

Using Models to Persuade

Joshua Schwartzstein
Harvard Business School

Adi Sunderam*
Harvard Business School

May 16, 2019

Abstract

We present a framework for analyzing “model persuasion,” where persuaders influence people’s beliefs by proposing models (likelihood functions) that suggest how to organize past data (e.g., on investment performance) to make predictions (e.g., about future returns). Receivers are assumed to find models more compelling when they better explain the past. A key tradeoff persuaders face is that models that better fit the data induce less movement in receivers’ beliefs. Model persuasion sometimes makes the receiver worse off than if he interprets data through the lens of a default model. The receiver is most misled by persuasion when there is a lot of publicly available data that is open to interpretation and exhibits randomness, as this gives the persuader “wiggle room” to highlight false patterns. Even when the receiver is exposed to the true model, the wrong model often wins because it better fits the past. Competition more broadly pushes persuaders towards overfitting available data and as a result tends to neutralize the data by leading receivers to view it as unsurprising. With multiple receivers, a persuader is more effective when receivers share similar priors and default interpretations. We illustrate with examples from finance, politics, and law.

*We thank Ned Augenblick, Pedro Bordalo, Erik Eyster, Tristan Gagnon-Bartsch, Nicola Gennaioli, Robert Gibbons, Brian Hall, Sam Hanson, Botond Koszegi, Deepak Malhotra, Sendhil Mullainathan, Matthew Rabin, Jesse Shapiro, Andrei Shleifer, Jeremy Stein, Dmitry Taubinsky, and seminar participants at UC Berkeley, MIT, and Harvard for helpful comments.

1 Introduction

Persuasion frequently involves an expert providing a “model” of the world, an interpretation of known data that favors the outcome the expert desires. When real estate agents tell potential home buyers, “House prices in this neighborhood are rising because of the schools,” they are supplying a model: home buyers who care about the resale value of their homes should pay attention to local schools, which are an important determinant of house price appreciation. Potential presidential candidates who do poorly in the Iowa caucuses often point donors to the New Hampshire primary saying, “They pick corn in Iowa and presidents in New Hampshire,” emphasizing that Iowa results should not figure in donors’ model of ultimate campaign success. In these examples, an expert “makes the case” using data their audience may already be aware of. The key persuasive element is not the information itself. It is that the expert highlights a relationship between outcomes and variables in a way that logically leads the audience to take an action the expert favors.

This kind of persuasion using models is ubiquitous. In finance, when recent market performance is better than historical performance, bullish traders argue “this time is different”. Stock market analysts use technical analysis to argue that patterns in prices and trading volume identify profit opportunities. In the debate surrounding climate change, one side might argue that extreme weather events provide evidence in favor of global warming, while the other might argue that they largely reflect “noise” in an inherently unpredictable process. In politics, there are “spin rooms” where participants seek to influence interpretations of debate performances. In law, the defense and prosecution build their cases around the same evidence. Recall the famous line from the O.J. Simpson trial that “If it [the glove] doesn’t fit, you must acquit.” In advertising, producers propose frames that positively highlight known aspects of their products. The car-rental company Avis, famously lagging behind Hertz in sales, ran a well-known campaign with the slogan “When you’re only No. 2, you try harder”. When social scientists want to build the case for a particular conclusion, they may draw curves through data points in ways that make the conclusion more visually compelling. (Figure 1 provides a humorous illustration of this point.) Despite the pervasiveness of persuasion using models, economists’ understanding of persuasion (DellaVigna and Gentzkow 2010) has typically focused on the disclosure of information (e.g., Milgrom 1981; Kamenica and Gentzkow 2011) rather than its interpretation.¹

In this paper, we present a formal framework for studying “model persuasion.” We consider the problem of a decision maker or “receiver”, who needs to interpret a history of outcomes that may be informative about a payoff-relevant state of nature before taking an action. Persuaders propose models for interpreting the history to the receiver. A model is a likelihood function that maps the

¹The few exceptions (e.g., Mullainathan, Schwartzstein, and Shleifer 2008) are described in more detail below. There is also work (e.g., Becker and Murphy 1993) studying the idea that persuasion directly operates on preferences.

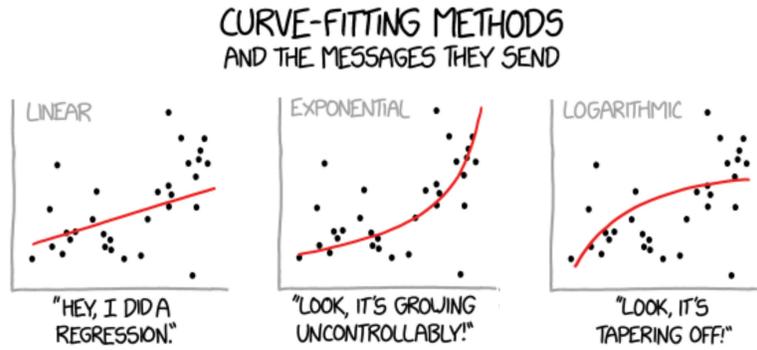


Figure 1: Example of Model Persuasion from xkcd.com
Source: <https://xkcd.com/2048/>

history to posterior beliefs for the receiver, which in turn leads the receiver to take certain actions. The persuader’s incentives are to propose models that generate particular receiver actions, but the persuader cannot influence the data itself. In other words, the persuader helps the receiver make sense of the data. The persuader is constrained to propose models that the receiver is willing to entertain, which we take as exogenous, and that are more compelling in the data than other models the receiver is exposed to, which we endogeneize.

A key ingredient is that we assume a proposed model is more compelling than an alternative when it fits the receiver’s knowledge (i.e., the data plus the receiver’s prior) better than the alternative. Essentially, we assume that the receiver performs a “Bayesian hypothesis test”, adopting from the set of models he is exposed to the one that makes the observed data most likely given his prior. Formally, we assume model m (associated with likelihood function π_m) is more compelling than model m' (associated with likelihood function $\pi_{m'}$) given data h and prior μ_0 over states ω when:

$$\Pr(h|m) = \int \pi_m(h|\omega)d\mu_0(\omega) > \int \pi_{m'}(h|\omega)d\mu_0(\omega) = \Pr(h|m').$$

This assumption loosely corresponds to various ideas from the social sciences about what people find persuasive, including that people favor models which (i) have high “fidelity” to the data as emphasized in work on narratives (Fisher 1985); (ii) help with “sensemaking” as discussed in work on organizational behavior and psychology (Weick 1995; Chater and Loewenstein 2016); and feature the most “determinism”, a property of intuitive theories documented in work on developmental and cognitive psychology (Schulz and Sommerville 2006; Gershman 2018).²

²There is related work in economic theory that assumes people favor explanations that maximize the likelihood of the data, for example Levy and Razin (2019) which analyzes how humans combine expert forecasts. There is also much work that draws out implications of related assumptions, including Epstein and Schneider (2007) which

To illustrate some of the framework’s basic insights, consider a simple example, which we will return to throughout the paper. We consider the problem of an investor deciding whether to invest in an entrepreneur’s new startup, based on the entrepreneur’s past history of successes and failures. As shown in Figure 2a, the entrepreneur’s first two startups failed, and the last three succeeded. The investor’s problem is to predict the probability of success of the sixth startup. The investor’s prior is that the probability of success of that startup, θ , is uniformly distributed on $[0, 1]$. Assume that, in the absence of persuasion, the investor would adopt the default view that the same success probability governs the success of all of the entrepreneur’s startups. Also assume for the purpose of the example that this is the true model.

The persuader wants the investor to invest, and thus wishes to propose models that maximize the investor’s posterior belief about θ . Suppose the receiver is willing to entertain the possibility that “this time is different”. That is, the receiver will entertain models suggesting that the entrepreneur’s success probability changed at some point, so that only the most recent startups are relevant for estimating θ . Assuming these are the only models the receiver will entertain, the persuader will propose the model that the entrepreneur’s last three startups are relevant, but the first two are not. As shown in Figure 2b, under the default model that the success probability is constant over time, the receiver predicts the success probability of the next startup to be 57%. Under the persuader’s proposed model, the receiver instead predicts it to be 80%. Crucially, the persuader’s model is more compelling in the data than the default, true model. The probability of observing the data under the true model is 1.7%, while the probability under the persuader’s model is 8.3%. A likelihood ratio (or, more precisely, Bayes Factor) test would strongly favor the persuader’s model over the true model, and thus the receiver would adopt the persuader’s model.

This simple example illustrates three key intuitions. First, there are times when a wrong model that benefits the persuader will be more compelling than the truth. Second, when the data are quite random under the true model, there will frequently be a wrong model more compelling than the true model. Third, persuasion can generate large biases in the receiver’s beliefs.

A few important assumptions drive the results. First, persuaders are more able than receivers to come up with models to make sense of data. Household investors rely on advisers to help interpret returns data; voters rely on pundits to shed light on polling data about the viability of political candidates; jurors rely on experts and lawyers to interpret evidence at a trial; patients rely on doctors to help interpret medical test results; people need scientists and statisticians to help interpret climate-change data. People may discard certain stories because they “do not hang together”—in our framework, receivers may not be willing to consider every possible model. And they may interpret data through the lens of a default model. But, crucially, receivers do not generate

studies learning under ambiguity; Ortoleva (2012) which studies “paradigm shifts”; and Gagnon-Bartsch, Rabin, and Schwartzstein (2018) which studies when people “wake up” to their own errors.

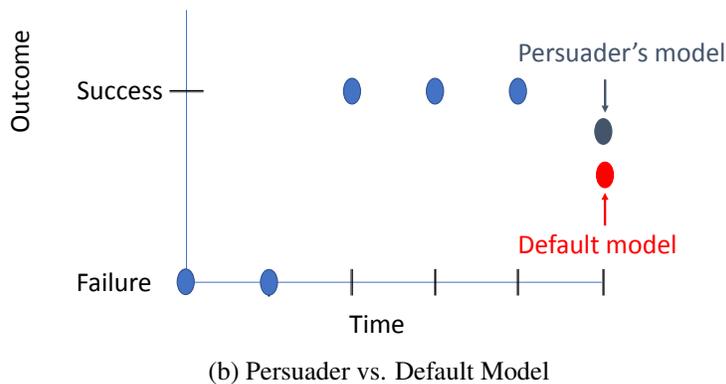
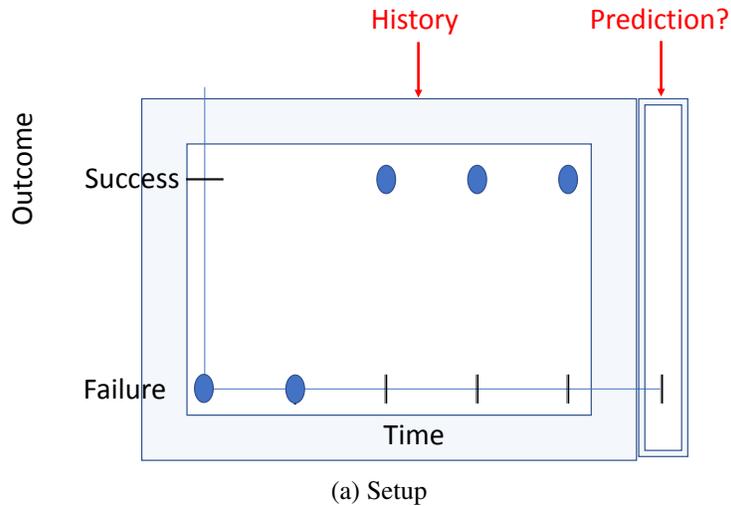


Figure 2: Predicting the success of an entrepreneur's next startup

new stories themselves. Rather, they need experts to supply them.³ Second, because receivers need persuaders to supply models, they do not have a prior over models. Instead, a receiver only judges models by how well they fit the combination of the data and the receiver's prior belief over states. Third, receivers do not discount stories just because they are supplied by biased experts—though they do discount stories if they are not compelling given the facts. As we discuss further below, it is difficult to specify exactly how more sophisticated receivers would behave in this setting. However, our results are qualitatively robust to a simple modification requiring models proposed by more biased experts to satisfy a stricter goodness-of-fit test than models proposed by less biased experts. Finally, receivers do not take into account persuaders' flexibility in proposing models after seeing the facts. Even in the social sciences it is often difficult to fully appreciate the dangers of multiple hypothesis testing, data mining, and data snooping. For example, the movement for

³This is analogous to what makes comedians different from typical audience members. While audience members are able to judge whether a given joke is funny, comedians are better at coming up with jokes.

experimental economists to publicly pre-register hypotheses is relatively recent. Moreover, even when such issues are understood, it is also non-trivial to correct for them: methods in machine learning and economics are still being developed to deal with these issues.

Section 2 sets up our general framework and provides some basic properties. We show that there is a tension between the information and framing components of persuasion. The information component of persuasion—the change in the receiver’s utility from interpreting the data with the true model—is always positive. In contrast, the framing component—the change in the receiver’s utility from using a persuader-proposed model rather than the true model—is always negative. The latter component means that model persuasion may reduce receivers’ welfare relative to if they went with a default interpretation. This is consistent with long-standing worries about the impact of persuasion (e.g., Galbraith 1967) but inconsistent with belief-based persuasion where receivers hold rational expectations (reviewed in, e.g., DellaVigna and Gentzkow 2010).

Section 3 considers two questions: what can receivers be persuaded of and when are they persuadable. Persuaders face a key tradeoff: the better a model *fits* the data plus the receiver’s prior, the less the model *moves* the receiver’s beliefs away from his prior. Intuitively, models that fit well say data is unsurprising, which means beliefs should not move much in response to data. The constraint that a persuader’s model be as compelling as the receiver’s default thus restricts the interpretations of the data the persuader is able to induce. For instance, a persuader is unable to convince a receiver that making a free throw signals that a basketball player is the next LeBron James: making a free throw is common both in reality and under any realistic default interpretation. If it were diagnostic of being the next LeBron James, it would have to be next to impossible, since LeBron Jameses are exceedingly rare. Thus, the “next LeBron James” interpretation is not compelling given the receiver’s knowledge.

Receivers are more persuadable when they have greater difficulty explaining the data under their default interpretation. Hearing someone consistently say “crazy things” opens the door to all sorts of interpretations of the data, including that the person is a genius. Receivers are also more persuadable when they are open to a larger number of different interpretations of the data, i.e., when they are willing to entertain a larger set of possible models.⁴ For both of these reasons, more publicly available data may not limit the impact of persuasion: with more data the receiver’s default interpretation may fit less well, increasing the number of alternative models the receiver finds compelling. For instance, in the example of the entrepreneur above, a longer history benefits the persuader because there are more opportunities to say “this time is different”. Of course, if the receiver is exposed to a lot of data that has only one interpretation, the scope for persuasion based

⁴Receivers may be willing to entertain more models because the available information is “vague” in the sense of Olszewski (2018) or because finding relevant characteristics in large data sets is a challenging task (Aragones, Gilboa, Postlewaite, and Schmeidler, 2005).

on other data that is open to interpretation is limited.

Section 4 asks when the wrong story wins. We consider the impact of model persuasion in the special case where the receiver’s default model is the true model. That is, receivers are exposed to a truth-teller (e.g., a watchdog) and only adopt the persuader’s model when it is more compelling than the truth in the data. One insight from this analysis—illustrated through analytical results as well as simulations—is that persuaders are fairly unconstrained by needing their model to be more compelling than the truth: the wrong story often wins. This is particularly true when data is highly random under the true model (as in financial markets) because it allows the persuader to invite receivers to extract signal from noise. Persuaders also have more space to frame histories that contradict the receiver’s prior under the true model: It is surprising when a prior belief turns out to be incorrect, so receivers will tend to find false models that say the data are consistent with their prior more compelling than a true model that says the data contradicts their prior.⁵

Section 5 then considers the impact of greater competition between persuaders on how receivers interpret data in equilibrium. Competition pushes persuaders to propose models that overfit the data. If a persuader proposes a model that does not fit the data well, this creates space for a competitor to win the battle over models by proposing a better fitting model. Thus, competition pushes persuaders towards models that make the data seem unsurprising in hindsight. Following this logic, a persuader who wants the receiver to hold correct beliefs is often better off proposing an untrue model that leads to those beliefs while overfitting past data—this protects against competing persuaders proposing models that better fit those data.

By leading receivers to adopt models under which the data is unsurprising, competition also leads receivers to underreact to evidence. If the data are not surprising, receivers should not update much in response. In other words, persuasion often neutralizes the data, preventing information from changing minds in equilibrium. This may shed light on why people’s beliefs seem so stubborn in the real world, while they also seem to move a lot in response to individual persuaders (e.g., Broockman and Kalla 2016; Pons 2018). More broadly—and reminiscent of the intuition in Gentzkow and Shapiro (2006)—a persuader is at an advantage when, relative to other persuaders, he does not want to move the audience’s beliefs far from their prior. Models that lead to conclusions that receivers are predisposed to believe are more compelling.

Section 6 asks when persuaders are constrained by needing to send the same message (or menu of messages) to heterogeneous receivers. We show examples where the persuader is able to send individualized, private messages to two receivers that lead each to take a desired action (e.g., to make a desired investment from the perspective of the persuader), but is unable to send common

⁵Conspiracy theories like QAnon (<https://www.vox.com/policy-and-politics/2018/8/1/17253444/qanon-trump-conspiracy-theory-reddit>) provide dramatic illustrations of how successful narratives often overfit. They often try to make sense of random events, for example cryptic comments of politicians that almost surely mean nothing.

public messages that lead both receivers to take the desired action. The key factor is the similarity between the receivers' priors and default interpretations: it is harder to simultaneously persuade dissimilar audience members.

Section 7 illustrates these results with examples in finance and in law.

