used in this model is similar to that used in some recent research in finance, which discusses stock price movements in response to non-information. Daniel and Titman (2006) suggest that over-pricing of some growth stocks is best understood as an over-reaction to non-fundamental “intangible” information, which is subsequently reversed. Analytically, our view of over-reaction as an unjustified response to news about items within the same category is distinct from the interpretation of over-reaction in Mullainathan (2000) and Mullainathan-Shleifer (2002), which look at it in terms of category change. The perspective here is much closer to the finance perspective, which talks about stock price moves unjustified by fundamental news.

As this discussion makes clear, the results in our model rely crucially on assumptions about which situations belong together in a category. This raises a question: what determines categories? In our formal model, we take such categorizations to be exogenous. People already bear in mind that judges and grandmothers are trustworthy. Indeed, the perspective that people already have systems of mental categories and

associations that the advertiser needs to tap into is not unusual (see, e.g., Zaltman 1997, Rapaille 2006). More generally, however, as Lakoff (1987) illustrates with dozens of examples, categories are not fixed, but malleable, over-lapping, varying across cultures and over time, and often created through persuasion itself. Mullainathan and Shleifer (2006) suggest along these lines that individuals co-categorize investing with other activities that make you rich and free *to* do things when stock prices are rising rapidly, but with protection and activities that keep you free *from* adversity in more normal circumstances. Moreover, some members of categories are more central (prototypical) than others, so agents who reason categorically do not treat them symmetrically. Football is a more prototypical example of a competitive sport than international trade. These are important issues, but we abstract away from them in the model.

Before turning to the analysis, we make three remarks about the interpretation of the model. First, the level of coarseness is an important variable. Individuals tend to differentiate more and use more refined categories when they are more engaged in a situation. In this case, many of our results will be mediated by a person's level of involvement. The more involved is the decision maker, the closer is his inference to that of a Bayesian in our model. To the extent that firms can influence the level of involvement, this can be a useful choice variable to consider in future work⁴.

There is a close relationship between coarse thinking and what marketing specialists call *low involvement* decisions. Roughly speaking, low involvement decisions are those that consumers do not think hard about. These might be choices among similar consumer products, or even among political candidates, where the effect on an

⁴ The observation that firms might want to keep the level of involvement low relates to the work of Gabaix and Laibson (2006) on shrouded attributes.

individual's welfare is small. (Low involvement is clearly related to low stakes, although some research shows that even such high stakes decisions as portfolio allocation are often made with minimal involvement; Bernartzi and Thaler 2006.) The marketing literature sees persuasive advertising as particularly effective for low involvement decisions, whereas high involvement decisions require informative advertising (e.g., Sutherland and Sylvester 2000). This observation in fact emerges from our model, since a full Bayesian is not vulnerable to non-informative persuasion. Our point is *not* the view that people are incapable of rational thinking, but rather that in many instances they do not engage in such thinking, perhaps because it is not worth it. It is precisely in those instances of coarse thinking that persuasion pays.

Second, the message itself may affect categorization. Take, for example, the famous advertising campaign run by the Avis Car Rental Company for several decades. Avis, which was for decades second to Hertz in the car rental business, advertised itself as "We're number 2. We try harder." In some respects, this ad fits perfectly into our framework, as it invites the audience to apply the information helpful in some situations (e.g., evaluating athletes) to another situation (evaluating Avis), where the information might be useless. In other respects, however, the example does not fit the model, because the co-categorization of rental car choice with other situations where the underdog is a good bet is entirely the invention of the persuader. This co-categorization is created by the message, which is the persuader's choice. This type of effect, though surely present (particularly in creative advertising), is beyond the scope of our model⁵. We focus on situations where the persuader is powerless to create new categories or to re-allocate his

⁵ Along these lines, Aragonés et al (2001) study persuasion as re-organization of available data. In our model, if messages create categories then the set of co-categorized situations becomes $C(s,m)$ rather than $C(s)$, and the analysis becomes correspondingly more complex.

products through the message to new categories. In other words, we take the category structure as given and explore how messages can harness this structure.

Third, in modeling inference as a particular form of a deviation from Bayesian rationality, we avoid the critical point that much of persuasion is partly or entirely emotional. Indeed, writers such as Zaltman (1997) think of associational inference, which is closely related to co-categorization, as largely emotional. We agree with this observation, but nonetheless offer a model using economists' standard toolbox.

Persuasion

The persuader maximizes the individual's expected quality, net of persuasion costs. Presumably, the higher the expected quality, the greater the support the persuader receives, whether through sales, votes, or membership in his organization.⁶ The persuader can alter messages. Specifically we assume that the persuader observes the signal m that the individual would see absent intervention (for example, if the shampoo appears on a store shelf without advertising). He can then intervene prior to the individual observing m , and create an altered message m' . The individual then observes m' and never sees the original m .

We assume that the persuader can alter messages in two ways. First, he can costlessly "drop" information on any particular dimension i by replacing message m_i with the empty message \emptyset . Second, he can manipulate the content of messages. To capture the idea that persuaders cannot lie about easily verifiable facts without consequence, we make the stark assumption that messages on *non-informative* dimensions (i.e., those i for

⁶ In a Bayesian framework, Holden and Kamenica (2006) examine persuasion where the objective is to change the consumer's decision rather than his beliefs about the state of the world.

which $f_i(m_i, s) = 0$ for all m_i) can be manipulated completely, while messages on *informative* dimensions can only be dropped. For any message on a non-informative dimension, by paying a cost c , the persuader can replace it with anything else. An advertiser of athletic shoes can omit the information about their price from the ads, but he cannot lie about it. In contrast, at some cost, he can hire any athlete to profess fondness of those shoes. This assumption treats manipulability of a message on a dimension and non-informativeness of that dimension as the same. This is plausible, but there are important exceptions to this assumption. For example, past returns on a mutual fund cannot be manipulated, but they provide little if any information about future returns.

The persuader's strategy is determined by how individuals respond to the messages. In this section, we make the *face value assumption*, namely that individuals, including both Bayesians and coarse thinkers, take the message m' they see and update as if it were the original message. In other words, they do not make the extra inference that, if they see m' , it tells them that the original message might have been a different m . In the next section, we allow for strategic inference by the individual, and show that the results presented in this section are largely robust to it.

A face value Bayesian updates according to:

$$E(q | \mathbf{m}', s) = f(\mathbf{m}', s) \tag{18}$$

In contrast, a face value coarse thinker updates according to:

$$E(q | \mathbf{m}', s) = k(\mathbf{m}', s) \tag{19}$$

The persuader's payoff, which he is assumed to maximize, is given by:

$$E(q | \mathbf{m}', s) - cX(\mathbf{m}, \mathbf{m}') \tag{20}$$

where $X(\mathbf{m}, \mathbf{m}')$ is a variable that equals the number of messages in \mathbf{m} that have been replaced to form \mathbf{m}' . That is, he maximizes expected quality net of persuasion costs.

We now study the optimal persuasion strategy of the persuader. To do so, we introduce some notation and a few simplifying assumptions. Denote by $M_i^*(\mathbf{m})$ the set of optimal messages for the persuader on dimension i given that the original message was \mathbf{m} , so that $M_i^*(\mathbf{m}) = \operatorname{argmax}_{m_i'} \{E[q_i | m_i', s] - c \mathbf{1}_{m_i' \neq m_i}\}$, where $\mathbf{1}$ is an indicator variable.

We study the properties of $M_i^*(\mathbf{m})$ for Bayesian and coarse thinkers. To understand how the persuader responds to changes in the underlying distribution of situations, we use the expression “changing $p(s)$ by ε ” to mean that the underlying distribution of situations is replaced by the probability distribution p' where $p'(s) = p(s) + \varepsilon$ and, for all other s' not equal to s ,

$$p'(s') = p(s') - \frac{\varepsilon p(s')}{p(C(s)) - p(s)}.$$

This definition allows us to examine how a persuader’s strategy changes as one situation becomes more or less prominent.

To simplify the analysis, we make three further assumptions. First, we assume that if the persuader is indifferent between altering or dropping a message and leaving it unchanged, he leaves it unchanged. Second, since we are interested in the incentives of persuaders to generate false messages, we assume that for every dimension i and situation s , there exist messages m_i' and m_i'' such that $k_i(m_i', s) > 0$ and $k_i(m_i'', s) < 0$. In other words, there exist potential positive and negative messages for the persuader to create along every dimension.⁷ These assumptions yield the following results:

⁷ Two points are worth noting about this assumption. First, weakening this assumption does not substantively change our results. For dimensions where this assumption is not true, we simply get the result that the persuader never attempts to create messages. Second, we require this assumption only for uninformative dimensions, since these are the only ones for which the persuader can create messages.