Related Literature

Our paper is related to several strands of the economics and psychology literatures. At a basic level, many of the logical stories, narratives, analogies, and metaphors people tell themselves are models to make sense of the data (e.g., Lakoff and Johnson 1980; Bruner 1991; Chong and Druckman 2007; Shiller 2017). While even without persuasion people engage in such sensemaking, for example “seeing” non-existent patterns in the data, persuasion impacts the patterns they see (e.g., Andreassen 1990; DiFonzo and Bordia 1997). Indeed, in almost every situation, people are somewhat uncertain about the right model to use, which opens the door for persuaders to encourage the use of models they favor. In this way our model connects to Mullainathan, Schwartzstein, and Shleifer (2008), who consider a situation where one of the ways that persuasion works is through providing advantageous “frames” of known aspects of a product.⁶ In this paper, we provide a more portable and systematic treatment of this idea that goes back at least to Goffman (1974). Model persuaders aim to “make the truth work” for them.

Our paper also connects to contemporaneous formal models of narratives. Benabou, Falk, and Tirole (2018) explore the role of narratives and imperatives in moral reasoning. Eliaz, Spiegler, and Thyssen (2019) study a model akin to ours where persuaders seek to influence receivers' understanding of messages. Barron and Powell (2018) theoretically analyze markets for rhetorical services. While we study what makes messages persuasive, these papers start from the premise that messages are persuasive and seek to understand its implications. In perhaps the closest paper to ours, Eliaz and Spiegler (2019) draw on work from the Bayesian Networks literature to formalize narratives as causal models (directed acyclical graphs) in the context of understanding public-policy debates. While we consider when wrong stories are more compelling in the data than correct stories, they assume that the public favors “hopeful narratives”.

Our paper is also related to the literature on Bayesian Persuasion that begins with Kamenica and Gentzkow (KG, 2011). Persuaders in our model act very differently from persuaders in KG and in generalizations of their framework such as Alonso and Camara (2016) and Galperti (2016). KG's persuaders influence by providing information, fixing the models receivers use to interpret

⁶At a broad level our work connects to a growing literature on how people learn when they follow misspecified models (e.g., Barberis, Shleifer, and Vishny 1998; Eyster and Piccione 2013; Schwartzstein 2014; Acemoglu et al. 2016; Spiegler 2016; Esponda and Pouzo 2016; Heidhues, Koszegi, and Strack 2018; Gagnon-Bartsch, Rabin, and Schwartzstein 2018). While those frameworks take as given the models people follow, ours considers the role of persuasion in promoting misspecified models.

information; ours influence by providing models, fixing the information receivers have access to. The traditional Bayesian framework, including KG, assumes that the receiver is dogmatic that they are using the right model, while our sharpest and most portable analytical results are for the case where the receiver is willing to entertain a rich set of models that roughly includes every interpretation of the data.

Finally, model persuasion is related to, but distinct, from theories of persuasion through cherry picking and slanting (Hayakawa 1939; Milgrom 1981; Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006), which are about selective reporting of facts. Model persuasion can sometimes operate similarly to cherry picking by proposing models where only certain facts are relevant. However, it frequently operates quite differently, by proposing models where consider all the relevant evidence, but interpret it differently than they would under the true model. For instance, a cherry picker notes that 1999 is a great year for corporate earnings, while a model persuader encourages receivers to look at all years and note how different the years 1995-1999 look from all previous years. This distinction is particularly stark with competition. A receiver who faces many cherry-picking persuaders would naturally combine the information provided by each persuader, creating a force towards uncovering the truth (e.g., as reviewed in Gentzkow and Shapiro 2008). In contrast, a receiver who faces many model persuaders goes with the model that best explains the data, which we show creates a force for neutralizing the data.

2 Model Persuasion

2.1 General Setup

Persuaders are interested in influencing the beliefs of receivers, whose beliefs in turn depend on both the past history of outcomes, as well as the models this history to make forecasts about future outcomes. We start by considering the situation with a single persuader and a single receiver, where the receiver only has access to two models: a default model and the model proposed by the persuader. We later consider competing persuaders.

Broadly, the setup is as follows. The persuader proposes a model to the receiver. If the receiver finds the proposed model more compelling than the default model, meaning that the proposed model better explains available data, then the receiver adopts it. The receiver consequently updates his beliefs about the state of the world using the adopted model and takes an action that maximizes his utility given those beliefs. The persuader's aim is to propose a model that induces the receiver to take an action that maximizes the persuader's utility rather than the receiver's.

Formally, given beliefs over states of the world ω in finite set Ω , the receiver chooses action a

from compact set A to maximize $U^R(a, \omega)$.⁷ The persuader tries to alter the receiver's beliefs about ω to induce the receiver to take an action that maximizes $U^S(a, \omega)$. The persuader and receiver share a common prior $\mu_0 \in \text{int}(\Delta(\Omega))$ over Ω .

Both the persuader and receiver observe a history of past outcomes, h , drawn from finite outcome space H . Given state ω , the likelihood of h is given by $\pi(\cdot|\omega)$. The true model m^T is the likelihood function $\{\pi_{m^T}(\cdot|\omega)\}_{\omega \in \Omega} = \{\pi(\cdot|\omega)\}_{\omega \in \Omega}$. We assume that every history $h \in H$ has positive probability given the prior and the true model. The persuader knows the true model m^T and after observing the history uses Bayes' rule to update his beliefs to μ_h . The receiver, on the other hand, does not know the true model. He either (i) updates his beliefs based on a default model $\{\pi_d(\cdot|\omega)\}_{\omega \in \Omega}$, which is potentially a function of h or (ii) updates his beliefs based on a model m proposed by the persuader to organize the history, where m is taken from compact set M (unless we state otherwise) and indexes a likelihood function $\{\pi_m(\cdot|\omega)\}_{\omega \in \Omega}$.

Given the history and model proposed by the persuader, the receiver adopts the persuader's model if it better explains the history than the default model. Formally, let $\mu(h, \tilde{m})$ denote the posterior distribution of ω over Ω given h and model $\tilde{m} \in M \cup \{d\}$, as derived by Bayes' rule. We assume the receiver adopts the persuader's model m and hence posterior $\mu(h, m)$ if

$$\underbrace{\Pr(h|m)}_{=\int \pi_m(h|\omega) d\mu_0(\omega)} > \underbrace{\Pr(h|d)}_{=\int \pi_d(h|\omega) d\mu_0(\omega)} \quad (1)$$

and adopts the default model and hence posterior $\mu(h, d)$ if the inequality is reversed. We assume that in the case of a tie the receiver goes with the default model. We sometimes allow the persuader to "say nothing", $m = \emptyset$, which is formally equivalent to assuming $d \in M$.

Upon adopting a model \tilde{m} , the receiver uses Bayes' rule to form posterior $\mu(h, \tilde{m})$ and takes an action that maximizes his expected utility given that posterior belief:

$$a(h, \tilde{m}) \in \arg \max_{a \in A} \mathbb{E}_{\mu(h, \tilde{m})} [U^R(a, \omega)],$$

breaking ties in favor of the persuader and choosing an arbitrary action if there are remaining ties.

The persuader proposes a model to induce the receiver to take an action that the persuader favors, solving⁸

$$m(h) \in \arg \max_{m \in M} \mathbb{E}_{\mu_h} [U^S(a(h, m), \omega)],$$

subject to (1). The persuader breaks ties involving the true model in favor of that model.

A few points about the default model merit discussion. First, we allow the default to be history

⁷In applications, we will sometimes relax the assumption that Ω is finite.

⁸Note that when $U^S(\cdot)$ is independent of ω then the persuader's optimal action is independent of the true likelihood π . In these situations, it is without loss of generality to assume the persuader knows the data-generating process.

dependent to capture the idea that a person might only come up with a default explanation for the data after seeing the data. Second, absent competition, we primarily analyze two “extremes” for the default. At one extreme, the default renders the data uninformative. Under such an *uninformative default*, the receiver sticks with his prior in the absence of persuasion (i.e., $\mu(h, \emptyset) = \mu_0 \forall h$) and finds any model in M more compelling than the default.⁹ This extreme is more likely to hold in situations where the receiver would ignore data in the absence of persuasion because he would be at a loss to interpret it. For example, a patient often requires a doctor’s guidance to interpret medical test results. At the other extreme, which we analyze in Section 4, the default is the true model. This extreme is more likely to hold in situations where the true model is readily accessible, perhaps because there are academics or watchdogs actively pushing it. Third, we try to emphasize lessons that hold across a range of assumptions about the default.

2.2 Examples

Example 1 (Highlighting strips of data). We now sketch two brief examples to show how they map into the general framework. Our first example involves highlighting strips of data. The setup captures the entrepreneurship example from the introduction, in addition to a variety of other situations in finance. For instance, as described by Kindleberger and Aliber (2010), the history of the technology bubble in the late 1990s fits the setup:

The authorities recognize that something exceptional is happening and while they are mindful of earlier manias, ‘this time it’s different’, and they have extensive explanations for the difference. The Chairman of the US Federal Reserve, Alan Greenspan, discovered a surge in US productivity in 1997... the increase in productivity meant that profits would increase at a more rapid rate, and the higher level of stock prices relative to corporate earnings might not seem unreasonable.

The notion that a structural shift had occurred, allowing the “new economy” to grow very rapidly due to technology, was popularized in part by financial analysts with incentives that rewarded high stock prices.¹⁰

To put such examples in the notation of the general framework, suppose there is a coin (investment) that yields heads (success) with probability $\theta \in (0, 1)$, where θ is drawn once and for all at the beginning of time from a density ψ that is strictly positive over $[0, 1]$. Suppose the persuader knows that the probability of heads is constant over time, but the receiver is willing to entertain the

⁹Under the uninformative default, d is a function of h and has the feature that $\pi_{d(h)}(h|\omega) = \varepsilon \approx 0$ for all h, ω .

¹⁰As one analyst described, “It’s nothing short of a revolution. The realities of the technology revolution will be realized in both the U.S. and Canada beyond the year 2000, which is going to lead to a sharp acceleration in earnings growth across the board.”<http://money.cnn.com/1999/04/30/economy/neweconomy/>

possibility that it was drawn again from ψ at some date. In the notation of the general model, the state space is $\Theta = [0, 1]$ and the prior is ψ . We assume that the receiver has incentives to correctly estimate the success probability and hence uncover the correct value of θ , while the persuader wants to inflate its value. Formally, the receiver’s payoff is given by $U^R(a, \theta) = -(a - \theta)^2$, and the persuader’s is given by $U^S(a, \theta) = a$.

The persuader can propose models of the form: “the last K periods are relevant for whether the next flip comes up heads”. Denote these K -models, where for example the 1-model is the model where only the last flip matters. The persuader is restricted to only propose models that state at least the last \underline{K} flips matter. The case where $\underline{K} = 0$ corresponds to the persuader being able to “say nothing” when the default is uninformative.

If the persuader proposes the K -model, where S of those K flips came up heads, and the receiver adopts it, then his belief the next flip will be heads with approximately (in the sense of Diaconis and Freedman 1990) probability S/K .¹¹ For example, if $\psi = \text{Uniform}[0, 1]$ and m is the K -model, then the receiver’s posterior expectation of the probability of heads is $\hat{\theta} \equiv \mathbb{E}[\theta|m, h] = (S + 1)/(K + 2)$. The persuader chooses the model that maximizes $\hat{\theta}$ subject to (1).

Example 2 (Highlighting characteristics). Our second example involves highlighting characteristics. For instance, Tesla is both a technology company and an automotive company. Persuaders may propose models in which one or both characteristics are relevant.

To capture such an example, suppose that a person assesses the likelihood that an actor (e.g., a business, investment, worker, politician) will be successful ($y = 1$) or not ($y = 0$), where the actor has characteristics x taken from finite set X . The true likelihood that the actor is successful is given by probability $\theta(x)$, where $\theta(x)$ is drawn from strictly positive density $\psi(\cdot)$ on $[0, 1]$. In the notation of the general model, the state space is $\Theta(x) = [0, 1]$ and the prior is ψ . We again assume the receiver is interested in correctly assessing the success probability, while the persuader wants to inflate it: $U^R(a, \theta) = -(a - \theta(x))^2$, while $U^S(a, \theta) = a$.

Both the persuader and the receiver observe a history $h = (y_k, x_k)_{k=0}^{t-1}$ of successes and failures of previous actors with various characteristics. The persuader can influence the probability the receiver attaches to the actor being successful by proposing models of which characteristics are relevant to success. Models group together actors with particular characteristics and assert that these actors all have the same success probability. In effect, the models are partitions of X , where x and x' share the same success probability if they are in the same element of the partition.¹² For

¹¹Suppose the total history is length t and contains l total heads, and the persuader proposes the K -model, where S of those K flips came up heads. Then the likelihood function is given by

$$\Pr(h|K \text{ model}) = \left(\int_0^1 \theta^S (1 - \theta)^{K-S} \psi(\theta) d\theta \right) \cdot \left(\int_0^1 \theta^{l-S} (1 - \theta)^{t-l-K+S} \psi(\theta) d\theta \right).$$

¹²Formally, we write $c_m(x)$ to denote the element of the partition that contains x under model m . To illustrate,

instance, a persuader aiming to inflate Tesla’s value would propose a partition grouping Tesla with technology companies. We analyze this example in Appendix A.

2.3 Additional Discussion of Assumptions

The key assumptions of the model were discussed in the introduction. Some other assumptions are worth discussing. First, as in the vast cheap talk literature that begins with Crawford and Sobel (1982), we assume that the persuader’s incentives can differ from the receiver’s. As noted by Shiller (2015), “There are still many people (indeed, the stock brokerage and mutual fund industries as a whole) who benefit from telling stories that suggest that the market will go up further.” Mutual funds want to drum up business. Analysts at banks with underwriting relationships with firms they are reporting on may want to induce positive beliefs about those firms. A politician could find it advantageous to stump for a measure that is not best for his constituents.

Second, while the receiver does not know how to interpret data, he does have a prior over payoff-relevant parameters. Even a casual investor may understand that a mutual fund cannot be expected to outperform the market 100% of the time. Similarly, a voter is unlikely to place high probability on a third party candidate winning a presidential election no matter how the candidate invites the receiver to interpret polls.

Third, the receiver is naive in the sense that he does not take the persuader’s incentives into account when reacting to proposed models. We make the naivete assumption for a few reasons: First, we think it is broadly realistic, as much evidence suggests that receivers underreact to persuader’s incentives (e.g., Malmendier and Shanthikumar, 2007; DellaVigna and Kaplan, 2007; Cain, Loewenstein, and Moore, 2005). Second, it more sharply captures the idea that receivers do not know how to interpret data without the help of a persuader than would alternatives that assume sophistication. Third, it makes the model quite transparent and tractable.

That said, receivers are unlikely to take everything persuaders say at face value. Section 4 explores the case where persuaders have to compete with truth-tellers, such as consumer watchdogs, and receivers only adopt a persuader’s proposed model when it is more compelling than the truth. We show how our results carry over to this case. Our results are also qualitatively robust to modifying Eq. (1) so that persuaders’ whose incentives are known to be misaligned with receivers’ must satisfy tighter constraints. This will tend to reduce receivers’ sensitivity to all data, so receiver skepticism may in fact backfire, while still leaving room for misleading persuasion. We show this formally in Appendix B.

if each x is described as a vector of attributes, $x = (x_1, x_2, \dots, x_J)$, and m is a model where only the first three attributes are relevant to success, then $c_m(\tilde{x}) = \{x \in X : (x_1, x_2, x_3) = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)\}$. If the receiver adopts model m , the success probability he ascribes to each element of the partition is based on the number of successes in the data within that element of the partition.

Fourth, we assume that receivers select rather than average models. This is consistent with some psychological evidence on “thinking through categories”, for example as discussed in Mulainathan (2002). It is also natural given we assume that receivers do not have a prior over models.

Finally, our assumption that receivers adopt models that fit the data well, embodied in Eq. (1), drives many of our results. Our formulation is formally equivalent to the receiver starting from a flat “prior” over the models he is exposed to, with models he is not exposed to getting zero weight, and then selecting the model that has the greatest posterior probability.¹³ This notion of fit is independent of the receiver’s incentives. An alternative, “value-adjusted” notion of fit takes into account the impact of adopting a model on the receiver’s decisions. We briefly analyze this alternative in Appendix B.¹⁴

The idea that people find stories compelling when they explain the existing data well is intuitive and related to evidence (briefly described in the introduction) from psychology and the broader literature on what makes narratives or theories persuasive. In addition, it is consistent with the degree to which people “see” patterns in the data, especially with the help of stories (e.g., Andreassen 1990; DiFonzo and Bordia 1997). There are articles questioning whether we *should* find a good model fit persuasive (e.g., Roberts and Pashler 2000), but as far as we can tell little debate that we *do* find a good fit persuasive.

2.4 Basic Framework Properties

In our setup, persuasion has two effects. First, it enables receivers to act on more information. Second, it frames that information.

Let

$$V^R(h, m) \equiv \mathbb{E}_{\mu_h} [U^R(a(h, m), \omega)]$$

denote the receiver’s true payoff (computed using posterior beliefs μ_h on the distribution of ω under the true model after observing history h) as a function of the action a taken because the receiver observed history h and adopted model m . Similarly, let $V^S(h, m) \equiv \mathbb{E}_{\mu_h} [U^S(a(h, m), \omega)]$ denote the sender’s payoff as a function of the history and adopted model.

¹³As is well known from the literature on Bayesian Model Selection (e.g., Kass and Raftery 1995), our formulation then does not necessarily favor “more complex” models. For example, after seeing an entrepreneur fail two times, a two-parameter model where the entrepreneur’s success probability is independently drawn each period from a uniform distribution over $[0, 1]$ fits worse than a one-parameter model where the success probability is drawn once and for all: $\Pr(2 \text{ failures} | 2 \text{ parameter model}) = 1/4 < 1/3 = \Pr(2 \text{ failures} | 1 \text{ parameter model})$.

¹⁴This alternative favors models that prescribe actions that have greater utility consequences when those models are true. For example, according to the notion of value-adjusted fit, the receiver may find a model compelling when it says the evidence supports taking a low-cost herbal treatment to prevent cancer even if it doesn’t fit the data plus prior as well as a model that says the treatment is ineffective: Failing to take the pill when the model is true has far greater utility consequences than taking the pill when the model is not true.

Define the *impact of persuasion* for agent j as the expected change in j 's payoff: $\mathbb{E}[V^j(h, m(h)) - V^j(h, d)]$, where $\mathbb{E}[\cdot]$ is taken over the true distribution of histories h , i.e., with respect to the prior and the true likelihood (μ_0, π) . The impact of persuasion on the receiver is positive if in expectation he is made better off by being exposed to data plus persuasion, relative to only having access to data and his default model. We can break it into two terms:

$$\underbrace{\mathbb{E}[V^R(h, m^T) - V^R(h, d)]}_{\text{information component}} + \underbrace{\mathbb{E}[V^R(h, m(h)) - V^R(h, m^T)]}_{\text{framing component}}.$$

The first term, the information component, is the value to the receiver of operating under the true model rather than the default model. This component is the one typically emphasized in the economics literature, namely that persuasion allows the receiver to make more informed decisions. For example, when d is the uninformative default, the information component is equivalent to the impact of acting on correctly-interpreted information. The second term, the framing component, is the value to the receiver of operating under the persuader's proposed model rather than the true model. This is the more novel feature of our framework and captures the idea that persuasion also influences how the receiver reacts to publicly available data.

The following are some basic framework properties. (All proofs are in Appendix C.)

Observation 1 (Model Properties).

1. The information component of persuasion is positive for the receiver: $\mathbb{E}[V^R(h, m^T) - V^R(h, d)] \geq 0$.
2. Assume the persuader can propose the true model, $m^T \in M$, and $\Pr(h|m^T) \geq \Pr(h|d)$. The framing component of persuasion is positive for the persuader and negative for the receiver: $\mathbb{E}[V^j(h, m(h)) - V^j(h, m^T)]$ is positive for $j = S$, negative for $j = R$, and strictly positive for $j = S$ if and only if it is strictly negative for $j = R$.

The first property is that the information component of persuasion is positive for the receiver: the receiver clearly cannot be made worse off on average by using the true model instead of the default model. The second property is that the framing component of persuasion is positive for the persuader and negative for the receiver whenever the persuader could get the receiver to adopt the true model. It is positive for the persuader because he always has the option of proposing the true model and will only propose a different model when it improves his payoff; it is negative for the receiver because she cannot be better off acting on the wrong model instead of the true model. We will see below that the premise that the persuader can get the receiver to adopt the true model is substantive: there are natural cases (e.g., where receivers are prone to overfitting data on their own) where default models fit better than the true model.

As we will also see below, the two components of persuasion go hand in hand. Persuasion can simultaneously benefit receivers relative to their potentially-incorrect default models, while making them worse off relative to the true model. Thus, our model provides a framework for thinking about long-standing concerns on negative consequences of persuasion (e.g., Galbraith 1967), while also showing that receivers are not necessarily led astray by persuasion.

3 What Can Receivers be Persuaded of and When Are They Persuadable?

In this section, we consider what receivers can be persuaded of and when they are persuadable. To illustrate some key intuitions, we start by considering the following simple example. Pat is considering investing in an actively managed mutual fund. The active fund is either good, meaning that future returns will be high (better than a passive index fund alternative), or bad, meaning they will be low (worse than a passive alternative). A broker would like to convince Pat to invest in the active fund, and therefore wants to convince him that it is likely to be good. Pat's prior is that the probability of the fund being good is 20%—think of this as being pinned down by the empirical distribution of historical fund returns across all funds,¹⁵ and he will invest only if his belief moves to at least 50%. For simplicity, suppose the only data available to Pat and the broker is whether the active fund had high returns or not last year—think of this as a restriction on the set of models M ; Pat is unwilling to believe that other data (e.g., returns from prior years, the fund manager's educational background) is relevant. Pat's default model is that past returns are somewhat informative. He believes that good funds have a higher probability of high past returns than bad funds: $\pi_d(\text{high returns}|\text{good}) = \pi_d(\text{low returns}|\text{bad}) = 75\%$. Thus, under the default model, Pat believes the probability that a given fund delivered high returns last year is

$$\Pr(\text{high returns}|d) = \pi_d(\text{high returns}|\text{good}) \times 20\% + \pi_d(\text{high returns}|\text{bad}) \times 80\% = 35\%.$$

Suppose that the active fund Pat and the broker are considering has high past returns. Under his default model, Pat will still not invest in the fund because

$$\Pr(\text{good}|\text{high returns}, d) = \pi_d(\text{high returns}|\text{good}) \frac{\mu_0(\text{good})}{\Pr(\text{high returns}|d)} = 75\% \frac{20\%}{35\%} = 43\%.$$

Intuitively Pat thinks that good active funds are unconditionally too rare, and high past returns are not informative enough about the quality of the fund. What can the broker convince Pat of?

¹⁵24% of all actively managed have outperformed their passive benchmarks over the last 10 years (Morningstar Active/Passive Barometer, 2019).

First, note that broker cannot get Pat to believe anything she wants. For instance, she cannot simply assert that the percentage of good funds is much higher than 20%, i.e., that Pat’s prior is incorrect, without referencing the data. Our model is about framing the data; Pat’s beliefs only change in response to data, framed by the broker.

Even though the broker has great flexibility to frame the data, a key intuition is that Eq. (1) limits how much Pat’s beliefs can change in response to the data. For instance, suppose the broker tries to convince Pat that the high past returns mean the active fund is good for sure: $\pi_m(\text{high returns}|\text{good}) = 100\%$, $\pi_m(\text{high returns}|\text{bad}) = 0$. Under this model, the probability of high past returns is

$$\Pr(\text{high returns}|m) = \pi_m(\text{high returns}|\text{good}) \times 20\% + \pi_m(\text{high returns}|\text{bad}) \times 80\% = 20\%,$$

which is less than the probability of high returns under Pat’s default model. If the broker tries to tell Pat that high past returns are strongly associated with a relatively rare event (a good active fund), then Pat finds his default model—that high past returns are not that rare but also not perfectly informative about the quality of the fund—more compelling. Intuitively, under his default, Pat thinks high returns are less rare than good funds. Therefore, the default provides a better explanation of seeing high returns than an alternative that says only good funds have high returns.

To beat the default model, the broker must propose a model where

$$\Pr(\text{high returns}|m) = \pi_m(\text{high returns}|\text{good}) \times 20\% + \pi_m(\text{high returns}|\text{bad}) \times 80\% > 35\%.$$

If she proposes a model with $\pi_m(\text{high returns}|\text{good}) = 100\%$, this means she must set $\pi_m(\text{high returns}|\text{bad})$ above 18.75%. Under the most favorable such model to the broker, $\Pr(\text{good}|\text{high returns}) = 57\%$, and Pat will invest in the active fund. Intuitively, the broker can propose a model that, relative to the default, provides a better explanation of seeing a given fund have high past return, while implying a stronger relationship between returns and fund quality. Such a model acknowledges that high returns do not imply that the fund is necessarily good, but says this is exactly what you would expect to see if the fund is good. To Pat, this is a better explanation than his default model—it makes it less surprising that he is seeing an active fund with high returns, and it is an explanation that leads him to invest.

A second key intuition is that Pat’s prior restricts the stories that will resonate with him and thus the actions he will take. Imagine that Pat is more pessimistic about active funds: his prior is that 10% of active funds are good instead of 20%. Then there is no model that the broker can propose that gets Pat to invest. If Pat believes good active funds are very rare, he will not find any model that says high past returns are informative enough to dictate investing in the fund more compelling than his default. Given such a low prior on the likelihood of good active funds, such a

model will provide a relatively poor explanation of seeing a high past return.

A third key intuition is that the broker has more flexibility when the data is unusual under the default model. For instance, suppose that past returns can either be low, high, or very high. Further, suppose that Pat’s default model says that very high returns are no more informative than high returns of fund quality, but are rarer. Now, if the active fund has very high returns, the broker can convince Pat that this is perfectly diagnostic of the fund being good.¹⁶ Because Pat’s default does not explain the occurrence of very high returns well, he is open to alternative explanations of the data. He finds the broker’s alternative model—that very high returns are less rare than he thinks and diagnostic of good active funds—to be more compelling than his default.

The next two subsections explore in greater detail what receivers can be persuaded of and when they are persuadable.

3.1 What Can Receivers Be Persuaded Of?

We saw in the above example that receivers cannot be persuaded of everything, even when the space of models they are willing to consider is rich. What can they be persuaded of?

Persuaders face a basic tradeoff between how well a model *fits* the data plus the receiver’s prior and how much the model causes *movement* in the receiver’s beliefs in response to the data. Formally, let

$$\text{Fit}(\tilde{\mu}; h, \mu_0) \equiv \max_m \Pr(h|m, \mu_0) \text{ such that } \mu(h, m) = \tilde{\mu}$$

denote the maximum fit of *any* model, i.e., the maximum across all likelihood functions, that induces posterior belief $\tilde{\mu}$ given data h . Fit varies between 0—the data is impossible under any model that induces $\tilde{\mu}$ —and 1—the data is inevitable under a model that induces $\tilde{\mu}$. Further, let

$$\text{Movement}(\tilde{\mu}; \mu_0) \equiv \max_{\omega \in \Omega} \tilde{\mu}(\omega) / \mu_0(\omega)$$

be a measure of the change in beliefs from prior μ_0 to posterior $\tilde{\mu}$. Movement varies between 1 ($\tilde{\mu} = \mu_0$) and ∞ ($\tilde{\mu}$ places positive probability on a state that prior μ_0 says is zero probability).

Lemma 1. *Fixing history h , $\text{Fit}(\tilde{\mu}; h, \mu_0) = 1/\text{Movement}(\tilde{\mu}; \mu_0)$.*

¹⁶Formally, Pat’s default is given by: $\pi_d(\text{very high returns}|\text{good}) = 15\%$, $\pi_d(\text{high returns}|\text{good}) = 60\%$, $\pi_d(\text{low returns}|\text{good}) = 25\%$, $\pi_d(\text{very high returns}|\text{bad}) = 5\%$, $\pi_d(\text{high returns}|\text{bad}) = 20\%$, $\pi_d(\text{low returns}|\text{bad}) = 75\%$. Relative to the previous version with only high and low returns, we have modified 1/5 of the occurrences of high returns to be very high returns when the fund is both good and bad. That is, we split $\pi_d(\text{high returns}|\text{good}) = 75\%$ into $\pi_d(\text{very high returns}|\text{good}) = 15\%$, $\pi_d(\text{high returns}|\text{good}) = 60\%$. And we split $\pi_d(\text{high returns}|\text{bad}) = 25\%$ into $\pi_d(\text{very high returns}|\text{bad}) = 5\%$, $\pi_d(\text{high returns}|\text{bad}) = 20\%$. The alternative model that good funds always reveal themselves with very high returns and bad funds never deliver such returns (i.e., $\pi_m(\text{very high returns}|\text{good}) = 100\%$, $\pi_m(\text{very high returns}|\text{bad}) = 0$) is more compelling than the default: $\Pr(\text{very high returns}|m) = 20\% > \Pr(\text{very high returns}|d) = 7\%$.

Intuitively, when the data fit a particular model well, the data are not surprising under that model. But if the data are not surprising, they are not very informative, and thus cannot move beliefs far from the prior. On the other hand, any model that leads beliefs to react a lot to the data cannot fit the data well. In other words, if the persuader needs to fit the data well, she is constrained to propose models that induce beliefs that are close to the receiver’s prior. This constraint pushes persuaders towards models that feature a kind of hindsight bias (Fischhoff 1975). Models that fit well say the past was completely unsurprising given prior beliefs, implying that those beliefs should not move.

An implication is that the requirement that the persuader’s proposed model fits the data better than the receiver’s default model places restrictions on beliefs the persuader is able to induce. To clarify these restrictions, it is useful to characterize the set of beliefs the persuader is able to induce, independent of exogenous constraints on the set of models M the receiver is willing to entertain.

Definition 1. Receivers are *maximally open to persuasion* when M is such that for any likelihood function $\{\tilde{\pi}(\cdot|\omega)\}_{\omega \in \Omega}$, there is an $m \in M$ with $\{\pi_m(\cdot|\omega)\}_{\omega \in \Omega} = \{\tilde{\pi}(\cdot|\omega)\}_{\omega \in \Omega}$.

Being maximally open to the persuasion means that the set of models the receiver is willing to believe is large and flexible enough that any likelihood function over histories can be implemented. It is of course unrealistic to assume that receivers are maximally open to persuasion. We develop results for this case because it clarifies constraints derived from the requirement that models are compelling in the data. We also think—and to some extent show in simulations below—that the comparative statics derived assuming that receivers are maximally open to persuasion likely extend to more realistic situations where receivers entertain only a subset of possible models.

Proposition 1. Fix d , μ_0 , and h . There is a model space M under which the persuader is able to induce target belief $\tilde{\mu} \in \Delta(\Omega)$ if and only if

$$\tilde{\mu}(\omega) < \frac{\mu_0(\omega)}{\Pr(h|d)} \forall \omega \in \Omega. \quad (2)$$

Remark 1. Equivalently, assume the receiver is maximally open to persuasion and fix d , μ_0 , and h . The persuader is able to induce target belief $\tilde{\mu} \in \Delta(\Omega)$ if and only if (2) holds.

Remark 2. Letting $m(\mu)$ be the best-fitting model that induces belief μ , this result also trivially implies the following: Fix d , μ_0 , and h . The persuader is able to induce target belief $\tilde{\mu} \in \Delta(\Omega)$ if $m(\tilde{\mu}) \in M$ and (2) holds.

Proposition 1 follows directly from the lemma on fit versus movement. The better the default model fits the data, the more constrained the persuader is because the persuader must propose a model that fits the data even better. And the better the model fits the data, the less the persuader

is able to convince the receiver in an *ex ante* unlikely state ω . Remark 1 notes that this intuition applies exactly when the set of models the persuader can propose is completely flexible. Remark 2 supplies a partial characterization of which beliefs are implementable when the set of models the persuader can propose is restricted.

Returning to the example of Pat and the active mutual fund, what is the maximum fit of any model that for sure induces Pat to invest if he adopts it? From Lemma 1, we know that $\text{Fit} = 1/\text{Movement}$. Since the target belief $\mu(\text{good})$ is 50% and the prior is $\mu_0(\text{good})$ is 20%, movement equals $.5/.2 = 5/2$. This means by the lemma that the maximum fit is $\Pr(\text{high returns}) = 2/5 = 40\%$. To induce this belief, the broker argues that $\pi_m(\text{high returns}|\text{good}) = 1$ and $\pi_m(\text{high returns}|\text{bad}) = (40\% - 1 \times 20\%)/80\% = 25\%$. That is, the broker argues that high past returns are exactly what one would expect to see from this fund if good, but are fairly likely either way. The broker will only persuade Pat to invest using this argument when Pat's default does not well explain high returns: $\Pr(\text{high returns}|d) < 40\%$.

So even when the broker uses the best possible argument and Pat is willing to entertain any model, the analysis in this section shows that he is only persuadable to invest under certain conditions: namely, when the fit of Pat's default model is under 40%. For this to be the case, then Pat must first believe that there is data (e.g., past returns) relevant to predicting whether the active fund's future returns will be high. Under his default, the probability of the particular realization of this data observed must also be sufficiently low. These conditions are more likely to be met if he considers a lot of data relevant, is open to different interpretations of this data, and views such data as generated randomly. They are less likely to be met if his default overfits, for instance if he comes up with his own explanations for why the data confirms his prior.

3.2 When Are Receivers Persuadable?

More broadly, there are four major factors that influence the scope for persuasion:

1. The difficulty receivers have explaining the data under their default interpretation.
2. The (ex ante) expected difficulty receivers will have explaining the data under their default interpretation, which in natural cases is increasing in the randomness inherent in the data given the true process.
3. The degree to which data is open to interpretation.
4. The amount of unambiguous (i.e., closed-to-interpretation) data available to receivers, relative to the amount the amount of ambiguous (i.e., open-to-interpretation) data available.

To illustrate each of these points, we make use of the following definition: We say persuasion *is ineffective at history h given default d* when $\Pr(h|d) \geq \Pr(h|m)$ for all $m \in M$ satisfying $V^S(h, m) > V^S(h, d)$. That is, persuasion is ineffective when the persuader is unable to convince the receiver of any interpretation of the data more favorable to the persuader than the receiver's default interpretation.

The first factor affecting persuadability is how well the receiver's default model fits the history. Holding fixed history h , consider defaults d and d' such that d' fits the history better but induces the same posterior: $\Pr(h|d') > \Pr(h|d)$ and $\mu(h, d') = \mu(h, d)$. If persuasion is ineffective at history h given default d , then it is also ineffective at h given default d' . This means that when receivers have defaults that overfit the data, they are hard to persuade. For instance, academics and benevolent financial advisers have a hard time convincing individual investors that stock returns are unpredictable because individual investors falsely perceive patterns in stock prices.

The analysis is similar across histories. Under conditions we will make precise below, if persuasion is ineffective at history h given default $d(h)$, then it is also ineffective at any history \tilde{h} that given default $d(\tilde{h})$ that induces the same beliefs and fits better: $\mu(\tilde{h}, d(\tilde{h})) = \mu(h, d(h))$ and $\Pr(\tilde{h}|d(\tilde{h})) > \Pr(h|d(h))$. For instance, receivers are more persuadable that an abnormally cold month signals a hiatus in global warming than that an abnormally warm month signals a hiatus. A cold month poorly fits the default model that global warming is taking place, creating space for the persuader to propose an alternative. Similarly, in the US, a persuader would find it very difficult to convince a receiver that red traffic lights mean go because the default model that red traffic lights mean stop (together with knowledge of the law, incorporated in the prior) fit the data very well. In contrast, the default model that the speed limit of 55 is followed fits the data less well.