Proposition 1. Suppose individuals are face value Bayesians and the original message is $\mathbf{m}=(m_1, \dots, m_n)$. Then for all uninformative dimensions j :

$$M_j^*(\mathbf{m}) = \{m_j\}. \quad (21)$$

For all informative dimensions i :

$$M_i^*(\mathbf{m}) = \begin{cases} \{\emptyset\} & \text{if } f_i(m_i, s) < 0 \\ \{m_i\} & \text{if } f_i(m_i, s) \geq 0. \end{cases} \quad (22)$$

In other words, for an audience of face value Bayesians, the persuader only omits informative signals that are bad news and replaces them with nothing. He never pays the cost to create an uninformative signal and simply reports the truth on those dimensions.

Proposition 2. Suppose individuals are face value coarse thinkers. Then, for all informative dimensions i :

$$M_i^*(\mathbf{m}) = \begin{cases} \{\emptyset\} & \text{if } k_i(m_i, s) < 0 \\ \{m_i\} & \text{if } k_i(m_i, s) \geq 0 \end{cases} \quad (23)$$

For all uninformative dimensions j , the persuader may choose to create messages.

Specifically if $\max_{m_j'} k_j(m_j', s) - c > 0$, then

$$M_j^*(\mathbf{m}) = \begin{cases} \arg \max_{m_j'} k_j(m_j', s) & \text{if } k_j(m_j, s) < \max_{m_j'} k_j(m_j', s) - c \\ \{m_j\} & \text{if } k_j(m_j, s) \geq \max_{m_j'} k_j(m_j', s) - c \end{cases} \quad (24)$$

If $\max_{m_j'} k_j(m_j', s) - c < 0$, then

$$M_j^*(\mathbf{m}) = \begin{cases} \{\emptyset\} & \text{if } k_j(m_j, s) < 0 \\ \{m_j\} & \text{if } k_j(m_j, s) \geq 0 \end{cases}$$

Finally, if $p(s)$ is changed by ε , $\max_{m_j} k_j(m_j, s) - c$ tends towards zero as ε increases.

Thus, if initially (at $\varepsilon = 0$) $\max_{m_j} k_j(m_j, s) - c > 0$ and it is therefore sometimes optimal for the persuader to fabricate messages, for high enough ε $\max_{m_j} k_j(m_j, s) - c$ becomes negative and fabrication of messages ceases to ever be optimal.⁸

The contrast between the two propositions is interesting. Persuaders hide negative information from the face-value audience, whether it consists of Bayesians or coarse thinkers. But what constitutes negative information for these two groups is not the same. The criterion for hiding information from face-value Bayesians is whether $f_i(m_i, s) < 0$. The parallel criterion for face value coarse thinkers is whether $k_i(m_i, s) < 0$. There may be signals that are negative for one type but not for the other. The informativeness of a signal in situation s for the coarse thinker depends on $C(s)$, the other situations in the same category. For example, consider a mutual fund *not* reporting past returns. While this signal (or the lack thereof) would be negative if mutual funds were considered in isolation, it need not be so if mutual funds are co-categorized with other situations, such as other professional services, where past performance data may not exist.

⁸ If initially (at $\varepsilon = 0$) $\max_{m_j} k_j(m_j, s) - c < 0$ and it is sometimes optimal to drop messages (but never to fabricate them), the set of messages that is optimally dropped only shrinks in the limiting case where $\varepsilon = p(C(s)) - p(s)$ and $p(s)/p(C(s)) = 1$ and thus the set of messages that is optimally altered is independent of ε for $\varepsilon < p(C(s)) - p(s)$. However, this result is dependant on the assumption that it is costless to drop messages. For any positive cost of dropping messages, the set of optimally dropped messages would be shrinking in ε .

Persuaders never manufacture messages for face-value Bayesians. The reason is due to our assumption that messages can only be manufactured along uninformative dimensions, and Bayesians – unlike coarse thinkers – do not update their beliefs from such messages. Persuaders do, however, manufacture messages for coarse thinkers. When they do so, they simply pick the most positive message possible. Their decision whether to manipulate the data reduces to a simple question: is the gap (in terms of the improved assessment of quality) between the best possible message and the unmanipulated message big enough to offset the cost of manipulation?

Persuaders are more likely to manufacture a message in situation s when the probability of s is lower within its own category. Since the motive for doing so is the transference from other situations in a category to the current one, a higher probability of these other situations increases transference and therefore the benefit from manufacturing messages. This point is again closely related to what advertisers refer to as consumer involvement. A high involvement product occupies a huge probability space in its category ($p(s)/p(C)$ is close to 1, a situation we model as an increase in ϵ), so the transference from other situations is small, and hence so is the benefit of non-informative advertising. Our model predicts, as the marketing literature recommends, that advertising in these instances should be informative (Sutherland and Sylvester 2000). In contrast, low involvement products are mixed up in consumers' minds with many similar situations and hence there is greater scope for persuasion, exactly as the marketing literature suggests. Silk in shampoo, American Express travelers' checks, and Schwarzenegger's defense of free trade are all consistent with this point.

Proposition 2 can also help illuminate the Culver “silk in a bottle” example. As we assumed earlier, suppose there is a category consisting of two situations: $s = \text{“hair”}$ and $s' = \text{“shampoo”}$. Suppose the only available messages are $m = \text{“silk”}$ and the empty message. For the case of hair, the silk message conveys positive information, so that $f(\text{“silk”}, \text{“hair”}) > 0$ and there is thus room for purely informational persuasion in our model, along the lines of Stigler (1987). For the case of shampoo, it conveys no information, so that $f(\text{“silk”}, \text{“shampoo”}) = 0$. Thus there is no benefit of creating the Culver ad for the Bayesian thinker. On the other hand, consider the categorical thinker. Since this thinker has an update function $k(\text{“silk”}, \text{“shampoo”})$ that is a combination of $f(\text{“silk”}, \text{“hair”})$ and $f(\text{“silk”}, \text{“shampoo”})$, he actually views the “silk” message as informative: $k(\text{“silk”}, \text{“shampoo”}) > 0$. For a naïve coarse thinker, the Culver ad actually affects beliefs: the informativeness of silkiness in one situation is being transferred to another. Proposition 2 points out that, in this case, persuaders choose to create non-information: the Culver ad.

3. Sophistication

The results in section 2 rely on a simplifying assumption that individuals take information at face value. Sophisticated individuals, however, make inferences from the observed message about what the underlying information of the persuader must have been, recognizing that he is a rational wealth maximizer. We now ask how the recognition of strategic behavior influences inference about the true state of affairs.

Suppose that the persuader has some strategy $\mathbf{m}'=N(\mathbf{m},s)$, which is a map from the set of messages M into itself. The persuader replaces the message \mathbf{m} with the message $\mathbf{m}'=N(\mathbf{m},s)$ in situation s . Define $N^{-1}(\mathbf{m}',s)=\{\mathbf{m}|N(\mathbf{m},s)=\mathbf{m}'\}$. Also define $N_i(\mathbf{m},s)$ to be the induced map from \mathbf{M} to M_i so that it describes the persuader's behavior on dimension i . Since our assumptions guarantee that, under any optimal strategy of the persuader, $N_i(\mathbf{m},s)$ is independent of m_{-i} , we sometimes write $N_i(m_i,s)$. We assume that individuals know this strategy and update their expectations of quality accordingly. A Bayesian in this case updates according to:

$$\begin{aligned}
E^{SB}(q_i | m_i, s) &= \int_{m_i' \in N_i^{-1}(m_i, s)} \int_{q_i} q_i p_i(q_i | m_i', s) \Pr(m_i' | m_i, s) dq_i dm_i' \\
&= \int_{m_i' \in N_i^{-1}(m_i, s)} \int_{q_i} q_i p_i(q_i | m_i', s) \left(\frac{p_i(m_i' | s)}{p_i(N_i^{-1}(m_i, s) | s)} \right) dq_i dm_i' \quad (25) \\
&= \frac{1}{p_i(N_i^{-1}(m_i, s) | s)} \int_{m_i' \in N_i^{-1}(m_i, s)} \int_{q_i} q_i p_i(q_i | m_i', s) p_i(m_i' | s) dq_i dm_i' \\
&= \frac{1}{p_i(N_i^{-1}(m_i, s) | s)} \int_{m_i' \in N^{-1}(m_i, s)} f_i(m_i', s) p_i(m_i' | s) dm_i' \\
&\equiv g_i(m_i, s)
\end{aligned}$$

where $\Pr(m_i' | m_i, s)$ denotes the probability that m_i' is the true message when m_i is observed, given the persuader's strategy. A sophisticated Bayesian facing message m_i

forms a linear combination of the update rules $f_i(m_i', s)$ across all m_i' that would have induced the persuader to generate m_i . When we speak of the optimal equilibrium strategy of the persuader against a sophisticated Bayesian, we are therefore considering Perfect Bayesian Nash Equilibria in which the individual knows the persuader's strategy $N(\mathbf{m}, s)$ and updates taking it into account. We define $g(\mathbf{m}, s)$ to be the optimal expectation of \mathbf{q} :

$$E^{SB}(q | \mathbf{m}, s) = \sum_{i=1}^n E(q_i | m_i, s) = \sum_{i=1}^n g_i(m_i, s) = \sum_{i=1}^n \frac{1}{p(N_i^{-1}(m_i, s) | s)} \int_{m_i' \in N_i^{-1}(m_i, s)} f_i(m_i', s) p(m_i' | s) dm_i' \equiv g(\mathbf{m}, s) \quad (26)$$

We derive the updating rule of the sophisticated coarse thinker similarly.

Specifically, define:

$$N_i^{-1}(m_i, C(s)) \equiv \bigcup_{s' \in C(s)} N_i^{-1}(m_i, s') \quad (27)$$

$$\Pr(m_i' | m_i, s) \equiv \frac{p_i(m_i' | C(s))}{p_i(N_i^{-1}(m_i, s) | C(s))}, \quad (28)$$

and

$$\Pr(m_i' | m_i, C(s)) \equiv \frac{\sum_{s' \in C(s)} \Pr(m_i' | m_i, s') p(s')}{p(C(s))}. \quad (29)$$

The sophisticated coarse thinker updates according to:

$$\begin{aligned} E^{SC}(q_i | m_i, s) &= \int_{m_i' \in N_i^{-1}(m_i, C(s))} \int_{q_i} q_i p_i(q_i | m_i', C(s)) \Pr(m_i' | m_i, C(s)) dq_i dm_i' \\ &= \int_{m_i' \in N_i^{-1}(m_i, C(s))} k_i(m_i', s) \Pr(m_i' | m_i, C(s)) dm_i' \\ &= \frac{1}{p(C(s))} \int_{m_i' \in N_i^{-1}(m_i, C(s))} k_i(m_i', s) \sum_{s' \in C(s)} \Pr(m_i' | m_i, s') p(s') dm_i' \\ &\equiv k_i'(m_i, s) \end{aligned} \quad (30)$$

We can then compute the expected overall quality as:

$$E^{SC}(q | \mathbf{m}, s) = \sum_{i=1}^n E(q_i | m_i, s) = \sum_{i=1}^n k_i'(m_i, s) \equiv k'(\mathbf{m}, s) \quad (31)$$

As before, the persuader chooses a strategy $\mathbf{m}' = N(\mathbf{m}, s)$ to maximize:

$$E(q | N(\mathbf{m}, s), s) - cX(\mathbf{m}, N(\mathbf{m}, s)), \quad (32)$$

where $X(\mathbf{m}, N(\mathbf{m}, s))$ is a variable that equals the number of messages that have been fabricated to transform \mathbf{m} to $N(\mathbf{m}, s)$. By the definition of Perfect Bayesian Nash Equilibrium, the persuader's strategy should be optimal for each message \mathbf{m} that he sees. When faced with message \mathbf{m} , it should be optimal in that situation to report $N(\mathbf{m}, s)$. The following propositions summarize the results.

Proposition 3. Suppose individuals are sophisticated Bayesians and that $p_i(m_i = \emptyset | s) = 0$ for all informative dimensions i . Then there exists a unique equilibrium in which the optimal strategy of the persuader dictates that $N_h(m_h, s) = m_h$ for all informative and non-informative dimensions h .⁹ In other words, unless the empty message occurs naturally, the optimal strategy of the persuader is not to drop any information and to create no messages.

Proposition 3 illustrates the central feature of the traditional economics of persuasion with sophisticated Bayesian consumers. In this case, firms face only one avenue for persuasion: full provision of all available information; they cannot use non-information to persuade. Uninformative persuasion has no effect here. Moreover, firms must reveal all the information they have since selective dropping of information results

⁹ The uniqueness of this equilibrium depends on the assumption that, when indifferent, the persuader chooses to not alter a message.

in a penalty. This replicates the standard Grossman/Hart (1980) – Grossman (1981) – Milgrom (1981) result that, with sophisticated Bayesian thinkers, it does not pay the persuader to create or even drop messages since, when messages are absent, the audience assumes the worst.

Proposition 4. Suppose individuals are sophisticated coarse thinkers and $p_i(m_i = \emptyset | s) = 0$ for all dimensions i that are informative for a particular state s but that $p(m_i = \emptyset | C(s)) > 0$. Suppose further that, for all co-categorized situations s' other than s , persuaders tell the truth (or the audience gets the truth directly). Then for each informative dimension i there exists a constant a_i such that:

$$N_i(m_i, s) = \begin{cases} \emptyset & \text{if } k_i(m_i, s) < a_i \\ m_i & \text{if } k_i(m_i, s) > a_i \end{cases} \quad (33)$$

For all uninformative dimensions j , suppose that $\max_{m_j'} k_j(m_j', s) - c > 0$ so that it may be profitable to fabricate messages (rather than to simply drop them) on this dimension. Then there cannot be an equilibrium in which no fabrication takes place. Moreover, any equilibrium satisfies the conditions that there exists a constant b_j^s and a set Y_j such that $N_j(m_j, s) = m_j$ if and only if $k_j(m_j, s) \geq b_j^s$. If $k_j(m_j, s) < b_j^s$, then $N_j(m_j, s) \in Y_j$. This set Y_j has the features that: (i) $\arg \max_{m_j'} k_j(m_j', s) \in Y_j$ and (ii) $k_j'(m_j', s)$ is constant for all $m_j' \in Y_j \setminus \{\emptyset\}$.

Proposition 5. With coarse thinkers, sophistication makes it harder to drop or fabricate messages but does not eliminate dropping or fabrication. Specifically, $a_i < 0$ and, in the

case where $\max_{m_j} k_j(m'_j, s) - c > 0$, $b_j^s < \max_{m_j} k_j(m'_j, s) - c$. In some instances, the persuader manipulates messages against a face-value coarse thinker but not against a sophisticated one.

In contrast to Proposition 3, Propositions 4 and 5 illustrate that sophistication does not eliminate the advantage of uninformative persuasion for coarse thinkers. With coarse thinkers, while sophistication, as opposed to belief in the face value of the message, renders persuasion less effective and therefore less likely to be used at a given level of cost (Proposition 5), it does not eliminate the incentive to manufacture data along objectively uninformative dimensions. Proposition 4 shows that persuasion is primarily a consequence of coarse thinking and not of the failure to understand the strategic motive of the persuader. The face value assumption simplifies our results, but it is not critical.

A key assumption in Proposition 4 is that persuaders in the other situations are simply telling the truth (or the audience gets it directly). This assumption is important since it mutes the strategic interaction between persuaders. When individuals are sophisticated but coarse thinkers, the persuasion strategy of one persuader can affect that of another. This idea is captured in Proposition 6.