Proposition 2. *Suppose persuasion is ineffective at history h given default d and prior μ_0 .*

1. *Persuasion is also ineffective at history h given default \tilde{d} and prior μ_0 , assuming (i) d induces the same posterior belief as \tilde{d} : $\mu(h, \tilde{d}) = \mu(h, d)$ and (ii) \tilde{d} fits h better than d fits h : $\Pr(h|\tilde{d}, \mu_0) > \Pr(h|d, \mu_0)$.*
2. *Persuasion is also ineffective at history \tilde{h} given default \tilde{d} and prior μ_0 , assuming (i) receivers are maximally open to persuasion, (ii) h given d induces the same posterior belief as \tilde{h} given \tilde{d} : $\mu(\tilde{h}, \tilde{d}) = \mu(h, d)$, and (iii) \tilde{d} fits \tilde{h} better than d fits h : $\Pr(\tilde{h}|\tilde{d}, \mu_0) > \Pr(h|d, \mu_0)$.*

This proposition formalizes the two way that increasing the fit of receivers' default interpretations reduces how persuadable they are. With a bit more structure, similar intuitions also apply if we modify the prior to change how well the same default interpretation fits the data. When the data and default interpretation imply something that the receiver viewed as ex ante unlikely, there is more space for persuasion. For instance, it is easier to persuade a voter that a worse gaffe by

the candidate he thought was competent is meaningless than a smaller gaffe by the candidate he thought was incompetent.¹⁷

The second factor affecting persuadability is the expected (ex ante, prior to h being realized) difficulty receivers will have explaining the data under their default interpretation.¹⁸ Receivers are persuadable that this time is different when interpreting financial market data (they are often puzzled by what they see); they are not when considering whether the sun will rise tomorrow (they always have ready explanations for what they see). Similarly, receivers find technical analysis compelling in interpreting prices and trading volumes in financial markets; they would not when explaining patterns in their bank-account balances.

The third factor affecting persuadability is how open to interpretation the data is. If persuasion is ineffective at history h and default $d(h)$ given model space M , then it is ineffective for any $M' \subset M$. The receiver's openness to different models for interpreting data creates space for misleading persuasion. When signals have a natural interpretation, the persuader cannot do much to change minds; vague signals (Olszewski 2018), on the other hand, are ripe to be framed. For example, the persuader would have a hard time convincing an audience that, all else equal, being older reduces mortality risk. Audiences are only willing to entertain models that suggest that mortality risk rises with age. On the other hand, a persuader has more wiggle room to convince an audience that consuming a specific food reduces mortality risk because audiences are willing to entertain a large set of models relating diet to mortality.

The fourth factor affecting persuadability is the amount of available data that is or is not open to interpretation. Suppose the history is comprised of unambiguous data, for which the receiver will only entertain the true-model interpretation, and ambiguous data, for which the receiver will entertain many interpretations. For concreteness, suppose $h = (h_1, h_2)$, and any model in the space M that the receiver is willing to entertain is representable as $\pi_m(h|\omega) = \pi_{m^T}(h_1|\omega)\pi_m(h_2|\omega)$. Here, h_1 is the data that is unambiguous and h_2 is the data that is ambiguous. In such cases, the scope for persuasion depends on the relative amounts of unambiguous and ambiguous data. If unambiguous data h_1 pins down the state, then persuasion is ineffective no matter h and $d(h)$. In contrast, if we fix unambiguous data that does not pin down the state, then for "enough" ambiguous data (i.e., where $\Pr(h_2|d)$ is sufficiently low) the persuader is able to induce any target belief

¹⁷As an illustration, suppose there are binary states, $\Omega = \{0, 1\}$, and the persuader's payoff equals $v > 0$ if $\mu(1) \geq k > 0$ and equals 0 otherwise. If persuasion is ineffective at history h given default d and prior μ_0 , then it is also ineffective at history \tilde{h} given default \tilde{d} and prior $\tilde{\mu}_0(1) < \mu_0(1)$, assuming (i) receivers are maximally open to persuasion, (ii) \tilde{h} given \tilde{d} and $\tilde{\mu}_0$ induces the same posterior belief as h given d and μ_0 : $\tilde{\mu}(\tilde{h}, \tilde{d}) = \mu(h, d)$, and (iii) \tilde{d} and $\tilde{\mu}_0$ fit \tilde{h} better than d and μ_0 fit h : $\Pr(\tilde{h}|\tilde{d}, \tilde{\mu}_0) \geq \Pr(h|d, \mu_0)$. To see this, the only non-trivial case is where $\mu(h, d)[1] = \tilde{\mu}(\tilde{h}, \tilde{d})[1] < k$. In this case, persuasion being ineffective at history h given default d and prior μ_0 means that $k > \mu_0(1)/\Pr(h|d, \mu_0)$ (applying condition (2) of Proposition 1). But this implies that $k > \tilde{\mu}_0(1)/\Pr(\tilde{h}|\tilde{d}, \tilde{\mu}_0)$ as well since $\tilde{\mu}_0(1)/\Pr(\tilde{h}|\tilde{d}, \tilde{\mu}_0) < \mu_0(1)/\Pr(h|d, \mu_0)$. So the conclusion follows from Proposition 1.

¹⁸In the limiting case that the world is deterministic under the receiver's default, i.e., (μ_0, π_d) places probability 1 on a single history so the set of possible histories H is a singleton, then persuasion is completely ineffective.

provided the model space M is sufficiently rich.¹⁹ For instance, receivers are more persuadable that a check-engine light signals the need for an expensive car repair than that an out-of-gas light signals the need for such a repair: the check-engine light is open for interpretation, but the gas light is not. While the gas light might be constraining, the amount of data that is open to interpretation also matters: Receivers are more persuadable that a car with every warning light on needs serious repairs than a car with only the check-engine or gas light on.

Formally, define sequences $(h_1^i)_{i=1,\dots,\infty}$ and $(h_2^i)_{i=1,\dots,\infty}$ with the properties that, for each $l \in \{1, 2\}$, (i) $\pi_{m^T}(h_l^i|\omega) \rightarrow 0$ and $\pi_{d(h_l^i)}(h_l^i|\omega) \rightarrow 0$ as $i \rightarrow \infty$ and (ii) there is a $\omega^T \in \Omega$ such that $\pi_{m^T}(h_l^i|\omega)/\pi_{m^T}(h_l^i|\omega^T) \rightarrow 0$ as $i \rightarrow \infty$ for all $\omega \neq \omega^T$. We interpret higher i as corresponding to more data: the first property says that, asymptotically, the likelihood of a history goes down as the length of the history increases; the second property says that, asymptotically, the true state is identified as the length of the history increases. Both of these properties hold for almost all sequences generated by π_{m^T} under standard further assumptions (e.g., we increase data by adding independent draws from a common underlying distribution). When the receiver has amount i of unambiguous data and amount j of ambiguous data, then his default model is represented as $\pi_{m^T}(h_1^i|\omega) \cdot \pi_d(h_2^j|\omega)$.

Proposition 3. *Suppose as described above there is unambiguous data, h_1^i , and ambiguous data, h_2^j : The receiver interprets h_1^i through the lens of the true model and is maximally open to interpretation regarding h_2^j .*

1. *Fixing ambiguous data h_2^j and any target belief $\tilde{\mu}$ with $\tilde{\mu}(\omega^T) < 1$, then there exists a \bar{i} such that the persuader is unable to propose a model that induces $\tilde{\mu}$ for any h_1^i with $i \geq \bar{i}$.*
2. *Fixing unambiguous data h_1^i and any target belief $\tilde{\mu}$, then there exists a \bar{j} such that the persuader is able to propose a model that induces $\tilde{\mu}$ for any h_2^j with $j \geq \bar{j}$.*

This proposition provides a sense in which more unambiguous data constrains the persuader, while more ambiguous data liberates the persuader. The first part says that, fixing the amount of ambiguous data, with enough unambiguous data the persuader is *unable* to get the audience to believe anything but the truth. The second part says that, fixing the amount of unambiguous data, with enough ambiguous data the persuader is able to get the audience to believe anything.

¹⁹The model proposed by the persuader must satisfy $\Pr(h|m) > \Pr(h|d)$, which in this context with ambiguous and unambiguous data implies that

$$\sum_{\omega} \pi_{m^T}(h_1|\omega) \mu_0(\omega) [\pi_m(h_2|\omega) - \pi_d(h_2|\omega)] > 0.$$

The first two terms in the summation are proportional to posterior from updating the prior μ_0 using unambiguous data h_1 under the true model. Thus, a receiver only finds an interpretation of ambiguous data compelling if it fits the data better than his default model after correctly accounting for unambiguous data. To the extent that unambiguous data pins down the state, it eliminates the space for effective persuasion. However, a greater quantity of ambiguous data increases the space for effective persuasion.

4 When the Wrong Story Wins

In this section, we assume the receiver’s default is the truth and ask: when does the wrong story win? We show that having to propose a model that is more compelling than the truth in the data often does not meaningfully constrain persuaders. In other words, the wrong story often wins.

When the true model is the default, the constraint that the persuader’s model be more compelling than the default (Eq. (1)) becomes

$$\Pr(h|m) > \Pr(h|m^T), \quad (3)$$

where m is the model proposed by the persuader. The persuader takes the true-model default into account in choosing which model to propose. Denote the model he proposes as $m^{tc}(h)$, where the “ tc ” superscript above is short-hand for “truthteller constrained”. For simplicity we are assuming that the true model is known by the persuader. In all of our examples where we assume the persuader’s payoff is independent of the state ω , the substantive part of this assumption is that the persuader knows the receiver’s default interpretation.²⁰

When the true model is that the default model, a trivial corollary of Proposition 1 characterizes the beliefs the persuader is able to induce.

Corollary 1. *Assume receivers use the true model as the default and receivers are maximally open to persuasion. Then the persuader is able to induce any beliefs $\mu(h, m) \in \Delta(\Omega)$ satisfying*

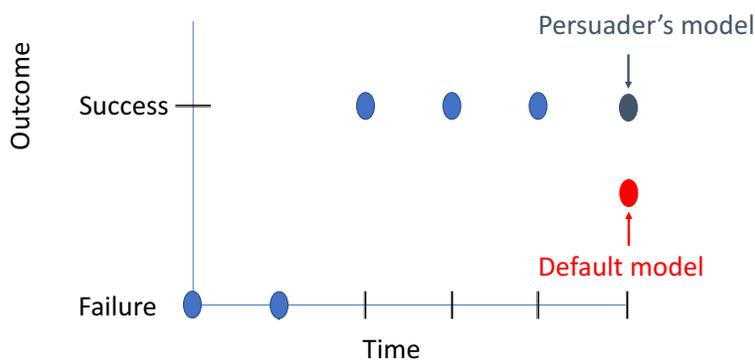
$$\mu(h, m)[\omega] < \frac{\mu_0(\omega)}{\Pr(h|m^T)} \quad \forall \omega \in \Omega \quad (4)$$

and is not able to induce beliefs that do not satisfy this inequality. As a result, persuasion is effective if and only if there exists an h in the support of (μ_0, π) such that some μ satisfying (4) yields a higher payoff to persuader than does μ_h .

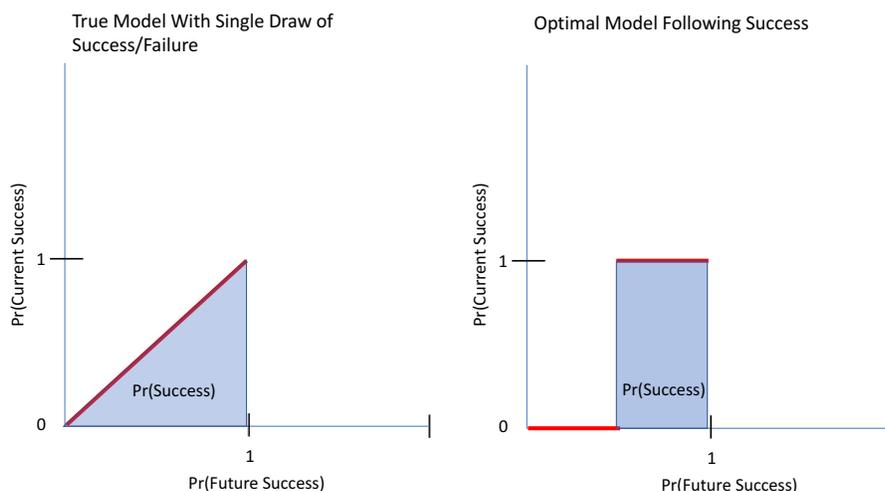
Corollary 1 makes it easy to compute the receiver’s beliefs under the optimal model when the receiver is maximally open to persuasion. As an illustration, return to the entrepreneur example from the introduction. We showed in the introduction that the persuader is able to get the investor to predict that the entrepreneur’s next startup will be successful with probability 80% if the receiver is only willing to entertain models of the form “this time is different”. Corollary 1 implies, and Figure 3a illustrates, that when the receiver is maximally open to persuasion, the persuader is able to get the investor to predict a much larger future success probability, 99%. Since the true model

²⁰In practice, the true data generating process is often not perfectly understood, even by experts. Taken literally, our assumption that the true model is known corresponds to the idea that experts are able to draw on a larger body of theory than non-experts. For example, in the case of climate change, experts might draw on thermodynamic theory, while non-experts might only draw on weather patterns.

does not fit the data all that well, the persuader is able to get the receiver's beliefs to move a lot in response to the data.



(a) Predicting success



(b) Optimal model when maximally open

Figure 3: Predicting the success of an entrepreneur's next startup when receivers are maximally open to persuasion

What do models look like when the receiver is maximally open? Continue with the entrepreneur example, but simplify the history so that the entrepreneur only has a single previous startup, which was successful. The left panel of Figure 3b shows the true model relating the probability of success of the entrepreneur's first startup and the probability of success of their second. Since a common success probability θ governs the success of each startup, the curve relating $\text{Pr}(\text{Current Success})$ to $\text{Pr}(\text{Future Success})$ is just the 45 degree line. Under this model, the investor estimates that the entrepreneur's next startup will be successful with probability $2/3$.

The right panel of Figure 3b shows the persuader's optimal model relating the probability of success of the entrepreneur's first startup and the probability of success of their second. Since the

persuader wants the investor to believe the entrepreneur is likely to be successful going forward, he has an incentive to propose a model where an initial success is inevitable when the likelihood of future success is greater than cutoff $\tilde{\theta}$, and initial success is impossible when the likelihood of future success is less than $\tilde{\theta}$. That is, the persuader proposes a model that “good entrepreneurs always reveal themselves by being successful early”. Under such a model, the investor estimates a probability of future success following an initial success of $(1 + \tilde{\theta})/2$. The persuader clearly wants cutoff $\tilde{\theta}$ as large as possible, but its magnitude is limited by the truth-teller constraint: the largest $\tilde{\theta}$ such that an initial success is as likely under the persuader’s as under the true model is $\tilde{\theta} = 1/2$. That is, the area under the rectangle in the right panel of the figure has to be at least as large as the area under the triangle in the left panel. Consequently, the persuader gets the investor to estimate that the entrepreneur’s next startup will be successful with probability $3/4$. The constraint that the persuader fits the data as well as the true model limits how much the persuader is able to move the receiver’s beliefs in response to the data. If the persuader tried to propose a higher cutoff $\tilde{\theta}$, the model would suggest that it is rare to see early successes because early successes only go with very rare, very talented entrepreneurs.

When the default model is the true model, the comparative statics on when receivers are persuadable we derived in Section 3 become statements about the true process. For instance, the more unlikely the history under the true model, the more space there is for misleading persuasion: Events that are truly surprising are ripe to be framed. By nature of being surprising under the true model, such events are not explained well by that model. This means the truth-teller constraint is relatively weak for such events. This result suggests that we should see a lot of persuasive activity to create narratives surrounding “tail events”, such as particularly good or bad performance of a company or worker.

In addition, applying the comparative statics from Section 3 when the true model is the default suggests that model persuasion is likely to be most effective in settings with a lot of randomness under the true model. The key advantage a persuader has relative to the truth-teller when there is randomness is that the persuader is able to tailor the model to the data. Knowing what the data say, the persuader can pick a model that is more compelling than the truth and makes the interpretation of the data favorable to the persuader.

4.1 An Illustration: Highlighting Strips of Data

The general results in the previous section suggested that a truth teller does not constrain the persuader much, particularly when the data are random. We next analyze the highlighting strips example from Section 2.2 using simulations to better understand the intuitions and magnitudes involved when receivers are open only to a limited set of models.

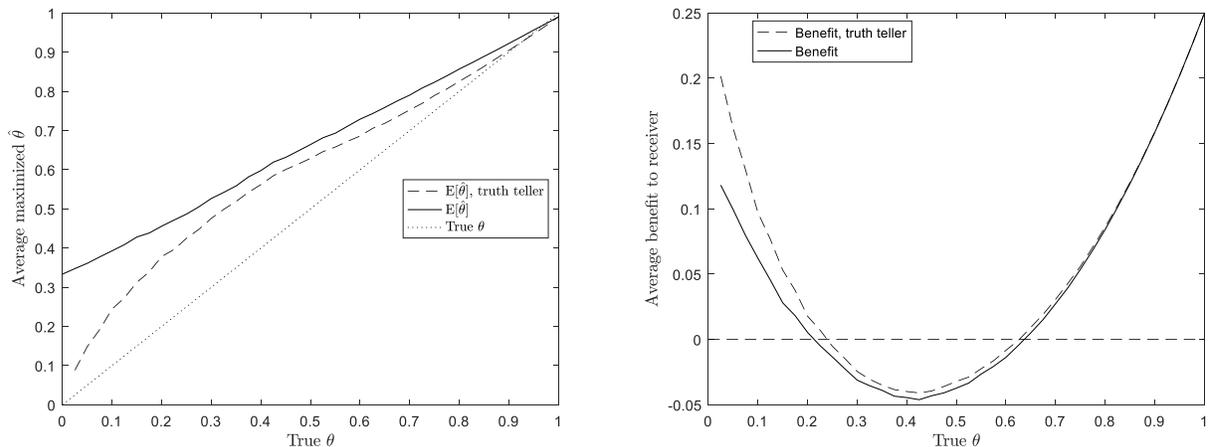


Figure 4: Simulated Impact of Persuasion on Beliefs and Welfare When The Receiver’s Default Model is the Truth

This figure presents results on the impact of persuasion from simulations of the coin-flipping example for the case where $\psi \sim U[0, 1]$ and $\underline{K} = 1$. For each of 40 values of θ , we plot the average post-persuasion beliefs of the receiver over 5,000 sample paths, each of length 100, comparing results when the default model is the true model versus when the default model is uninformative. The left panel plots the average post-persuasion beliefs of the receiver. The right panel plots the average post-persuasion benefit to the receiver.