Proposition 6. Suppose there are only two situations $s=1$ and $s=2$ which are co-categorized and that individuals are sophisticated coarse thinkers. Further, suppose that for a given informative dimension i $p_i(m_i = \emptyset|1) = 0$ and $p_i(m_i = \emptyset|2) \neq 0$, so that the probability of no message is zero in one situation but positive in the other. If the persuader in situation 1 optimizes and the persuader in situation 2 tells the truth,

Proposition 4 defines the constants a_i which determine the optimal dropping rule for persuader 1. Suppose now that the persuader in situation 2 drops a set of messages Q such that $p(Q|C(s)) > 0$ and $k_i(m_i, I) < 0$ for all $m \in Q$. In this case, the optimal dropping rule for persuader 1 is defined by constants $d_i < a_i$ such that

$$N(m_i, s) = \begin{cases} \emptyset & \text{if } k_i(m_i, s) < d_i \\ m_i & \text{if } k_i(m_i, s) > d_i \end{cases} \quad (34)$$

Proposition 6 highlights the importance of strategic interaction between persuaders *in different situations*. The more one persuader engages in strategic concealment of information, the narrower is the scope for the other persuader to do so. To understand this, consider how a sophisticated coarse thinker responds to seeing no message on a dimension. This thinker understands that one of two things has happened: (i) there was actually no signal, or (ii) the signal was dropped. The greater his belief in (ii), the more he downgrades his estimate of quality. This is the lemons effect (Akerlof 1970). What is interesting is that the coarse thinker calibrates his belief in selective dropping using the rate at which signals are dropped for *all situations* in a category. The more wide-spread is selective dropping in other situations, the greater is the lemons effect and the fewer firms choose to drop. This logic of dilution also has implications for the dynamics of advertising. The more persuaders exploit transference of the same message, the less effective persuasion becomes. As more and more products are advertised as silky, the benefits of such advertising are diluted and eventually disappear.

4. Application: Mutual Funds

We now apply this model to the situation of an individual picking a mutual fund. Our key assumption is that coarse thinking individuals do not sufficiently differentiate between choosing a mutual fund and selecting other professional advisors, such as doctors or lawyers. Specifically, we take expected performance from a service to be its quality q . For simplicity, we take q to be continuous and normally distributed with mean zero. We suppose that individuals face two types of situations in selecting professional help: choosing a mutual fund (to help with investments) and choosing another professional service. We refer to the former as situation 1, and the latter as situation 2. Assume that the probability of situation 1 is p , and that of situation 2 is $1-p$.

The different types of messages that the individual might see are as follows. First, he might see information about past performance, r . For a mutual fund, this could be the return on the fund in the previous year. For another professional service, this could be any available metric of past performance, including the history of success of a particular surgery or type of lawsuit. We assume that past performance is a normally distributed continuous variable with mean zero. Second, the individual might see information about the fee for the service, c . For mutual funds, this is a composite of loads and all fees charged by the fund. For another professional service, this information is the fee for the service. We assume that this fee is a normally distributed continuous variable. Third, the individual might see qualitative information about the service provider's expertise, e . For a mutual fund, this would be the "soft" data contained in the ads, such as the number of professionals the money management firm employs or a write-up in a magazine. For a

professional service, this information might include testimonials of past customers. We assume that expertise e is uniformly distributed between -1 and 1.¹⁰

We make the stark assumption that, for mutual funds, the joint distribution of q, r, c and e is such that only c is informative about q . Specifically, for mutual funds (i.e., in situation 1), when all data are available, the best estimate of quality is given by:

$$E[q | r, c, e, (1)] = -c \quad (35)$$

A Bayesian would focus only on fees. He recognizes that past mutual fund returns do not predict future returns, and that qualitative statements about expertise are meaningless. The assumption that past returns on mutual funds do not predict future returns might be too strong empirically, but it is approximately correct.¹¹ We assume that, absent intervention by the persuader, all these data are available at all times.

For the other professional service, we assume that the joint distribution of q, r, c and e is such that all of them are informative about future performance. Define $\delta_c, \delta_r, \delta_e$ to be indicator variables for whether r, c and e are available (i.e., whether they equal \emptyset or not). We assume that the best guess of quality in situation 2 is:

$$E[q | e, r, c, (2)] = \beta r \delta_r - c \delta_c + e \delta_e \quad (36)$$

In other words, when assessing the quality of this other professional service, a Bayesian takes into account past performance and qualitative signals about expertise together with fees. We further assume that some of this information might not be available in some instances. With independent probability g , information on a given dimension is

¹⁰ We make this functional form assumption for the distribution of expertise because of our extreme assumption that persuasion entails only a fixed cost. If e were unbounded (e.g., normally distributed), the persuader would always manufacture an infinite e .

¹¹ See Chevalier and Ellison (1997), Sirri and Tufano (1998), Carhart (1997), and Swensen (2005).

unavailable. In that case, the update rule simply omits that piece of data so, for example, if r is unavailable, but e and c are, a Bayesian updates according to

$$E[q \mid r=\emptyset, c, e, (2)] = e - c. \quad (37)$$

The assumption that not all dimensions are available all the time for professional services is important when considering the sophisticated coarse thinker. Recall that, when this thinker sees an empty returns signal, he asks why that signal is not there. If, *in both situations*, that signal naturally exists in the absence of persuasion, then the thinker infers that the only reason for its absence is persuasion itself. This in turn leads to an unraveling and full revelation of information. If, on the other hand, *in one of the situations*, there is an exogenous reason for the absence of such information, then *in both situations*, the persuader has more freedom to selectively drop inconvenient information.¹²

These are the updating rules for a Bayesian. Let us now consider the coarse thinker. We suppose that a coarse thinker combines these two situations – selecting a mutual fund and selecting another professional – into the same category of selecting professional advice. As a consequence, this individual updates according to:

$$E[q \mid r, c, e, (1)] = E[q \mid r, c, e, (2)] = (1 - p)(\beta r \delta_r + e \delta_e) - c \delta_c \quad (38)$$

In the terminology of the general model, only c is *informative* for mutual funds whereas r, c and e are *informative* for other services.

¹² This is worth contrasting to the sophisticated Bayesian. Because the sophisticated Bayesian treats each situation independently, it would not be enough to assume that these signals were unavailable in *only one* situation since there is no transference across situations. Instead, one would need to assume that such signals are unavailable *in both situations*.

We make two departures from the original model of Sections 2 and 3. First, we assume that a mutual fund cannot manufacture information about returns or fees, r or c . Since information on past returns is taken to be uninformative about quality for mutual funds, this assumption conflicts with our previous assumption that uninformative messages (and hence past returns) can be manufactured. Instead, we assume here that the persuader can only costlessly drop information about r and c (that is, replace such information with \emptyset). Consistent with the previous model, however, he can at cost $C < 1$ replace expertise information e with whatever he wants.

Second, we focus on the persuader's problem only for mutual funds, and assume that other professional service providers simply report all information available and do not manufacture any.

As before, we assume that the persuader is maximizing $E[q|N(r,c,e),s]$ net of any cost C . Under these assumptions, the following propositions follow.

Proposition 7. Suppose individuals are face value Bayesians. Then the optimal strategy for the mutual fund is:

$$\begin{aligned} N_r(r,c,e) &= r \\ N_c(r,c,e) &= \begin{cases} c & \text{if } c < 0 \\ \emptyset & \text{if } c > 0 \end{cases} \\ N_e(r,c,e) &= e \end{aligned} \tag{39}$$

Suppose individuals are sophisticated Bayesians. Then the optimal strategy is:

$$N(r,c,e) = (r,c,e) \tag{40}$$

In other words, when individuals are face value Bayesians, the fund simply drops high fees data. When individuals are sophisticated Bayesians, the fund truthfully reports

all the data, as in the standard information revelation literature. Moreover, sophisticated Bayesians update only from the fee data.

Proof: This result follows directly from Propositions 1 and 3.

Proposition 8. Suppose individuals are face value coarse thinkers. Then there exists a constant b_e^f such that the optimal persuasion strategy is:

$$\begin{aligned}
 N_r(r, c, e) &= \begin{cases} r & \text{if } r > 0 \\ \emptyset & \text{if } r < 0 \end{cases} \\
 N_c(r, c, e) &= \begin{cases} c & \text{if } c < 0 \\ \emptyset & \text{if } c > 0 \end{cases} \\
 N_e(r, c, e) &= \begin{cases} e & \text{if } e > b_e^f \\ 1 & \text{if } e < b_e^f \end{cases}
 \end{aligned} \tag{41}$$

Proof: This result follows directly from Proposition 2.

Because the coarse thinker groups situations into categories, he does not differentiate the problems of selecting a mutual fund and another expert. Mutual funds then attempt to capitalize on such coarse thinking through their persuasion strategies. According to Proposition 8, there are three features of such persuasion. First, when they charge high fees, funds choose not to report them. Second, when faced with low returns, funds choose not to report them. Third, funds may generate and report qualitative information about expertise and will only report that information when returns are poor and fees are high.

These predictions are broadly consistent with the evidence on mutual fund advertising. First, financial ads only rarely include information about fees (Cronqvist 2005). Only the funds with low fees, such as Vanguard, typically talk about them.

Second, as demonstrated by Mullainathan and Shleifer (2006), there is evidence that financial advertisements are more likely to include information on past returns when such information is favorable. The authors study financial advertising in *Money* and *Business Week* magazines during 1994-2003, roughly the full course of the internet bubble. The authors collect all the consumer-oriented financial ads from these magazines, classify them by content, and aggregate them in the quarterly series. One of the variables they consider is the share of ads that contain own past returns (absolute or relative) in all stock mutual fund ads. Figure 1 presents the relationship between this variable and the rolling one quarter lagged Standard and Poor's 500 Index return. As Figure 1 shows, on average only about 60 percent of the mutual fund ads present any data about own past returns, and the correlation between the inclusion of these data in the ads and past market returns is over .7 for both *Money* and *Business Week*. Indeed, Figure 1 makes clear that after the market crash, past returns disappear from the advertisements. This finding is broadly consistent with the prediction of our model.

Third, again as shown by Mullainathan and Shleifer (2006), much of the information contained in financial advertisements is soft (e.g., describing the alleged expertise and experience of the advertiser), and the incidence of such soft information rises after the market falls. Figure 2 classifies the ads by T. Rowe Price, a mutual fund complex and the most frequent advertiser in the sample, into those for growth funds, other mutual funds, and everything else, which tend to be softer ads about expertise.

Figure 2 shows how, after the market crashes in 2001, growth fund ads disappear and are replaced, in the first instances, by the softer ads. The post-crash judges and crossing guards campaign from Charles Schwab is representative of a broader phenomenon.

Although one can rationalize all these findings in a number of ways, the perspective of this paper, in which financial advertisers free ride on their audience's broader experience – and hence co-categorization -- with choosing experts, appears to explain all three aspects of the data most simply.

Proposition 9. Suppose individuals are sophisticated coarse thinkers. Then there exist constants $b_c^s > 0$, and $b_r^s < 0$, such that the optimal persuasion strategy is:

$$\begin{aligned} N_r(r, c, e) &= \begin{cases} r & \text{if } r > b_r^s \\ \emptyset & \text{if } r < b_r^s \end{cases} \\ N_c(r, c, e) &= \begin{cases} c & \text{if } c < b_c^s \\ \emptyset & \text{if } c > b_c^s \end{cases} \end{aligned} \quad (42)$$

On the expertise dimension, any equilibrium satisfies the conditions that there exist constants $b_e^s < b_e^f$ and $e^* < 1$ such that an optimal strategy is defined so that if $e < b_e^s$, the persuader replaces e by a number in $[e^*, 1]$ or drops it.

The key point here is the sign of the constants, namely $b_c^s > 0$, and $b_r^s < 0$. With a sophisticated audience, a fund actually reports some negative returns and high fees, but censors very negative returns or very high fees. The fund still engages in the production of qualitative information, but does so over smaller ranges. Based on the evidence in Mullainathan-Shleifer (2006), the data are easier to reconcile with the case of face value coarse thinkers, since there is no evidence of reporting of negative returns.

Proof: This result follows directly from Proposition 4.

6. Conclusion

This paper has supplied a simple formal model encapsulating a collection of ideas about inference and persuasion from such diverse fields as linguistics, psychology, philosophy, politics, and advertising. The basic building blocks of the model are categorization and transference: people group diverse situations into categories and use information about one member of a category to evaluate another, even if the information is objectively useless in the latter case. The model includes full Bayesian rationality as a special case, in which each situation is evaluated as if in its own category.

The model sheds light on a number of phenomena. Most importantly, it explains which types of uninformative persuasion work and which do not. A persuader will use specific types of messages that do not contain any objectively useful information. The model explains why uninformative persuasion is particularly important in low involvement situations, such as advertising cheap goods or political candidates, whereas high involvement choices require information. The model also helps understand the pervasive phenomenon of persuaders' omitting bad payoff-relevant news from their messages. This is possible in our model even if the audience thinks strategically. Empirically, the model helped us think about several otherwise hard-to-understand persuasive messages (such as putting silk in a shampoo), but also clarified a number of puzzling features of financial advertising, such as the omission of data about fees and returns (when the latter are bad), as well as the preponderance of "soft" messages that at least in the standard financial model should be of little interest to investors.

At the methodological level, our model suggests that key to understanding a range of empirically observed manifestations of “bounded rationality” are not either the violation of Bayes' rule, or even the failure of strategic thinking on the part of the decision maker. Rather, the failure seems to lie in the decision maker's use of wrong models, in the very specific sense of co-categorizing distinct situations. Even when decision makers apply the Bayes' rule and think strategically, but form coarse categories, their inference is flawed and they are vulnerable to persuasion.

As a final point, we reiterate that our paper is just a first step in the analysis of coarse thinking and persuasion. We have treated categories as rigid and exogenously fixed, and their members as completely symmetric in the mind of the decision maker. But as Lakoff (1987) shows, categories are far from rigid, they often overlap, their membership changes, and some members are more prototypical than others. These steps toward realism would obviously change the updating rules we wrote down. We hope to find out what additional insights could be gained from these improvements.

References

- Akerlof, George (1970), "The Market for Lemons," *Quarterly Journal of Economics* 84, 488-500.
- Aragonis, Enriqueta, Itzhak Gilboa, Andrew Postlewaite and David Schmeidler (2001), "Rhetoric and Analogies," Working Paper.
- Becker, Gary (2001), "Rational Indoctrination and Persuasion," University of Chicago Mimeo.
- Becker, Gary, and Kevin Murphy (1993), "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics* 108, 941-964.
- Becker, Gary, and George Stigler (1977), "De Gustibus Non Est Disputandum," *American Economic Review* 67, 76-90.
- Bernartzi, Shlomo, and Richard Thaler (2006), "Heuristics and Biases in Retirement Savings Behavior," Mimeo.
- Carhart, Mark (1997), "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57-82.
- Carpenter, Gregory, Rashi Glazer, and Kent Nakamoto (1994), "Meaningful Brands from Meaningless Differentiation: The Dependence on Irrelevant Attributes," *Journal of Marketing Research* 31, 339-350.
- Chevalier, Judith, and Glenn Ellison (1997), "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy* 105, 1167-1200.
- Cronqvist, Henrik (2005), "Advertising and Portfolio Choice," University of Chicago Mimeo.

- Daniel, Kent, and Sheridan Titman (2006), "Market Reactions to Tangible and Intangible Information," *Journal of Finance*, forthcoming.
- Dewatripont, Mathias, and Jean Tirole (1999), "Advocates," *Journal of Political Economy* 107, 1-39.
- Downs, Anthony (1957), *An Economic Theory of Democracy*, New York, NY: Harper and Row.
- Edelman, Gerald (1992), *Bright Air, Brilliant Fire: On the Matter of the Mind*, New York, NY: BasicBooks.
- Fryer, Roland, and Matthew Jackson (2004), "A Categorical Model of Cognition and Biased Decision-Making," Working Paper.
- Gabaix, Xavier, and David Laibson (2006), "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics* 121, 505-540.
- Gentzkow, Matthew, and Jesse Shapiro (2006), "Media Bias and Reputation," *Journal of Political Economy* 114, 280-316.
- Glaeser, Edward (2005), "The Political Economy of Hatred," *Quarterly Journal of Economics* 120, 45-86.
- Glaeser, Edward, Giacomo Ponzetto, and Jesse Shapiro (2005), "Strategic Extremism: Why Republicans and Democrats Divide on Religious Values," *Quarterly Journal of Economics* 120, 1283-1330.
- Glazer, Jacob, and Ariel Rubinstein (2001), "Debates and Decisions: On a Rationale of Argument Rules," *Games and Economic Behavior* 36, 158-173.