Recall that in the example the coin is flipped t times, where it yields heads with probability θ . While θ is drawn once and for all at the beginning of time from a density ψ , the persuader can propose models of the form “the last \underline{K} periods are relevant for whether the coin comes up heads.” We denote the receiver’s posterior expectation of the probability of heads as $\hat{\theta}$. In our simulations, we pick a value of the true θ , and draw $t = 100$ random coin flips where the probability of heads is θ . We then find the optimal model for the persuader to propose, subject to the constraints that $\underline{K} = 1$, so that the persuader cannot “say nothing”, and that the persuader’s model must be more compelling than the truth. Finally, we compute the receiver’s post-persuasion beliefs assuming $\psi \sim U[0, 1]$. We run 5,000 simulations and report statistics aggregating across those simulations.

The left panel of Figure 4 shows the receiver’s average post-persuasion beliefs as a function of the true probability of heads θ . It draws a curve depicting the situation where the receiver has an uninformative default, so that he believes anything the persuader says, as well as a curve depicting the situation where the receiver’s default is the true model. When the true model is the default, the receiver’s post-persuasion beliefs are lower. The truthteller constraint prevents the persuader from proposing models that focus on very short favorable sequences. This reduces the scope for persuasion, particularly for low values of the true probability of heads θ . However, the figure

shows that the scope for persuasion remains substantial, particularly for intermediate values of θ . Intuitively, there is always positive probability of a history with a long string of tails followed by a long string of heads, i.e., $(0 \dots, 0, 1, \dots 1)$. A politician can point to their recent “momentum” and thus limit voters’ attention to a window of recent polls. In expectation, this increases voters’ assessment of the politician’s likelihood of winning. Similarly, a mutual fund company will choose to advertise with frames such as “be bullish” that emphasize past performance only when that past performance boosts investors’ beliefs that future returns will be high (Mullainathan, Schwartzstein, and Shleifer 2008; Phillips, Pukthuanthong, and Rau, 2016; Koehler and Mercer, 2009).

When the data is closer to random, i.e., $\theta \approx 0.5$, the truth-teller is not very helpful, and the persuader retains significant flexibility. The right panel of the figure shows impact of persuasion on the receiver’s payoff, defined here as $-(\hat{\theta} - \theta)^2 + (1/2 - \theta)^2$.²¹ Adding a truth-teller benefits receivers when θ is low, but has little benefit for intermediate or high values of θ .²²

Two other patterns from the left panel of the figure are worth noting. First, persuaders are constrained in the beliefs they can induce: on average, the receiver’s estimate $\hat{\theta}$ is increasing in the true θ because it influences the expected number of heads in the history. A politician with a greater chance of winning will on average be more successful at increasing voters’ assessments of her likelihood of winning. Similarly, a mutual fund with past successes will on average be more successful at increasing investors’ assessment of future returns. Second, persuasion tends to attenuate the relationship between people’s beliefs and the truth. For example, by inflating all political candidates’ perceived chances of winning, persuasion reduces the average reaction of perceptions to reality.

We next study how the impact of persuasion varies with sample size. Figure 5 plots the average difference between the receiver’s post-persuasion beliefs and the econometrician’s beliefs, $\mathbb{E}_{\psi}[\hat{\theta}] - \mathbb{E}_{\psi}[\theta]$, as a function of the length of the sample. Strikingly, we see that additional data actually *benefits* the persuader at the expense of the receiver. The intuition is that more data gives the persuader flexibility to propose compelling models that highlight favorable sequences—that is, to propose models that are beneficial to the persuader and overfit the historical data.²³ For instance,

²¹Applying our general formula for the impact of persuasion with a true-model default would, for given h , yield $-(\hat{\theta} - \theta)^2 + (\mathbb{E}_{\psi}[\theta|h] - \theta)^2$. This would obviously be negative in expectation for large enough t since $\mathbb{E}_{\psi}[\theta|h]$ converges to θ .

²²For sufficiently long histories, model persuasion not only leads to bias, but also to more variable beliefs relative to when receivers use the true model to interpret data. This arises because persuasion focuses the receiver’s attention on finite data when infinite data is available. This is consistent with the view of Akerlof and Shiller (2015) in the context of finance, who argue “Asset prices are highly volatile... sales pitches of investor advisors, investment companies, and real agents, and narratives of riches from nowhere are largely responsible.” In short histories, however, persuasion can sometimes reduce variance of beliefs. For instance, if the persuader’s incentive is to inflate estimates of θ and the true θ is large, the persuader is pulling in the “right” direction, which can reduce volatility.

²³On the other hand, the impact of persuasion does not go up with the data if the number of models the receiver is willing to consider decreases or stays the same as the amount of data increases. For example, the effectiveness of persuasion weakly decreases if the persuader can only choose between models that throw out the first 1, 2, or 3 flips

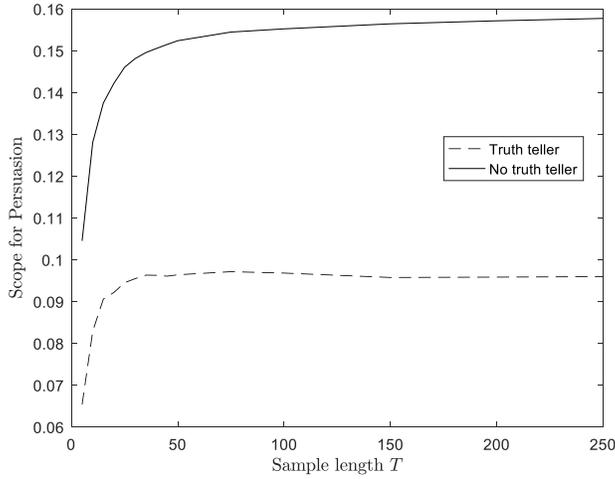


Figure 5: Simulated Impact of the Truthteller Constraint on the Scope for Persuasion

This figure presents results on the impact of persuasion from simulations of the coin-flipping example for the case where $\psi \sim U[0, 1]$ and $\underline{K} = 1$. For each sample length, we compute the average over θ of the difference between the receiver’s average post-persuasion beliefs and the econometrician’s beliefs.

for political candidates, a longer history in the public eye is both a blessing and a curse. The candidate has a larger set of positives to highlight, but their opponent also has a larger set of potential negatives to highlight. This finding contrasts with an intuition from “information persuasion” that more publicly available data if anything limits the scope for persuasion.

A final straightforward property of the framework is worth highlighting in this setting: The model the persuader proposes reveals the bias in the receiver’s beliefs. That is, $\hat{\theta} > \mathbb{E}_{\psi}[\theta|h]$ if and only if the persuader proposes a $k < t$ model. This result follows from a simple revealed preference argument: If the persuader proposes a $k < t$ model, it must be the case that $\mathbb{E}_{\psi}[\theta|h, k - \text{model}] > \mathbb{E}_{\psi}[\theta|h]$, since otherwise the persuader would propose the t -model. When a campaign urges people to neglect a bunch of polls saying they are biased, this probably signals that their candidate is in trouble. Similarly, the venture capitalist Marc Andreessen describes the way valuations are formed as follows: “What actually happens: (1) Observe current market valuation; (2) Construct theory and model to explain that valuation. . . At cyclical bottom, low prices drive creation of theories to explain permanent future misery; positive investors and analysts get fired. Therefore a boom in theories of how everything’s a bubble and certain to crash is evidence of a cyclical bottom, not a cyclical top.” In other words, when incentives are strong to propose pessimistic models, there will be many pessimistic models, which in fact indicate a more positive future.

($\underline{K} = t - 3$). In this case, the impact of persuasion will go away for large t since the impact of the first three flips will become negligible. Such restrictions seem less plausible than the setting we consider.

5 Competition Between Persuaders

This section considers competition between persuaders who are not restricted to tell the truth. To incorporate competition between persuaders, we suppose that a receiver who entertains multiple models adopts the one with the highest associated likelihood given the history. So, for example, if the receiver is exposed to one persuader who proposes m and another who proposes m' , then the receiver adopts posterior $\mu(h, m)$ if

$$\Pr(h|m) > \Pr(h|m') \tag{5}$$

and adopts posterior $\mu(h, m')$ if the inequality is reversed (assuming both models fit better than the default). The extension to more than two persuaders is obvious—the receiver goes with the proposed model, including the default, that maximizes $\Pr(h|\cdot)$. To break ties we assume that, in the case of equality, the receiver goes with the true model if (i) it is the default or is proposed by one of the persuaders and (ii) it provides the best fit among models the receiver is exposed to. Otherwise, we assume that the equilibrium determines the receiver’s tie-breaking procedure.

We assume that each message is determined by Nash Equilibrium and begin with a basic result.

Proposition 4. *Fix history h and suppose there are at least two persuaders. If \tilde{m} solves*

$$\max_{m \in M \cup \{d(h)\}} \Pr(h|m)$$

and m^T does not, then there is an equilibrium where the receiver holds beliefs $\mu(h, \tilde{m})$.

If a model maximizes the probability of seeing a history, then there is an equilibrium where receivers interpret data through the lens of that model. While it is often not the *only* equilibrium model, the proposition indicates that competition can push persuaders to propose models that best fit the data, even though such a model is rarely the one a single persuader would want to propose.²⁴ The intuition is that no persuader has an incentive to unilaterally deviate from proposing a model that best fits the historical data if another persuader is proposing it: the receiver will not find any other model more compelling.

²⁴Following such a model may lead receivers astray. To illustrate, suppose a person assesses which of several investments to make and has access to historical data on the success of investments with various characteristics. If the person will only entertain models of the form “characteristic A forecasts success” or “characteristic A does not forecast success”, then she will find it compelling that A forecasts success when it indeed does so in the historical data. If, however, characteristic A forecasts success only because it is correlated with a different characteristic that truly matters for success, then it may be a mistake to make investment decisions based on this characteristic. Similarly, a bad causal account that matches the empirical distribution will be more compelling than a correct causal account that does not quite match the empirical distribution. As we saw above with a truth-teller, a model may beat another even when it leads to worse decisions.

To place more structure on the full set of equilibrium beliefs and comparative statics, we now turn to the situation where receivers are maximally open to persuasion. In this case, the persuader's success will be a function of its incentives, the data, and the receivers' prior beliefs. We write the payoff to persuader j as V^j and use $V^j(\mu, h)$ as shorthand for $V^j(m(\mu), h)$, where $m(\mu)$ is a model that induces belief μ .

Proposition 5. *Suppose the receiver is maximally open to persuasion and there are multiple persuaders. μ is an equilibrium belief given history h if and only if (i) $\text{Fit}(\mu; h, \mu_0) > \Pr(h|d)$ and (ii) for all persuaders $j = 1, 2, \dots, J$*

$$V^j(\mu', h) > V^j(\mu, h) \Rightarrow \text{Movement}(\mu'; \mu_0) > \text{Movement}(\mu; \mu_0), \quad (6)$$

recalling that $\text{Movement}(\mu; \mu_0) = \max_{\omega \in \Omega} \frac{\mu(\omega)}{\mu_0(\omega)}$ is a measure of the movement from μ_0 to μ .

This result is a simple application of Lemma 1 and provides a necessary and sufficient condition for checking whether a belief is an equilibrium belief. One implication of this result is that a persuader is at an advantage when she wants to persuade the audience to reach a conclusion it is predisposed to believe.

Another implication of this result speaks to whether competition leads to more accurate beliefs. While this is often true with information or Bayesian persuasion (e.g., Milgrom and Roberts 1986; the conscientious reader example of Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2008; Gentzkow and Kamenica 2017), it is not true with model persuasion. The reason is that, as we saw in the previous section, receivers often do not find the true model the most compelling.

Corollary 2. *Suppose the receiver is maximally open to persuasion.*

1. *If there is a single persuader, the prior belief μ_0 may not be a solution to the persuader's problem given history h . However, when there are at least two persuaders, then μ_0 is an equilibrium belief given h .*
2. *Moreover, if prior belief μ_0 is the only equilibrium belief given history h , then it is the only equilibrium belief given h when more persuaders are added to the existing set of persuaders.*
3. *However, if true belief μ_h is an equilibrium belief given history h , then it may not be an equilibrium belief given h when more persuaders are added to the existing set of persuaders.*

This result implies that competition between model persuaders does not robustly lead receivers to more accurately interpret the data. Rather, as we will see illustrated in Section 7, it is a force towards leading receivers to adopt models that overfit the past, thus rendering the past as uninformative about the state. The idea is that a model which says that the past was inevitable in hindsight

will win out over other models—so this will be the equilibrium model if some persuader benefits from receivers adopting it.²⁵ And such a model promotes underreaction to data since it frames the data as completely unsurprising. Intuitively, competition promotes such narratives that explain everything in hindsight and consequently predict little.

As an illustration, consider the general phenomenon of associating market movements with narratives, which is a noted tendency among popular observers of financial markets:²⁶

In [the “Abreast of the Market”] column, you can also read selected post-mortems from brokerage houses, stock analysts and other professional track watchers explaining why the market yesterday did whatever it did, sometimes with predictive nuggets about what it will do today or tomorrow. This is where the fascination lies. For no matter what the market did—up, down or sideways—somebody will have a ready explanation.

As another illustration, return to the entrepreneurship example from the introduction, modifying it to consider two persuaders: one who wants the investor to invest in the entrepreneur’s next startup and another who wants the investor to not invest. That is, one persuader’s payoff is strictly increasing in the receiver’s posterior probability the startup succeeds, $\hat{\theta}$, and the other is strictly decreasing in $\hat{\theta}$. Then Proposition 5 implies that, in equilibrium, the investor will not react to the entrepreneur’s historical successes and failures at all: the investor predicts the future probability of success to be the prior probability of 50%. This situation is depicted in Figure 6. Competition between persuaders with opposing interests pushes receivers to adopt models that view the data as uninformative. That is, competition *neutralizes the data*.

This result may shed light on why some beliefs in the real world (e.g., on climate change) seem so stubborn in the face of facts, despite the presence of plenty of persuaders who have an interest in moving beliefs. This evidence seems particularly puzzling in light of recent work showing that short conversations are surprisingly effective at changing minds about issues like transgender rights (Broockman and Kalla 2016) and political candidates (Pons 2018). Our results suggest that when all persuaders have identical incentives they will indeed have a lot of success getting receivers to adopt models that lead them to overreact to data. However, when they have very different incentives (as in many competitive situations), they will end up persuading receivers to adopt models that lead them to underreact to data.

A third implication of Proposition 5 sheds light on the interaction between a persuader and a *strategic* truth-teller who wants the receiver to hold correct beliefs but is not constrained to propose the true model. Specifically, assume the strategic truth-teller’s payoff equals $v > 0$ if the receiver

²⁵However, in many cases μ_0 is not the only or the most natural equilibrium belief, for example in situations where all persuaders want receivers to hold optimistic beliefs.

²⁶Vermont Royster (Wall Street Journal, “Thinking Things Over Aaft of the Market,” January 15, 1986).

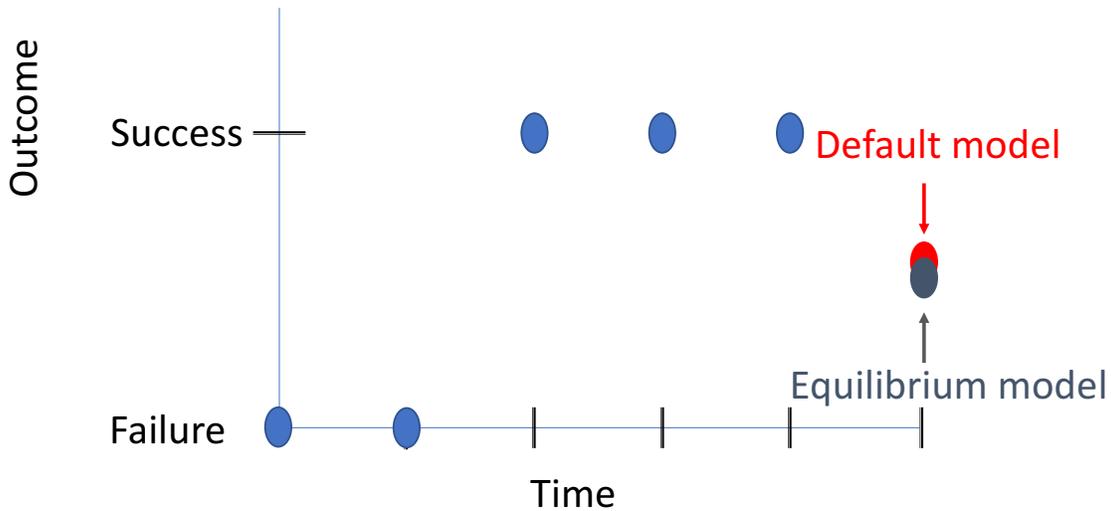


Figure 6: Competition between persuaders. One wants an investor to believe the entrepreneur’s next startup will be successful and the other that it will be unsuccessful.

holds beliefs μ_h and equals 0 otherwise. Does the strategic truthteller constrain equilibrium beliefs by more than the non-strategic truthteller? That is, might the strategic truthteller be more effective at inducing true beliefs by proposing untrue models? The answer is yes whenever the true model cannot perfectly explain the data.