- Glazer, Jacob, and Ariel Rubinstein (2004), "On Optimal Rules of Persuasion," *Econometrica* 72, 1715-1736.
- Grossman, Sanford (1981), "The Informational Role of Walrasian and Private Disclosure about the Product Quality," *Journal of Law and Economics* 24, 461-483.
- Grossman, Sanford, and Oliver Hart (1980), "Disclosure Laws and Takeover Bids," *Journal of Finance* 35, 323-334.
- Holden, Richard, and Emir Kamenica (2006), "How to Persuade a Bayesian," Mimeo.
- Kahneman, Daniel, and Amos Tversky (1982), *Judgment Under Uncertainty: Heuristics and Biases*, New York, NY: Cambridge University Press.
- Lakoff, George (1987), *Women, Fire, and Dangerous Things*, Chicago, IL: The University of Chicago Press.
- Luntz, Frank (2006), *Frank Luntz Strategy Report*, Frank Luntz Research Companies.
- McCloskey, Donald, and Arjo Klamer (1995), "One Quarter of GDP is Persuasion," *American Economic Review Papers and Proceedings* 85, 191-195.
- Milgrom, Paul (1981), "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics* 12, 380-391.
- Milgrom, Paul, and John Roberts (1986), "Relying on the Information of Interested Parties," *Rand Journal of Economics* 17, 18-32.
- Mullainathan, Sendhil (2000), "Thinking through Categories," MIT Mimeo.
- Mullainathan, Sendhil, and Andrei Shleifer (2002), "Media Bias," NBER Working Paper no. 9295.
- Mullainathan, Sendhil, and Andrei Shleifer (2005), "The Market for News," *American Economic Review* 95, 1031-1053.

- Mullainathan, Sendhil, and Andrei Shleifer (2006), "Persuasion in Finance," NBER Working Paper no. 11838.
- Murphy, Kevin, and Andrei Shleifer (2004), "Persuasion in Politics," *American Economic Review Papers and Proceedings* 94, 435-439.
- Okuno-Fujiwara, Masahiro, Andrew Postlewaite, and Kotaro Suzumura (1990), "Strategic Information Revelation," *Review of Economic Studies* 57, 25-47.
- Peski, Marcin (2006), "Categorization," University of Chicago working paper.
- Posner, Richard (1995), "Rhetoric, Legal Advocacy, and Legal Reasoning," in *Overcoming Law*, Cambridge, MA: Harvard University Press.
- Sirri, Erik, and Peter Tufano (1998), "Costly Search and Mutual Fund Flows," *Journal of Finance* 53, 1589-1622.
- Stigler, George (1961), "The Economics of Information," *Journal of Political Economy* 69, 213-225.
- Stigler, George (1987), *The Theory of Price*, 4th ed., New York, NY: MacMillan Publishing.
- Sutherland, Max, and Alice Sylvester (2000), *Advertising and the Mind of the Consumer*, 2nd ed., St. Leonards, Australia: Allen & Unwin.
- Swensen, David (2005), *Unconventional Success: A Fundamental Approach to Personal Investment*, New York: Free Press.
- Tullock, Gordon (1967), *Toward a Mathematics of Politics*, Ann Arbor: University of Michigan Press.
- Zaltman, Gerald (1997), "Rethinking Market Research: Putting People Back In," *Journal of Marketing Research* 34, 424-437.

Zaltman, Gerald (2003), *How Customers Think*, Boston, MA: Harvard Business School Press.

Proof of Proposition 1. We know that the face value Bayesian updates according to:

$$\begin{aligned} E[q | \mathbf{m}', s] &= f(\mathbf{m}', s) \\ &= f_1(m_1', s) + f_2(m_2', s) + \dots + f_n(m_n', s) \end{aligned}$$

and so the persuader is maximizing

$$E[q | \mathbf{m}', s] - cX(\mathbf{m}, \mathbf{m}') = f_1(m_1', s) + f_2(m_2', s) + \dots + f_n(m_n', s) - cX(m, m')$$

Along informative dimension i , the only decision by assumption is whether or not to costlessly drop the message. Recall that $f_i(\emptyset, s) = 0$. So if $f_i(m_i, s) < 0$, dropping the message increases posteriors at zero cost and hence increases the payoff. If $f_i(m_i, s) > 0$, dropping the message only decreases posteriors and hence decreases payoffs. Finally if $f_i(m_i, s) = 0$, dropping the message does not affect posteriors. Since we have assumed that the persuader, when indifferent, does not alter the message, it follows that dropping a message is optimal if and only if $f_i(m_i, s) < 0$.

Along any uninformative dimension j , $f_j(m_j', s) = 0$ for all messages m_j' . Thus, beliefs are independent of the message the persuader chooses to create, and the persuader is indifferent between all possible messages. Since we have assumed that, when indifferent, the persuader does not alter the message, it follows that the persuader does not change the message on dimension j . ♦

Proof of Proposition 2. The face value coarse thinker updates according to:

$$\begin{aligned} E[q | \mathbf{m}', s] &= k(\mathbf{m}', s) \\ &= k_1(m_1', s) + k_2(m_2', s) + \dots + k_n(m_n', s) \end{aligned}$$

so the persuader is maximizing

$$E[q | \mathbf{m}', s] - cX(\mathbf{m}, \mathbf{m}') = k_1(m_1', s) + k_2(m_2', s) + \dots + k_n(m_n', s) - cX(m, m')$$

Along informative dimension i , the only decision by assumption is whether or not to costlessly drop the message. Recall that $f_i(\emptyset, s) = 0$ for all states, so $k_i(\emptyset, s) = 0$. If $k_i(m_i, s) < 0$, dropping the message would then increase posteriors at zero cost and hence increase payoffs. If $k_i(m_i, s) > 0$, dropping the message only decreases posteriors and hence payoffs. Finally, if $k_i(m_i, s) = 0$, dropping the message does not affect posteriors. Since we have assumed that the persuader, when indifferent, does not alter the message, it follows that dropping a message is optimal if and only if $k_i(m_i, s) < 0$.

Along any uninformative dimension j , the persuader has a more complicated calculus. He must choose between: (i) doing nothing, (ii) replacing the message, or (iii) dropping the message. Several observations are in order. First, if he does choose to replace the message, he should replace it with the best possible message: $\arg \max_{m_j} k_j(m_j', s)$. Second, if he drops the message then, since $k_i(\emptyset, s) = 0$, he receives payoff 0 on that dimension. Therefore, if he does alter the message, he drops it if $\max_{m_j} k_j(m_j', s) - c < 0$ and replaces it if $\max_{m_j} k_j(m_j', s) - c > 0$. We consider each of these cases in turn.

Suppose $\max_{m_j} k_j(m_j', s) - c < 0$. Then the persuader's choice is between dropping and doing nothing. Since doing nothing yields a payoff of $k_j(m_j, s)$ and dropping the message yields a payoff of 0, the persuader drops if and only if $k_j(m_j, s) < 0$. Now suppose $\max_{m_j} k_j(m_j', s) - c > 0$. Doing nothing yields a payoff of $k_j(m_j, s)$, whereas altering the message yields a payoff of $\max_{m_j} k_j(m_j', s) - c$. Therefore the persuader replaces if and only if $\max_{m_j} k_j(m_j', s) - c > k_j(m_j, s)$.

Finally, recall that by definition:

$$k_j(m_j, s) = f_j(m_j, s) \frac{p(s)}{p(C)} + \sum_{s' \in C(s), s' \neq s} f_j(m_j, s') \frac{p(s')}{p(C)}.$$

Along uninformative dimensions (where $f_j(m_j, s) = 0$), if $p(s)$ is changed by ε :

$$\begin{aligned} & |k_j(m_j, s; \varepsilon)| \\ &= \left| \sum_{s' \in C(s), s' \neq s} f_j(m_j, s') \frac{p(s') - \frac{\varepsilon p(s')}{p(C(s)) - p(s)}}{p(C(s))} \right| \\ &= \left| k_j(m_j, s; \varepsilon = 0) - \sum_{s' \in C(s), s' \neq s} f_j(m_j, s') \frac{\frac{\varepsilon p(s')}{p(C(s)) - p(s)}}{p(C(s))} \right| \\ &\rightarrow 0 \text{ as } \varepsilon \rightarrow p(C(s)) - p(s) \end{aligned}$$

Consider the case where $\max_{m_j'} k_j(m_j', s; \varepsilon = 0) - c > 0$, so it is initially better for the persuader to replace a given message with the best possible message rather than to drop it. As ε increases, $\max_{m_j'} k_j(m_j', s; \varepsilon) - c$ decreases and it is easy to see that there is some $\varepsilon^* \in (0, p(C(s)) - p(s))$ such that $\max_{m_j'} k_j(m_j', s; \varepsilon^*) - c = 0$. Thus, for high enough ε , it is never profitable for the persuader to fabricate messages (though it is still optimal to drop some). ♦

Proof of Proposition 3: Consider an uninformative dimension j . Recall that:

$$g_j(m_j, s) = \frac{1}{p_j(N_j^{-1}(m_j, s) | s)} \int_{m_j' \in N_j^{-1}(m_j, s)} f_j(m_j', s) p(m_j' | s) dm_j'$$

Since j is uninformative, it follows that, irrespective of the persuader's strategy, $g_j(m_j, s) = 0$. Consequently, in the optimal strategy the persuader does not drop or alter messages on uninformative dimensions.

Consider now an informative dimension i . Recall that along informative dimensions the persuader's only decision is whether or not to drop the signal. Suppose that the persuader in fact finds it optimal in some equilibrium to drop some message m_i . Let Z_i be the set of such messages that are dropped in this equilibrium. Let $m_i \in Z_i$ be a dropped message. We make two observations. First, we wish to show that if m_i' is such that $f_i(m_i', s) \leq f_i(m_i, s)$, then in any equilibrium such that $m_i \in Z_i$, $m_i' \in Z_i$ as well. To see this, note that if $m_i \in Z_i$ in equilibrium, it must be the case that $g_i(\emptyset, s) > f_i(m_i, s)$. Consequently, $g_i(\emptyset, s) > f_i(m_i', s)$ and hence it is optimal to also drop m_i' . Second, since

$$g_i(\emptyset, s) = \frac{1}{p(N_i^{-1}(\emptyset, s) | s)} \int_{m_i' \in N_i^{-1}(\emptyset, s)} f_i(m_i', s) p(m_i' | s) dm_i'$$

it follows that $f_i(m_i, s)$ must be constant for all messages in Z_i . Otherwise, $\max_{m_i \in Z_i} f_i(m_i, s) > g_i(\emptyset, s)$ and the persuader is better off deviating from his strategy when facing $\arg \max_{m_i \in Z_i} f_i(m_i, s)$, contradicting this being an equilibrium. These two observations together imply that, at most, Z_i includes only the set of messages m_i that minimize $f_i(m_i, s)$. But, in this case, $f_i(m_i, s) = g_i(\emptyset, s)$ for all such messages and the persuader is indifferent between dropping and not dropping these messages. Hence, the supposition that Z_i is non-empty contradicts the assumption that when indifferent the persuader does not drop messages. It is easy to see that, in fact, $N_h(m_h, s) = m_h$ constitutes an equilibrium strategy and, from the above argument, it is the only strategy that satisfies the assumption that when indifferent the persuader does not drop messages. ♦

Proof of Proposition 4. Consider an informative dimension i . Recall that along informative dimensions the persuader's only decision is whether or not to drop the message. It is useful to re-order the messages on this dimension using a map J from $x \in [0,1]$ to $m_i \in M_i$ so that $k_i(J(x),s) < k_i(J(x'),s)$ if and only if $x < x'$. Let Z be the set of messages that are to be dropped. Upon seeing \emptyset , the sophisticated coarse thinker updates according to:

$$\begin{aligned}
E[q_i | (m_{-i}, \emptyset), s] &= k_i'(\emptyset, s) \\
&= k_i(\emptyset, s) \Pr(\emptyset | \emptyset, C(s)) + \int_{x \in Z} k_i(J(x), s) \Pr(J(x) | \emptyset, C(s)) dx \\
&= \int_{x \in Z} k_i(J(x), s) \Pr(J(x) | \emptyset, C(s)) dx \quad (\text{from the assumption that } k_i(\emptyset, s) = 0) \\
&= \int_{x \in Z} k_i(J(x), s) \frac{p_i(J(x) | C(s))}{p_i(Z | C(s)) + p_i(\emptyset | C(s))} \frac{p(s)}{p(C(s))} dx.
\end{aligned}$$

The set Z is such that if $x \in Z$ in equilibrium then all $x' < x$ must also be in Z . To see this, note that the payoff from not dropping x' is $k_i'(J(x'), s)$. Since messages can only be dropped and all other persuaders are telling the truth, it follows that $k_i'(J(x'), s) = k_i(J(x'), s)$. If it is profitable to drop x , then $k_i(J(x), s) < k_i'(\emptyset, s)$. But, since $k_i'(J(x'), s) = k_i(J(x'), s) < k_i(J(x), s)$, it follows that $k_i(J(x'), s) < k_i'(\emptyset, s)$ and it is profitable to drop x' as well.

Defining Z therefore amounts to finding a threshold, call it r^* , such that x is in Z if and only if $x < r^*$. To establish this threshold, we know that a threshold r^* is optimal if $E[q_i | J(x), s] < E[q_i | \emptyset, s]$ if and only if $x < r^*$. Consider the difference between $E[q_i | J(x), s]$ and $E[q_i | \emptyset, s]$ at a potential threshold r :

$$L(r) \equiv k_i(J(r), s) - k_i(\emptyset, s) = k_i(J(r), s) - \int_{x \in [0, r]} k_i(J(x), s) \frac{p_i(J(x) | C(s))}{p_i([J(0), J(r)] | C(s)) + p_i(\emptyset | C(s))} \frac{p(s)}{p(C(s))} dx.$$

Several features of this function $L(r)$ are worth noting. First, $L(0) < 0$. To see this, note that at $r=0$ the minimum message has negative information content by assumption and the second term is zero. Second, $L(r)$ is increasing. To see this, note that we have rank ordered messages so that $k_i(J(r), s)$ is increasing and, while the subtracted term increases in r as well, it does not increase by enough to offset the increase in $k_i(J(r), s)$. Finally, note that $L(1) > 0$. This follows because the maximum message has positive information content by assumption, whereas the second term equals zero since we have assumed that $E[q_i] = 0$. Thus, since this continuous function begins negative, increases monotonically and ends positive, it must have a value of r^* such that $L(r^*) = 0$. By defining $a_i = k_i(J(r^*), s)$, we get the first part of the proposition.

Now, consider an uninformative dimension j . As in the proof of the first part of the proposition, re-order the messages on this dimension using a map J from $x \in [0, 1]$ to $m_j \in M_j$ so that $k_j(J(x), s) < k_j(J(x'), s)$ if and only if $x < x'$. Assuming that $\max_{m_j} k_j(m_j, s) - c > 0$, there cannot be an equilibrium in which the persuader fails to drop messages. To see this, suppose that the persuader chooses not to replace any messages in equilibrium. Then, upon receiving message $J(0)$, it is profitable for him to deviate from his strategy and replace $J(0)$ with $J(1)$, a contradiction to his strategy being optimal.

Thus, in any equilibrium, the persuader replaces messages in some non-empty set Z_j with messages in some non-empty set Y_j (Y_j may or may not include \emptyset). These sets

must have the properties that: (i) $Z_j = [J(0), J(x^l)]$, (ii) $Y_j \setminus \{\emptyset\} = (J(x^s), J(1)]$ and (iii) $k_j'(m_j', s)$ is constant for all $m_j' \in Y_j \setminus \{\emptyset\}$.

To see (i), suppose that, in equilibrium, $J(x)$ is replaced by some message y_j in Y_j . Then it must be the case that the payoff from not replacing $J(x)$, $k_j'(J(x), s)$, is less than the payoff from replacing $J(x)$, $k_j'(y_j, s) - c$. First note that this means that, in equilibrium, no message is replaced by $J(x)$ since replacing that same message with y_j would lead to a higher payoff. Hence, the equilibrium payoff from not replacing $J(x)$ is $k_j'(J(x), s) = k_j(J(x), s)$. Since for all $x' < x$, $k_j(J(x'), s) < k_j(J(x), s) < k_j'(y_j, s) - c$, it is profitable to replace all $x' < x$ as well so it must be the case that such messages are replaced in equilibrium. Thus, defining Z_j amounts to finding a threshold x^l such that $J(x)$ is in Z_j if and only if $x < x^l$.

To see (ii), suppose that, in equilibrium $y_j \in Y_j \setminus \{\emptyset\}$ so it is profitable to replace some message x with y_j . Now suppose that y_j' is not in $Y_j \setminus \{\emptyset\}$ for some $y_j' > y_j$. This cannot be the case in equilibrium because it would be profitable for the persuader to deviate from this strategy when facing message x and replace it with y_j' rather than y_j . Thus, $Y_j \setminus \{\emptyset\}$ must equal either $[J(x^s), J(1)]$ or $(J(x^s), 1]$ for some x^s . To rule out the first set, suppose that x^s is in $Y_j \setminus \{\emptyset\}$. Then $k_j'(x^s, s) < k_j(x^s, s)$ and, by the continuity of $k_j(\cdot, s)$, $k_j'(x^s, s) < k_j(x_j, s) = k_j'(x_j, s)$ for some $x_j < x^s$. This cannot be the case in equilibrium because it would be profitable for the persuader to deviate from this strategy when facing some message in $N_j^{-1}(x^s, s)$ and replace it with x_j instead. Hence, defining $Y_j \setminus \{\emptyset\}$ amounts to finding a threshold x^s such that $Y_j = (J(x^s), 1]$.