Corollary 3. *Consider competition between a persuader and a strategic truthteller.*

1. *Suppose the true model does not perfectly explain the data: $\pi(h|\omega) < 1 \forall \omega \in \Omega$. If $\mu \neq \mu_h$ is an equilibrium belief then it also satisfies the (non-strategic) truthteller constraint. However, there is a belief μ that satisfies the truthteller constraint but is not an equilibrium belief.*
2. *Suppose $\max_{\omega \in \Omega} \pi(h|\omega) = 1$. In this case μ is an equilibrium belief if and only if it satisfies the truthteller constraint.*

This result says that, with competition, the most persuasive way to get someone to hold accurate beliefs μ_h is not necessarily to push the true model: the true model may create too much space for another persuader to propose a model that better fits the past. As Lakoff (2004) writes: “the truth alone will not set you free ... You need to frame the truths effectively from your perspective.” For a simple illustration, suppose the truth is that the data is uninformative because it is perfectly random given the true state. In this case, the persuader who wants to convince the audience that the data is uninformative is better off telling a story where the data is uninformative because the results were inevitable no matter the true state.

As another illustration, return to the entrepreneur example assuming receivers are maximally open to persuasion. One can show that there is a model that leads to the true-model posterior that fits the historical data 28 times better than the true model. A persuader who wants to induce optimistic beliefs is much more constrained competing with a strategic truth-teller who promotes this model than with a non-strategic truth-teller promoting the true model. Indeed, the best the persuader can do against this model is induce a belief that there's a 76% chance the entrepreneur will be successful in her next project—well below the 99% forecast the persuader is able to induce if competing with a non-strategic truth-teller.

This result suggests that persuaders are at a significant rhetorical disadvantage if they are committed to telling accurate stories to induce accurate beliefs. A climate scientist, for example, may be at a disadvantage in persuading the audience if she feels compelled to point out that some high-frequency temperature variation is likely just noise. Likewise, an advocate that A does not cause B (e.g., vaccines do not cause autism) is at a disadvantage if they are unwilling to propose alternative stories for what does cause B. And, as we discuss below, an unbiased investment advisor might be at a disadvantage if she is committed to providing academic arguments that passive investing is attractive because stock returns are hard to forecast. Instead, she may be better off supplying the more persuasive (though less scientifically accurate) stories explaining that the only traders who beat the market have insider information.

6 Multiple Receivers

This section considers what happens when there are multiple receivers, relaxing the assumption that everyone shares the same prior and/or default interpretation. For example, an entrepreneur might need to give the same pitch to multiple potential investors. Or an investment advisor might detail her philosophy on active investing in a newsletter that multiple people read. When does this constrain the persuader relative to if he could tailor his message to the receiver?

We begin with a simple illustration, building on the example in Section 3. As before, there is a broker who is incentivized to get investors to invest in an actively managed mutual fund, which is either good or bad. We now suppose that there are two investors, Pat and Oscar. As before, Pat is relatively pessimistic—his prior belief that the active fund is good is 20%. Oscar is more optimistic—his prior belief that the active fund is good is 40%. Pat's default is the same as before: he believes past returns are somewhat informative:

$$\pi_d^{\text{Pat}}(\text{high returns}|\text{good}) = \pi_d^{\text{Pat}}(\text{low returns}|\text{bad}) = 75\%.$$

Oscar has an uninformative default and believes that

$$\pi_d^{\text{Oscar}}(\text{high returns}|\text{good}) = \pi_d^{\text{Oscar}}(\text{high returns}|\text{bad}) = 64\%.$$

Each will only invest if, after persuasion, the probability they put on the active fund being good is above 50%. Finally, assume as before the active fund has high past returns.

By Proposition 1, if the broker can propose a different model to Pat and Oscar, she can get both to invest. This is not the case if the broker must propose the same model to both. To see why, note that the movement-maximizing model that gets Oscar to invest sets $\pi_m(\text{high returns}|\text{good}) = 1$ and $\pi_m(\text{high returns}|\text{bad}) = (64\% - 1 \times 40\%)/64\% = 37.5\%$. For Pat, this model implies the probability of observing high returns is

$$\Pr^{\text{Pat}}(\text{high returns}|m) = \pi_m(\text{high returns}|\text{good}) \times 20\% + \pi_m(\text{high returns}|\text{bad}) \times 80\% = 50\%. \quad (7)$$

So Pat finds this model more compelling than his default (as we saw in Section 3 the fit of Pat's default is 35%). But when Pat updates with this model he has

$$\Pr^{\text{Pat}}(\text{good}|m, \text{high returns}) = \pi_m(\text{high returns}|\text{good}) \frac{\mu_0^{\text{Pat}}(\text{good})}{\Pr^{\text{Pat}}(\text{high returns})} = 40\%. \quad (8)$$

Intuitively, Oscar's default fits so well that the model that Oscar finds compelling does not induce much movement. Even if Pat finds this model compelling, it does not induce enough movement to get him to invest, since he was relatively skeptical to start with. Any model that induces more movement for Pat will fit worse for Oscar, and then fit worse than Oscar's default model. Consequently, any model that gets Pat to invest will not be compelling to Oscar, and therefore not induce Oscar to invest.

Even more starkly, the broker can be in the tough situation in which persuasion can backfire: the model that gets one person to invest causes the other to stop investing. To see this, modify the example so that Pat's prior is slightly more optimistic: $\mu_0^{\text{Pat}}(\text{good}) = 25\%$. With this modification, Pat will invest under his default model in the absence of persuasion. The fit of Pat's default is now $\Pr^{\text{Pat}}(\text{high returns}|d) = 37.5\%$. If the broker again proposes the movement-maximizing model that gets Oscar to invest, Pat will find that model compelling by (7) and will now not invest by (8).²⁷

These examples show two instances where the persuader is constrained by her inability to send separate messages to different audience members. To develop intuition for when this is a

²⁷Intuitively, to get Oscar to invest, the broker must propose a model that fits well, i.e., a model that implies that high returns are frequent. Any such model must involve a relatively high probability of high returns for bad funds, $\pi_m(\text{high returns}|\text{bad})$. Pat finds such models compelling because his prior belief is that bad funds are common, so such models suggest that high returns are frequent and unsurprising. The combination of his prior and such a model implies that high returns are not informative enough about the quality of the fund to get him to invest.

problem for the persuader, we generalize these examples slightly. Retain a binary state space with $\Omega = \{b, g\}$ (e.g., “bad”, “good”) with binary actions $a \in \{0, 1\}$ (e.g., “not invest”, “invest”), and a binary history $h \in \{\underline{h}, \bar{h}\}$ (e.g., “low return”, “high return”). Suppose there are two receivers, “pessimist” and “optimist”, both of whom care about the true state ω :

$$U^{\text{optimist}}(a, \omega) = U^{\text{pessimist}}(a, \omega) = \begin{cases} 1 & \text{if } a = 0 \text{ when } \omega = b \text{ or } a = 1 \text{ when } \omega = g \\ 0 & \text{otherwise.} \end{cases}$$

The persuader is trying to make both receivers choose $a = 1$ (e.g., invest): $U^S(a, \omega) = a$. The receivers can have different priors, with the optimist being weakly more optimistic that $\omega = g$: $\mu_0^{\text{optimist}}(g) \geq \mu_0^{\text{pessimist}}(g)$. They can also have different default models, labeled π_d^{optimist} and $\pi_d^{\text{pessimist}}$.

Proposition 6. *Suppose, under their priors, neither the optimist nor the pessimist would choose $a = 1$, but both are persuadable to choose $a = 1$ at history h : $1/2 > \mu_0^j(g) > \Pr^j(h|d)/2$ for $j = \text{optimist, pessimist}$. The persuader is able to send a menu of public messages at history h that gets both receivers to take the action $a = 1$ if and only if (i) the optimist takes action $a = 1$ under their default interpretation or (ii) there is a message that is both compelling to the optimist and gets the pessimist to take action $a = 1$:*

$$\frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{optimist}}(g)}{1 - \mu_0^{\text{optimist}}(g)} < \frac{\mu_0^{\text{pessimist}}(g)}{1 - \mu_0^{\text{pessimist}}(g)}. \quad (9)$$

Corollary 4. *Under the assumptions of Proposition 6, the persuader is able to send a message that gets both receivers to take action $a = 1$ if they share the same prior, $\mu_0^{\text{optimist}} = \mu_0^{\text{pessimist}}$, or the same default interpretation, $\pi_d^{\text{optimist}} = \pi_d^{\text{pessimist}}$.*

The proposition and corollary imply that when there are multiple receivers, persuasion is more effective when the receivers share similar priors and default interpretations. The proposition strengthens the above examples by characterizing instances where the persuader is unable to send even a *menu* of public messages that simultaneously persuades individually persuadable receivers who have sufficiently different priors and default interpretations. In such instances, the persuader would benefit from being able to send private, individually-tailored, messages to the receivers. As emphasized in communications textbooks such as Severin and Tankard (2001) there are benefits to sending targeted messages, e.g., through face-to-face conversations, when the audience is diverse.

7 Examples

In this section, we give three brief examples of applications of our model.

7.1 Technical Analysis

Technical analysis in financial markets illustrates many of the key intuitions that arise from model persuasion. Technical analysis aims to identify trading opportunities by finding patterns in prices and trading volumes. Figure 7a shows a common type of technical analysis, identifying prices of “support” and “resistance” for a stock. Support is a price point, \$465 per share in the figure, at which there is posited to be high latent demand, which prevents prices from falling further. Resistance is a price point, \$580 in the figure, at which there is posited to be high latent supply, which prevents prices from rising further. These points are determined by examining the historical price path of the stock.

While Figure 7a is an illustrative example from the brokerage center Fidelity’s “Learning Center” for investors, Figure 7b shows a real world example of technical analysis. The analysis was done by a brokerage firm that sells investment advice and services to clients. Using data on Amazon’s stock price from December 2018 through the end of January 2019, the brokerage suggests going long (buying) Amazon stock because it was close to its support price on January 29, so it is likely to rise going forward.

This kind of analysis is extremely common in financial markets. All of the major brokerage services catering to individual investors, including Fidelity, E-Trade, Charles Schwab, Merrill Lynch, and TD Ameritrade, offer their client tools for technical analysis. And it is commonly featured on finance television programs like Jim Cramer’s Mad Money on CNBC. The practice is not restricted to amateurs—a variety of surveys find that over 30% of professional investors such as equity mutual fund managers and foreign exchange (FX) traders use technical analysis.²⁸ The ubiquity of technical analysis is puzzling given that it is arguably ineffective in actually producing trading profits (Lo, Mamaysky, and Wang 2000; Bajgrowicz and Scaillet 2012).

Why is technical analysis so common if it does not reliably generate profits? Basic lessons of our framework may shed light on this question: A key advantage of any model persuader is that they can tailor models to the data. Knowing what the data say, a persuader is able to pick model that is more compelling than the truth. Indeed, the support and resistance model looks so compelling in Figures 7a and 7b because the historical data are used to determine the support and resistance levels ex post. The advantage a persuader has in tailoring models to data is particularly large when there is a lot of randomness under the true model, as is the case with financial market prices.

Technical analysis is also a situation where there are many competing narratives. Figure 7c

²⁸For instance, in a sample of more than 10,000 portfolios, about one-third of actively managed equity funds use technical analysis (Smith, Faugere, and Wang 2013). About 60% of commodity trading advisors heavily or exclusively use technical analysis (Billingsley and Chance 1996). 90% of London-based FX traders put some weight on technical analysis (Taylor and Allen 1992), while 30% of US-based FX traders report that technical analysis is their “dominant strategy” (Cheung and Chinn 2001).



(a) Illustration of Support and Resistance. Source: Fidelity Investments



(b) Technical Analysis: Long Amazon. Source: TradingView.com



(c) Technical Analysis: Short Amazon. Source: TradingView.com

Figure 7: Technical Analysis

shows another real world example of technical analysis using the same data as Figure 7b on Amazon’s stock price from December 2018 through the end of January 2019. In this case, the analyst suggests selling Amazon’s stock short (i.e., betting on a decline) because its price has fallen below the “neckline” in a “head and shoulders” pattern. There is not direct evidence on which analysis receivers find more compelling. However, our framework suggests that, in equilibrium, a persuader

is at an advantage when he does not want to move the receiver’s beliefs far from her prior. So for instance, the analyst proposing Figure 7c will be at an advantage relative to the analyst proposing Figure 7b if the receiver is predisposed to be pessimistic about Amazon stock.

7.2 Persuading a Jury

Consider a prosecutor and defense attorney trying to convince a jury of the guilt or innocence of a defendant, along the lines of the example in Kamenica and Gentzkow (2011). Unlike that model, which concerns the ability of the defense and prosecution to selectively collect and reveal evidence to boost their respective cases, our interest is in the ability of the defense and prosecution to frame evidence to boost their case. For example, closing arguments are used not to introduce new evidence, but rather to push narratives for interpreting the evidence. This view of juror decision-making as being influenced by stories or narratives to explain the evidence—sometimes referred to as the “story model” for juror decision-making (Pennington and Hastie 1986; 1988; 1992)—has a long history in scholarship on psychology and law. A main point of our analysis is that, in equilibrium, evidence that is informative under the true model but open to interpretation does not influence juror decision-making: If anything, the model the juror finds most compelling frames such evidence to reinforce his prior beliefs.

Specifically, suppose there is a representative juror, a defense attorney, and a prosecutor who share prior μ_0 over the guilt ($\omega = g$) or innocence ($\omega = ng$) of a defendant. The juror gets payoff v if he convicts a guilty defendant or acquits an innocent defendant; payoff $-a$ if he convicts an innocent defendant; and payoff $-b$ if he acquits a guilty defendant. The juror will then optimally follow a cutoff rule where he convicts a defendant if and only if beliefs μ satisfy $\mu(g) \geq (v + a)/(2v + a + b) \equiv k(a, b)$. In the natural case where the juror is particularly wary of convicting an innocent defendant ($a > b > 0$), then $k(a, b) \in (1/2, 1)$. The prosecutor’s payoff is v if the defendant is convicted and 0 otherwise, while the defendant’s payoff is v if the defendant is acquitted and 0 otherwise. Kamenica and Gentzkow’s (2011) emphasis in a similar setup is that each side may tailor an investigative strategy (interviewing certain witnesses, etc.) to benefit from Bayesian persuasion because payoffs are naturally non-linear in prior beliefs. And when attorney’s compete, there is full revelation (Kamenica and Gentzkow 2012). What happens with model persuasion?

Suppose everyone sees the same evidence h (e.g., physical evidence, eyewitness testimonies, etc.) and the representative juror is maximally open to persuasion. Then Proposition 5 implies a sharp result: in equilibrium, the juror will make the same decision to acquit or convict as he would if he simply went with his prior. The impact of the competing narratives of the defense and prosecution is to divorce the juror’s decision from the evidence that is open to interpretation. To

see this, suppose that the juror's equilibrium beliefs μ supported convicting the defendant when his prior beliefs μ_0 supported acquitting her. Then condition (6) is violated because the defense would benefit from beliefs μ_0 , which are obviously closer to itself than to μ (i.e., $\text{Movement}(\mu_0; \mu_0) < \text{Movement}(\mu; \mu_0)$). This contradicts μ being the juror's equilibrium beliefs. While the impact of Bayesian persuasion is for the juror to make the correct decision in equilibrium, the impact of model persuasion is to move the juror's beliefs further in the direction of supporting the decision he would make in the absence of the evidence that doesn't have an obvious interpretation. Note that this result is independent of the specific data h : allowing the defense and prosecution to collect and reveal more data would not alter this conclusion, provided the evidence is maximally open to interpretation ex post.

The intuition is that models that frame the evidence to fit what the juror already believes to be true will resonate with him. The juror who thinks that the defendant is very unlikely to be guilty will find alternative explanations for that damning evidence compelling; the juror who thinks that the defendant is likely to be guilty will find it compelling that the same evidence is quite diagnostic of guilt. This basic force carries through to situations short-of-the-limit of jurors being maximally open to persuasion: model persuasion then mitigates but not eliminate the impact of evidence on juror decisions. To illustrate, we could imagine that evidence comes in two categories: facts that are not open to interpretation and softer or more circumstantial evidence. Our model then applies, taking the prior μ_0 as already incorporating the facts that are not open to interpretation.

The model suggests that the stories the defense and prosecution tell matter: jurors will be swayed separately by the defense's and prosecution's arguments to frame the evidence. However, the net effect (at least with skilled attorneys) will be for jurors to arrive at the same conclusion that they would in the absence of the arguments. This is broadly consistent with findings from the literature on competitive framing more generally (Busby, Flynn, and Druckman 2018), where equally strong competing frames "cancel out".

Relaxing the assumption of a common prior, the result may shed light on the importance of juror, judge, or arbitrator characteristics on outcomes (e.g., Anwar, Bayer, and Hjalmarsson 2012, 2014; Arnold, Dobbie, and Yang 2018; Egan, Matvos, and Seru 2018), despite how trials and arbitrations reveal evidence that Bayesians should agree on. It also suggests that parties will take juror and judicial characteristics into account when proposing narratives on how to interpret the facts.

7.3 Persuading an Investor

It is well known that household investors make mistakes in portfolio allocation decisions: they tend to be under-diversified, trade too much, and invest in dominated products like high-fee index

mutual funds (see Campbell 2006 for an overview). One often-stated reason is that investors follow biased recommendations of advisors, who have an incentive for the investors to invest in certain ways. For instance, brokers may earn high commissions for directing investors towards high fee mutual funds (Bergstresser, Chalmers, and Tufano 2008, Chalmers and Reuter 2012, Hackethal, Haliassos, and Jappelli 2012). But investors are likely exposed to financial advice from multiple competing sources, and some of these sources offer advice that, if followed, would lead to better decisions. This raises a question: Why do investors follow the biased advice? Our model offers a particular answer to the question: investors may find biased advice more compelling than the truth in the data.