Finally, to see (iii) suppose that $k_j(m_j', s)$ is not constant for all $m_j' \in Y_j \setminus \{\emptyset\}$ in equilibrium. That is, suppose that there exist $m_j', m_j'' \in Y_j \setminus \{\emptyset\}$ such that $k_j'(m_j', s) <$

$k_j'(m_j'', s)$. This cannot be the case because then it would be profitable for the persuader to deviate from this strategy when facing some message in $N_j^{-1}(m_j', s)$ and replace it with m_j'' instead. ♦

Proof of Proposition 5. To see that $a_i < 0$, recall that a_i is defined implicitly by $a_i = k_i(J(r^*), s)$, where r^* is the value of r that solves:

$$L(r) = k_i(J(r), s) - k_i'(\emptyset, s) = k_i(J(r), s) - \int_{x \in [0, r]} k_i(J(x), s) \frac{p_i(J(x) | C(s))}{p_i([J(0), J(r)] | C(s)) + p_i(\emptyset | C(s))} \frac{p(s)}{p(C(s))} dx = 0$$

Let r' be such that $k_i(J(r'), s) = 0$. At this r'

$$L(r') = - \int_{x \in [0, r']} k_i(J(x), s) \frac{p_i(J(x) | C(s))}{p_i([J(0), J(r)] | C(s)) + p_i(\emptyset | C(s))} \frac{p(s)}{p(C(s))} dx > 0$$

since $k_i(J(x), s) < 0$ for all $x \in [0, r']$. Since $L(r)$ is increasing in r , this means that $L(r^*) = 0$ for some $r^* < r'$ and, since the messages are ranked ordered, at this r^* , $a_i = k_i(J(r^*), s) < k_i(J(r'), s) = 0$.

To see that $b_j^s < \max_{m_j'} k_j(m_j', s) - c$, note that since some negative messages are replaced with $\arg \max_{m_j'} k_j(m_j', s)$ in the equilibrium described in Proposition 4,

$\max_{m_j'} k_j'(m_j', s) - c < \max_{m_j'} k_j(m_j', s) - c$ and thus the payoff from keeping the message x that yields $k_j(x, s) = \max_{m_j'} k_j(m_j', s) - c$ is higher than the payoff from replacing that message with any other. Hence the cutoff must be lower than $\max_{m_j'} k_j(m_j', s) - c$. ♦

Proof of Proposition 6. The proof proceeds similarly to the proof of Proposition 4.

Recall that if the persuader in situation 2 does not drop any messages then the dropping rule for persuader 1 is implicitly defined by finding the value of r , r^* , that solves:

$$L(r) = k_i(J(r),1) - k_i'(\emptyset,1) = k_i(J(r),1) - \int_{x \in [0,r]} k_i(J(x),1) \frac{p_i(J(x)|C(1))}{p_i([J(0),J(r)]|C(1)) + p_i(\emptyset|C(1))} \frac{p(1)}{p(C(1))} dx = 0$$

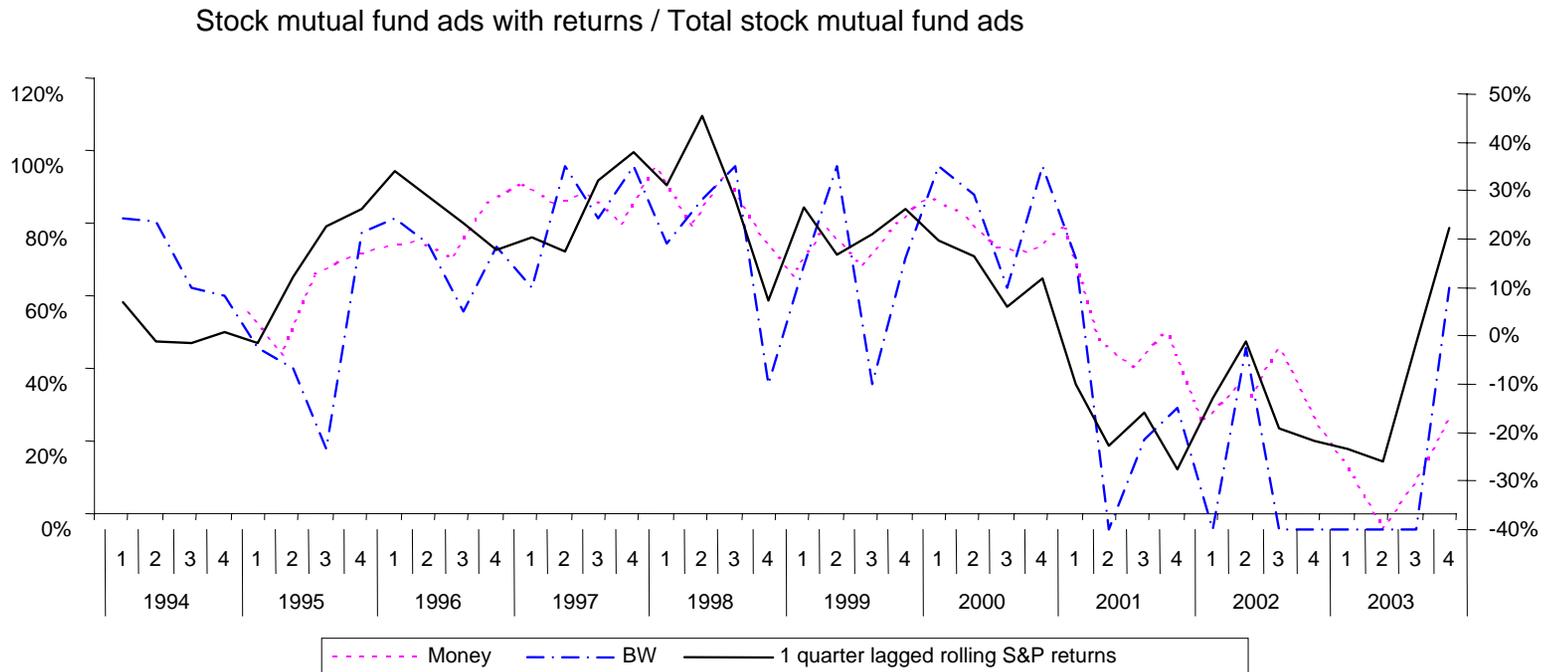
and then setting $a_i = k_i(J(r^*),1)$.

If the persuader in situation 2 does drop messages in Q , the new threshold is defined by solving a different equation:

$$\begin{aligned} L'(r) &= k_i(J(r),1) - k_i'(\emptyset,1) \\ &= k_i(J(r),1) - \int_{x \in [0,r]} k_i(J(x),1) \frac{p_i(J(x)|C(1))}{p_i([J(0),J(r)]|C(1)) + p_i(\emptyset|C(1))} \frac{p(1)}{p(C(1))} dx - \int_{m \in Q} k_i(m,1) \frac{p_i(m|C(1))}{p_i(Q|C(1)) + p_i(\emptyset|C(1))} \frac{p(2)}{p(C(1))} dx \\ &= 0, \end{aligned}$$

where the additional term reflects the extra update that a sophisticated coarse thinker makes upon seeing \emptyset . She now includes the possibility that the dropped message could be coming from the set of messages that persuader 2 drops. The value of r , $r^{*'}$, that solves the above equation implicitly defines d_i : $d_i = k_i(J(r^{*'}),1)$. We can show that $r^{*'}$ exists exactly as we did in the proof of Proposition 4. To see that $d_i < a_i$, note that $k_i(m,1) < 0$ for all $m \in Q$. Hence, at the original r^* that implicitly defines a_i , $L'(\cdot)$ is positive and thus $r^{*' < r^*$ so $d_i < a_i$ ♦

Figure 1



Correlations:

Money / S&P : 71.0%
BW / S&P: 74.3%

Means:

Money: 62.4%
BW: 59.0%

Figure 2

T.RowePrice Ad Composition (Money Magazine)