We provided a toy model of investment advice in Section 3. For a somewhat more elaborate formulation, suppose there are N investments with investment j having characteristics (x_1^j, \dots, x_K^j) . Each investment will either be successful or not. If the investor's investment is successful he gets a payoff of $s > 0$ and he gets a payoff of 0 otherwise. The investor may pay a cost $\chi \in (0, 1)$ to make an "active" choice in a particular investment or to pay a cost $\chi_L \in [0, \chi)$ to invest "passively" in one of the N investments selected at random. We normalize $\chi_L = 0$, so the investor will want to make an active choice of a particular investment if and only if he thinks he is able to predict which investment will be successful to an extent that justifies the cost χ . The person's prior μ_0 is that the success probability of each investment is independently drawn from a uniform distribution over a finite set centered around $1/2$.

This prior leaves open the possibility that the successes of previous investments with similar characteristics help predict which current investment will be successful. The investor has access to a database of the previous successes and failures of investments with different characteristics. There are T entries to the database, each of the form $(y_{jt}, x^j)_{j=1}^N$, where $y_{jt} = 1$ if investment j was successful in period $t \in \{1, \dots, T\}$ and 0 otherwise. In reality, this database is not helpful to predicting which current investment will be successful: the success probability of each investment is independently drawn each period from the uniform distribution.

But some investment advisors have an incentive for the investor to believe in predictability. In particular, assume that one advisor ("Active") gets a payoff of $v > 0$ if the investor incurs the cost χ to make an active investment and 0 otherwise, while another advisor ("Passive") gets a payoff of v if the investor makes a passive investment and 0 otherwise.

Suppose first that receivers are maximally open to persuasion and Passive acts as a non-strategic truth-teller who always proposes the true model that success is not predictable. In this case, a simple application of Proposition 1 shows that, when T is sufficiently large, Active will always convince investors to make an active investment: Active is at an advantage because there is so much room for overfitting.

Continue to assume that receivers are maximally open to persuasion, but suppose that Passive

acts as a strategic truth teller. In this case, a simple application of Proposition 5 shows that Passive will always convince investors to make a passive investment because Passive is at an advantage: she wants investors' beliefs to stay at the prior. But this is only an advantage if Passive is willing to propose the wrong model. In addition, Passive's advantage would disappear if the receiver's prior favored Active instead.

In summary, our model suggests that biased advice is followed because it often fits better with what investors know than unbiased advice. Biased advice typically comes from persuaders who have an incentive to propose models that overfit historical data; the unbiased advice may come from persuaders (e.g., academics) who have more of an incentive to push correct models than to push compelling models. This may help understand why investors may choose not to follow unbiased investment advice that would improve their portfolio performance even if they obtain it (e.g., Bhattacharya et al. 2012).

8 Discussion

This paper presents a framework for analyzing model persuasion, where persuasion operates through providing receivers with models for interpreting data they already have ready access to. Such persuasion is particularly effective when receivers have access to long and wide strips of data and, in natural cases, when outcomes are close to random. Persuaders' statements are predictive of the ultimate bias in receivers' beliefs. The presence of truth tellers does not eliminate the impact of misleading persuasion because there are wrong models that better fit the past than the correct model. And competition promotes models that explain too much in hindsight and cause beliefs to underreact to evidence.

The framework is amenable to a number of applications and extensions. Under suitable assumptions, for example, it could be used to study self persuasion. Endogeneizing default models and, in particular, better understanding when receivers "see patterns" and overfit the data on their own could shed more light on when it's particularly difficult to persuade others.

Having a static model allowed us to punt on some questions, for example what happens if the persuader proposes a different model today than yesterday, or if the persuader needs to commit to a model in advance of some data being released. It also abstracts from phenomena such as "preemptive framing", where a party benefits from proposing the first narrative surrounding new data. Likewise, having a fixed history abstracted from potential interactions between Bayesian and model persuasion. For example, the results on how more information may benefit the persuader (even in the presence of a truth teller) suggest that model persuaders prefer to collect and reveal a lot of information *ex ante* that they can then frame in beneficial ways *ex post*. We provide an illustrative example in Appendix B.3. Or, if receivers overfit on their own, persuaders may want

to provide fresh information that they get a first crack at spinning. But focusing on the case where data is publicly available and exogenous to the persuader allowed us to focus on a central feature of persuasion: Often, its impact comes through framing or telling stories about data—making the truth work—instead of generating the data itself.

A Highlighting Characteristics

This section considers the highlighting characteristics example. We simplify the analysis by assuming throughout that the history $h = (y_k, x_k)_{k=0}^{t-1}$ is infinite ($t = \infty$) and the receiver holds an uninformative default model. Let $p(x)$ denote the proportion of times the receiver encounters x in infinite h . The following lemma is useful for building intuition.

Lemma 2. *Consider the highlighting characteristics example with an infinite history ($t = \infty$). Supposing $V(a)$ is increasing in a , then the persuader will propose the model $m \in M$ that maximizes*

$$\hat{\theta}(x) = \max_{m \in M} \theta(x) \cdot p(x|c_m(x)) + \sum_{x' \in c_m(x) \setminus \{x\}} \theta(x') \cdot p(x'|c_m(x)). \quad (10)$$

We first consider the circumstances under which persuasion is most effective in this setting. Equation (10) highlights that the persuader takes into account both success probabilities and frequencies of occurrence in proposing a model that lumps x with other elements in $c_m(x)$. Lumping x together with a highly probable x' is beneficial to the extent that it reduces the impact of a relatively low $\theta(x)$ on beliefs — we call this the *dilution effect*. Lumping x together with a relatively successful x' is beneficial to the extent that drawing this analogy boosts the perception of x — we call this the *magnitude effect*. Because of the dilution effect, the framing component of persuasion will be large when $p(x)$ is small. Because of the magnitude effect, it will be large when $\theta(x)$ varies a lot in x . To confirm these intuitions, parameterize $p(x)$ by letting $p_\delta(x) = p(x) \cdot (1 + \delta)$ and $\forall x' \neq x, p_\delta(x') = p(x') \cdot \left[1 - \delta \frac{p(x)}{1-p(x)}\right]$. Increasing δ increases the frequency of x relative to $X \setminus \{x\}$ while keeping the relative frequencies of $x', x'' \in X \setminus \{x\}$ the same.

Proposition 7. *Consider the highlighting characteristics example with an infinite history ($t = \infty$). Parameterizing $p(x) = p_\delta(x)$ and supposing $V(a) = a$:*

1. *The framing component of persuasion is higher when the frequency of x is lower: For every $(\theta(x'))_{x' \in X} \in [0, 1]^{|X|}$, $\hat{\theta}(x)$ is (weakly) decreasing in δ .*
2. *The framing component of persuasion is higher when characteristics more strongly predict success: For every $p \in \Delta(X)$ and $\theta(x)$, in expectation $\hat{\theta}(x)$ is lower when the other $\theta(x')$, $x' \neq x$, are drawn from ψ than when they are drawn from a mean-preserving spread of ψ .*

The broad intuition here is well captured by the behavior of stock market analysts in the 1990s technology bubble. Incentivized to produce positive analysis for firms that did not perform well on traditional financial metrics, analysts “became bolder about relying on nonfinancial metrics

such as 'eyeballs' and 'page views.'"²⁹ For instance, a July 1998 report on Yahoo noted "Forty million unique sets of eyeballs and growing in time should be worth nicely more than Yahoo's current market value of \$10 billion." Later the same year, the same analyst assessed Yahoo along five key financial metrics, listing growth in page views first, before revenues or operating margins. By choosing a different set of valuation metrics, stock market analysts were able to (temporarily) justify high valuations for technology stocks.

Rare events can provide particularly strong fodder for a given viewpoint. A Financial Times article entitled "Wall Street works up new rationale for optimism" discusses financial analysts who argued that stock markets would rally as the Fed ended quantitative easing and raised interest rates by citing "the 1950s lifting of the financial repression that had helped fund the war effort with low bond yields. In that environment, steadily rising interest rates saw the beginning of a historic bull market."³⁰ The rare characteristic of government intervention in the bond market was highlighted by analysts to justify their viewpoint that stocks would rally.

We next study how what the persuader says relates to the bias she induces.

Proposition 8. *Consider the highlighting characteristics example with an infinite history ($t = \infty$), and suppose $V(a)$ is increasing in a . The framing component of persuasion is positive if and only if the proposed model is tailored to the characteristics of the actor: $\hat{\theta}(x) > \theta(x)$ for some x if and only if there exist $x, x' \in X$ such that $m(x) \neq m(x')$.*

Inconsistency is a hallmark of successful persuasion. The bad want to pool with the good, and the good don't want to pool with the bad. For instance, Zhou (2017) shows that companies tend to blame poor performance on external factors (i.e., macro- or industry-wide factors) on conference calls with investors and tend to have lower stock returns in the future when they do so. Blaming macro- or industry-wide factors is a way to pool with other firms. In contrast, firms tend to attribute good performance to their own idiosyncratic factors.

For related simulations, consider the case where there are two binary characteristics $(X_1, X_2) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Suppose the persuader is interested in maximizing the receiver's perception of the success probability of the characteristic bundle $(0, 0)$, $\hat{\theta}_{0,0}$. Suppose further that the true success probabilities are independent across bundles, so the best estimate is the value $\theta_{0,0}$, which we assume is observed without noise, since we are assuming there is an infinite history. In each simulation, we fix the value of $\theta_{0,0}$ and draw the values of the success probabilities $\theta_{0,1}$, $\theta_{1,0}$, and $\theta_{1,1}$ randomly. We assume that the probability that each characteristic is 0 is $\Pr(X_1 = 0) = \Pr(X_2 = 0) = \sqrt{0.5}$ so that the frequency of $(0, 0)$ is 50%, the frequencies of $(0, 1)$ and $(1, 0)$ are each 21%, and the frequency of $(1, 1)$ is 9%. The persuader proposes the partition of $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ that includes $(0, 0)$ and maximizes Equation (10).

²⁹http://archive.fortune.com/magazines/fortune/fortune_archive/2001/05/14/302981/index.htm

³⁰<https://www.ft.com/content/24f206e8-33ae-11e4-ba62-00144feabdc0>

Figure A1 panel (a) reports the receiver’s post-persuasion beliefs as a function of $\theta_{0,0}$. Beliefs are averaged over 5,000 simulations, each of which draws $\theta_{0,1}$, $\theta_{1,0}$, and $\theta_{1,1}$ independently from a uniform distribution on $[0, 1]$. The figure shows that persuasion is again effective: the persuader succeeds in pushing the receiver’s post-persuasion belief $\hat{\theta}_{0,0}$ above the true value $\theta_{0,0}$. Panel (b) of the figure shows the average post-persuasion benefit to the receiver, relative to their utility from using prior belief $1/2$. For high values of $\theta_{0,0}$, where the persuader’s incentives are aligned with the truth, the average benefit is positive. For low values of $\theta_{0,0}$, where the persuader’s incentives are not aligned with the truth, the average benefit is smaller. Relative to the coin flipping example, the information component of persuasion is weaker here. Even though we require that the proposed partition includes $(0, 0)$ and we have assumed $(0, 0)$ has a high frequency, for low values of $\theta_{0,0}$ the information component of persuasion is weaker than the framing component.

Figure A2 demonstrates the first part of Proposition 7. We compare the base case from Figure 5, where the frequency of $(0, 0)$ was 50%, to the case where the frequency of $(0, 0)$ is 75%.³¹ Panel (a) of Figure A2 shows that the receiver’s post-persuasion average belief is higher and further from the true value $\theta_{0,0}$ in the base case when the frequency of $(0, 0)$ is lower. Panel (b) of Figure A2 shows that the benefit to the receiver is lower when the frequency of $(0, 0)$ is lower.

Figure A3 demonstrates the second part of Proposition 7. We compare the base case from Figure A1, where $\theta_{0,1}$, $\theta_{1,0}$, and $\theta_{1,1}$ are drawn independently from a uniform distribution on $[0, 1]$, to the case where each variable is independently drawn from the average of 4 uniform distributions on $[0, 1]$. Thus, each variable is still 0.5 in expectation; however, its variance is now lower, .02 rather than .08. Panel (a) of Figure A3 shows that the receiver’s post-persuasion average belief is higher and further from the true value $\theta_{0,0}$ in the base case where the variance of $\theta(x)$ is higher. Panel (b) of Figure A3 shows the corresponding comparison of benefits to the receiver. They are lower in the base case, where the variance of $\theta(x)$ is higher.

B Further Extensions

B.1 Value-Adjusted Fit

In the main text, we assume the receiver finds one model more compelling than another when it fits the data plus prior better, a notion of fit that is independent of the receiver’s incentives. An alternative, value-adjusted notion of fit, takes into account the impact of adopting the model on the receiver’s decisions.

³¹Specifically, we assume that the probability each characteristic is 0 is $\Pr(X_1 = 0) = \Pr(X_2 = 0) = \sqrt{0.75}$ so that the frequency of $(0, 0)$ is 75%, the frequencies of $(0, 1)$ and $(1, 0)$ are each 12%, and the frequency of $(1, 1)$ is 2%.

Recall that if the receiver adopts model \tilde{m} , then he chooses action:

$$a(h, \tilde{m}) \in \arg \max_{a \in A} \mathbb{E}_{\mu(h, \tilde{m})}[U^R(a, \omega)].$$

The expected payoff of taking this action after being exposed to multiple models depends on the receiver's posterior beliefs over those models. Suppose the receiver's posterior is generated as if the receiver has a flat prior the models \tilde{M} he's exposed to. That is, the receiver's posterior of model m given prior $\mu_0(\omega)$, history h , and set of models \tilde{M} he's exposed to is

$$\Pr(m|h, \mu_0, \tilde{M}) = \frac{\Pr(h|m, \mu_0) \frac{1}{|\tilde{M}|}}{\sum_{m' \in \tilde{M}} \Pr(h|m', \mu_0) \frac{1}{|\tilde{M}|}}.$$

The value-adjusted fit of model m given history h , prior μ_0 , and set of exposed-to models \tilde{M} is then

$$\text{VFit}(m|h, \mu_0, \tilde{M}) = \sum_{m' \in \tilde{M}} \Pr(m'|h, \mu_0, \tilde{M}) \mathbb{E}_{\mu(h, m')} [U^R(a(h, m), \omega)].$$

Contrast this with the non-value-adjusted fit of model m ,

$$\text{Fit}(m|h, \mu_0, \tilde{M}) = \Pr(h|m, \mu_0).$$

Note that the model that maximizes Fit is also the model that maximizes $\Pr(m|h, \mu_0, \tilde{M})$.

When is the model that maximizes VFit also the model that maximizes Fit? This is true, for example, when $|\tilde{M}| = 2$ and (i) $\omega \in [0, 1]$, $U^R(a, \omega) = -(a - \omega)^2$ or (ii) $\omega \in \{0, 1\}$, $U^R(a, \omega) = \text{Indicator}(a = \omega)$.³²

When does the model that maximizes VFit meaningfully differ from the model that maximizes Fit? One class of examples is situations where the receiver cares more about taking the correct action in some states than in others. For example, taking an herbal treatment to prevent cancer. When the treatment works is much more important to a patient than when it does not work.

To illustrate such a case, assume $|\tilde{M}| = 2$ and imagine $\omega \in \{0, 1\}$ and $U^R(a, \omega) = 100 \cdot \text{Indicator}(a = \omega = 1) + \text{Indicator}(a = \omega = 0)$. In this case, the receiver cares a lot more about taking the appropriate action in state $\omega = 1$ (remedy works) than in state $\omega = 0$ (remedy does not work). The model that maximizes VFit is often not the model that maximizes Fit. Consider a model m that induces the receiver to take the treatment ($a(h, m) = 1$, which implies $\mu(h, m)[1] > 1/101$) and a model m' , which induces the receiver to not take the treatment ($a(h, m') = 0$, $\mu(h, m')[1] <$

³²When $|\tilde{M}| > 2$, then even with these payoffs a person might take an action associated with a model with low posterior probability when this action is close-to optimal under other considered models.

1/101). VFit implies the receiver should go with m whenever

$$\Pr(m|h, \mu_0, \tilde{M}) > \frac{1 - 101 \cdot \mu(h, m')[1]}{101 \cdot (\mu(h, m)[1] - \mu(h, m')[1])}.$$

If $\mu(h, m)[1] = 1/10$ and $\mu(h, m')[1] = 0$, then this says that the receiver will take the treatment whenever he places posterior probability of at least $10/101$ (which is substantially below $1/2$) on m .

B.2 Receiver Skepticism

In the main text, we assume the receiver does not take persuaders' incentives into account in assessing proposed models. Alternatively, the receiver might be more skeptical of a persuader's proposed model when she knows that taking an action according to that model is in the persuader's interest.

Suppose the receiver is exposed to set of models \tilde{M} , which includes the receiver's default model given h . Let $m_j \in \tilde{M}$ denote the model proposed by persuader j . Say that *model m_j is in the persuader's interest given \tilde{M}* when $m_j \in \arg \max_{m_j \in \tilde{M}} V^j(h, m_j)$. That is, m_j is in the persuader's interest given \tilde{M} when, among models in \tilde{M} , it is the best one from the persuader's perspective.

Imagine that the receiver penalizes the persuader for proposing a model in her interest by requiring the model to fit the data *sufficiently* better than the default (or models proposed by other persuaders that are not in their interest). Specifically, denote the skepticism-adjusted fit of model m_j , SFit, by

$$\text{SFit}(m_j|h, \mu_0, \tilde{M}) = \begin{cases} (1 - \sigma) \cdot \Pr(h|m_j, \mu_0) & \text{if } m_j \text{ is in persuader } j\text{'s interest given } \tilde{M} \\ \Pr(h|m_j, \mu_0) & \text{otherwise (including for the default model),} \end{cases}$$

where $\sigma \in [0, 1)$. A σ -skeptical receiver discounts any model she is skeptical of by factor $(1 - \sigma)$. Higher σ corresponds to more skepticism on the part of the receiver.

A simple generalization of Proposition 1 characterizes beliefs the persuader is able to induce when the receiver is σ -skeptical and only has access to a default model in addition to the persuader's proposed model.

Proposition 9. *Fix d , μ_0 , and h and suppose the receiver is σ -skeptical. There is an M under which the persuader is able to induce target belief $\tilde{\mu} \in \Delta(\Omega)$ if and only if*

$$\tilde{\mu}(\omega) < \frac{\mu_0(\omega)}{\Pr(h|d)} \forall \omega \in \Omega \text{ and } V^S(m(\tilde{\mu}), h) < V^S(d(h), h) \quad (11)$$

or

$$\tilde{\mu}(\omega) < \frac{\mu_0(\omega)}{\Pr(h|d)} \cdot (1 - \sigma) \quad \forall \omega \in \Omega \text{ and } V^S(m(\tilde{\mu}), h) \geq V^S(d(h), h), \quad (12)$$

recalling that $m(\tilde{\mu})$ is a model that induces belief $\tilde{\mu}$.

This result collapses to Proposition 1 when the receiver is not skeptical ($\sigma = 0$). Greater receiver skepticism ($\sigma > 0$) places restrictions on beliefs the persuader is able to induce. When receiver skepticism is sufficiently large ($\sigma > 1 - \Pr(h|d)$), the receiver will never adopt the persuader's proposed model when it is known to be in the persuader's interest. When receiver skepticism is slightly smaller ($\sigma \approx 1 - \Pr(h|d)$), the receiver will only adopt a model that is known to be in the persuader's interest when it induces beliefs that are close to the receiver's prior. Indeed, when $\sigma = 1 - \Pr(h|d)$, then the only beliefs that satisfy (12) are $\tilde{\mu} = \mu_0$.

This result implies that the impact of model persuasion remains substantial even with significant receiver skepticism. A very skeptical receiver is willing to adopt a model known to be in the persuader's interest, only if it implies beliefs that are close to the receiver's prior. By pushing persuaders to propose models that say receivers should ignore data, receiver skepticism may then backfire—a very skeptical receiver is unwilling to consider objectively more accurate models that are in the persuader's interest and would lead him to change his mind.

B.3 Gathering, Revealing and Framing Data

This section extends our model to allow a persuader to gather and reveal evidence, which he can then frame ex post. We follow the Bayesian Persuasion literature by supposing the persuader must commit to revealing whatever information he collects. We depart from that literature by allowing the persuader to frame the evidence after he reveals it.

Suppose the persuader is able to gather and reveal *unambiguous* evidence that is not open to interpretation, as well as *ambiguous* evidence that is open to interpretation. Assume that the receiver is maximally open to persuasion in the face of any ambiguous evidence, but uses the true model as a default. Here, denote the ambiguous evidence by h .

To simplify the analysis and limit the number of cases considered, suppose that the state space is binary, $\Omega = \{0, 1\}$ and the persuader's objective is an increasing function of the probability the receiver attaches to the state being 1: $U^S(a, \omega) = f(\mu(1))$, where $f'(\cdot) > 0$. As an example, the states might correspond to whether product 0 or product 1 is better and the persuader might be selling product 1. It is natural that the receiver's demand for product 1 is increasing in the likelihood he attaches to the product being better.

Given these assumptions, Corollary 1 implies that the persuader's payoff for fixed likelihood

function π , prior μ_0 , and ambiguous data h is

$$f\left(\frac{\mu_0(1)}{\Pr(h|m^T)}\right).$$

This implies that the persuader's expected payoff is

$$\mathbb{E}_\pi \left[f\left(\frac{\mu_0(1)}{\Pr(h|m^T)}\right) \right],$$

given π .

When f is the linear function $f(x) = x$, the persuader's expected payoff reduces to

$$|H| \cdot \mu_0(1).$$

Here, $|H|$ equals the number of elements in the support of $\pi(\cdot|\omega)$ and can be thought of as a measure of the *amount* of ambiguous data that the seller reveals. This should be contrasted with the *informativeness* of the ambiguous data, which relates to what h reveals about ω under π . The ambiguous data is completely uninformative, for example, whenever $\pi(h|\omega)$ is independent of ω for all h .

When f is concave, the persuader's expected payoff is at most

$$f(|H| \cdot \mu_0(1)), \tag{13}$$

which is implemented by an ambiguous data process π that features $\pi(h|\omega) = 1/|H|$ for all h, ω .³³ So the persuader maximizes his payoff by collecting ambiguous data that is completely uninformative. Noting that (13) is increasing in $|H|$, we see that *the persuader maximizes his payoff by collecting and reporting as much of this uninformative, ambiguous data as possible*. Finally, noting that (13) is decreasing in mean-preserving spreads of $\mu_0(1)$ since f is concave, we see that *the persuader does not want to collect and reveal any unambiguous information*.

The intuition behind why the persuader wants to collect and reveal completely uninformative ambiguous information is that she is (weakly) risk averse and eliminates the risk of having difficulty framing the information by making it uninformative.³⁴ This is also the intuition for why the persuader, who is assumed to have no prior informational advantage over the receiver about ω , does not want to collect and reveal any unambiguous information. The intuition for why the persuader wants to collect and reveal as much uninformative ambiguous information as possible is

³³This π is not necessarily the unique solution to the maximization problem. But note that this π also minimizes the expected value of $\Pr(h|m^T)$ —creating the most expected space for persuasion—no matter μ_0 .

³⁴Note the contrast with the Bayesian Persuader, who wants to collect and reveal informative data.

that this maximizes the wiggle room the persuader has to frame the information.

C Proofs

Proof of Observation 1. Part 1. For each h , $a(h, m^T)$ must yield the receiver a higher payoff under belief $\mu(h, m^T)$ than $a(h, d)$ because the receiver could always choose $a(h, d)$. Averaging across h , the information component of persuasion must be positive for the receiver.

Part 2. It is obvious that the framing component of persuasion is weakly positive for the persuader and weakly negative for the receiver. So turn to the “if and only if” statement. Suppose first that the framing component is strictly positive for the persuader. Then $m(h) \neq m^T$ for some h in the support of (μ_0, π) . At this h , $a(h, m(h)) \neq a(h, m^T)$, which implies that $V^R(h, m(h)) < V^R(h, m^T)$. Since $V^R(\tilde{h}, m) \leq V^R(\tilde{h}, m^T)$ for all \tilde{h} , the framing component is strictly negative for the receiver.

For the other direction, suppose the framing component is strictly negative for the receiver. Then it must be that $m(h) \neq m^T$ for some h in the support of (μ_0, π) . At this h , $V^S(h, m(h)) > V^S(h, m^T)$ (otherwise the persuader would have chosen m^T). Since $V^S(\tilde{h}, m(\tilde{h})) \geq V^S(\tilde{h}, m^T)$ for all \tilde{h} , the framing component is strictly positive for the persuader. □

Proof of Lemma 1. To induce $\tilde{\mu}$,

$$\pi_m(h|\omega) = \frac{\tilde{\mu}(\omega)}{\mu_0(\omega)} \cdot K.$$

Here, K equals $\Pr(h|m, \mu_0)$. The maximum K such that $\pi_m(h|\omega) \leq 1 \forall \omega$ is $\min_{\omega \in \Omega} \mu_0(\omega) / \tilde{\mu}(\omega)$. □

Proof of Proposition 1. We’ll directly prove the result instead of invoking Lemma 1. Note that

$$\mu(h, m)[\omega] = \frac{\pi_m(h|\omega) \cdot \mu_0(\omega)}{\Pr(h|m)}$$

by Bayes’ Rule. Since $\pi_m(h|\omega) \leq 1$ and, under the constraint Eq (1), $\Pr(h|m) > \Pr(h|d)$, the persuader is not able to induce any beliefs that do not satisfy inequality (2). To see that for rich enough M the persuader is able to induce any beliefs that do satisfy this inequality, define m by

$$\pi_m(h|\omega) = \frac{\mu(h, m)[\omega]}{\mu_0(\omega)} \times \Pr(h|d) \forall \omega \in \Omega.$$

□

Proof of Proposition 2. Since persuasion is ineffective given (h, d, μ_0) , we have $\Pr(h|d) \geq \Pr(h|m)$ for all m such that $V^S(m, h) > V^S(d, h)$.

1. This means when $\Pr(h|\tilde{d}) > \Pr(h|d)$, we also have $\Pr(h|\tilde{d}) > \Pr(h|m)$ for all m such that $V^S(m, h) > V^S(d, h)$.

2. This means when receivers are maximally open to persuasion that no belief μ' simultaneously satisfies (2) and $V^S(\mu', h) > V^S(\mu(h, d), h)$. Given $\mu(\tilde{h}, \tilde{d}) = \mu(h, d)$ and $\Pr(\tilde{h}|\tilde{d}, \mu_0) > \Pr(h|d, \mu_0)$, this further implies that no belief μ' simultaneously satisfies (2) (now under $\tilde{h}, \tilde{d}, \mu_0$) and $V^S(\mu', \tilde{h}) > V^S(\mu(\tilde{h}, \tilde{d}), h) = V^S(\mu(h, d), h)$. This is because the modification to \tilde{h}, \tilde{d} tightens (2), so any belief that does not satisfy this constraint under h, d also does not satisfy it under \tilde{h}, \tilde{d} .

□

Proof of Proposition 3. Let $h^{i,j} = (h_1^i, h_2^j)$. We have

$$\begin{aligned} \mu(h^{i,j}, m)[\omega] &= \frac{\pi_{m^T}(h_1^i|\omega)\pi_m(h_2^j|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi_{m^T}(h_1^i|\omega')\pi_m(h_2^j|\omega')\mu_0(\omega')} \\ &< \frac{\pi_{m^T}(h_1^i|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi_{m^T}(h_1^i|\omega')\pi_d(h_2^j|\omega')\mu_0(\omega')}. \end{aligned} \quad (14)$$

The inequality follows from $\pi_m(h_2^j|\omega) \leq 1$ and

$$\Pr(h^{i,j}|m) = \sum_{\omega' \in \Omega} \pi_{m^T}(h_1^i|\omega')\pi_m(h_2^j|\omega')\mu_0(\omega') > \sum_{\omega' \in \Omega} \pi_{m^T}(h_1^i|\omega')\pi_d(h_2^j|\omega')\mu_0(\omega') = \Pr(h^{i,j}|d).$$

To establish the first part of the result, re-write the inequality above as

$$\mu(h^{i,j}, m)[\omega] < \frac{[\pi_{m^T}(h_1^i|\omega)/\pi_{m^T}(h_1^i|\omega^T)]\mu_0(\omega)}{\sum_{\omega' \in \Omega} [\pi_{m^T}(h_1^i|\omega')/\pi_{m^T}(h_1^i|\omega^T)]\pi_d(h_2^j|\omega')\mu_0(\omega')}.$$

For any $\omega \neq \omega^T$, the right hand side of this inequality tends to 0 as $i \rightarrow \infty$ by the fact that, for such ω , $[\pi_{m^T}(h_1^i|\omega)/\pi_{m^T}(h_1^i|\omega^T)] \rightarrow 0$ as $i \rightarrow \infty$. The result then follows.

To establish the second part of the result, first note that any belief satisfying (14) for all ω is implementable with model

$$\pi_m(h_2^j|\omega) = \frac{\mu(h, m)[\omega]}{\mu_0(\omega)\pi_{m^T}(h_1^i|\omega)} \cdot \Pr(h^{i,j}|d) \quad \forall \omega \in \Omega.$$

Second note that that, for fixed i , the right hand side of (14) tends to ∞ as $j \rightarrow \infty$, since $\pi_d(h_2^j|\omega') \rightarrow 0$ as $j \rightarrow \infty$. The result then follows from these two facts.

□

Proof of Corollary 1. Follows directly from Proposition 1.

□

Proof of Proposition 4. Suppose $m^T \notin \arg \max_{m \in M \cup \{d(h)\}} \Pr(h|m)$. (The proof of the remaining case is analogous.) Consider a candidate equilibrium where all persuaders propose \tilde{m} and receivers follow a tie-breaking rule on (5) where they adopt \tilde{m} in the case of a tie involving \tilde{m} . Then no persuader has an incentive to unilaterally deviate from \tilde{m} because no model in M fits h better than \tilde{m} : Any unilateral deviation does not impact the receiver's beliefs or actions, and hence the persuader's payoff.

□

Proof of Proposition 5. Suppose the fit requirement and condition (6) holds. Then there is an equilibrium where all persuaders propose an m that induces μ : condition (6) together with Lemma 1 implies that it is impossible for any persuader to unilaterally deviate to a model m' that benefits them for which $\Pr(h|m') \geq \Pr(h|m)$.

Conversely, it is clear that if the fit requirement does not hold then there is no equilibrium that induces μ . More interestingly, suppose μ is such that the fit requirement holds but condition (6) does not hold. Then there cannot be an equilibrium that induces μ : Suppose there was such an equilibrium and denote the equilibrium proposed model profile by (m^1, \dots, m^J) . Persuader j would have an incentive to deviate to proposing model \tilde{m}^j that induces a μ' satisfying $V^j(\mu') > V^j(\mu)$ and $\text{Movement}(\mu'; \mu_0) < \text{Movement}(\mu; \delta_0)$: by the first inequality the induced beliefs would be profitable for the persuader and by the second the model would in fact be adopted by the receiver. This contradicts the original profile being an equilibrium.

□

Proof of Corollary 2. The first part is obvious: A single persuader may be able to get the receiver to hold beliefs μ that the persuader prefers over μ_0 . Moreover, with competition, there are models the persuaders are able to propose that induce μ_0 and are more compelling than any model persuaders are able to unilaterally deviate to. In other words, it is obvious that μ_0 satisfies (6).

The second part follows from the fact that adding persuaders just adds more constraints that need to hold in order to satisfy (6).

For the third part, suppose μ_h is an equilibrium given h and a set of persuaders. Suppose further that the environment is such that it is possible for a persuader to strictly prefer belief μ_0 over all other beliefs given h . Now add such a persuader to the existing set of persuaders. Then μ_0 becomes the only equilibrium belief: it is the only belief that satisfies (6).

□

Proof of Corollary 3. This is a corollary of Propositions 1 and 5. By Equation (4), μ satisfies the (non-strategic) truth-teller constraint if and only if $\max_{\omega \in \Omega} \mu(\omega)/\mu_0(\omega) < 1/\Pr(h|m^T)$. By Equation (6), $\mu \neq \mu_h$ is an equilibrium belief with a strategic truth-teller if and only if $\max_{\omega \in \Omega} \mu(\omega)/\mu_0(\omega) < \max_{\omega \in \Omega} \mu_h(\omega)/\mu_0(\omega)$. It thus suffices to show that $\max_{\omega \in \Omega} \mu_h(\omega)/\mu_0(\omega) \leq 1/\Pr(h|m^T)$ with equality if and only if $\max_{\omega \in \Omega} \pi(h|\omega) = 1$. Note that the last inequality is equivalent to

$$\max_{\omega \in \Omega} \pi(h|\omega) \mu_0(\omega)/\mu_0(\omega) \leq 1,$$

which establishes the result.

□

Proof of Proposition 6. First we establish that Eq (9) is indeed the condition for there to both be a message that is compelling to the optimist and gets the pessimist to take action $a = 1$.

For a message to be compelling to the optimist, we need $\Pr^{\text{optimist}}(h|m) > \Pr^{\text{optimist}}(h|d)$, or,

equivalently,

$$\pi_m(h|b)(1 - \mu_0^{\text{optimist}}(g)) + \pi_m(h|g)\mu_0^{\text{optimist}}(g) > \Pr^{\text{optimist}}(h|d) \iff \quad (15)$$

$$\pi_m(h|b) > \frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{optimist}}(g)\pi_m(h|g)}{1 - \mu_0^{\text{optimist}}(g)}. \quad (16)$$

For a message to get the pessimist to take action $a = 1$, we need $\mu^{\text{pessimist}}(h, m)[g] \geq 1/2$, or, equivalently,

$$\frac{\pi_m(h|g)\mu_0^{\text{pessimist}}(g)}{\pi_m(h|g)\mu_0^{\text{pessimist}}(g) + \pi_m(h|b)(1 - \mu_0^{\text{pessimist}}(g))} \geq 1/2 \quad (17)$$

$$\frac{\mu_0^{\text{pessimist}}(g)\pi_m(h|g)}{1 - \mu_0^{\text{pessimist}}(g)} \geq \pi_m(h|b). \quad (18)$$

Since the right hand side of Eq (16) is decreasing in $\pi_m(h|g)$ and the left hand side of Eq (18) is increasing in $\pi_m(h|g)$, there is a message that simultaneously satisfies the two inequalities if and only if Eq (9) holds.

Now we establish that Eq (9) is a necessary and sufficient condition for there to be a message that gets both receivers to take action $a = 1$ when the optimist takes action $a = 0$ under their default interpretation. To establish sufficiency, first note that any message that gets the pessimist to take action $a = 1$ also gets the optimist to take action $a = 1$ if compelling to the optimist. It remains to show that there is such a message that is compelling to the pessimist. For a message to be compelling to the pessimist, we need

$$\pi_m(h|b) > \frac{\Pr^{\text{pessimist}}(h|d) - \mu_0^{\text{pessimist}}(g)\pi_m(h|g)}{1 - \mu_0^{\text{pessimist}}(g)}.$$

For there to be such a message that also gets the pessimist to invest we need the right hand side of this inequality to be less than the left hand side of Eq (18) when $\pi_m(h|g) = 1$. But this follows from the pessimist being persuadable.

To establish necessity, this is clear when both the optimist and pessimist take action $a = 0$ under their default interpretations. When only the optimist takes action $a = 0$ under their default interpretation we need to show that when Eq (9) fails to hold we cannot find a message that (i) is compelling to the optimist, (ii) gets the optimist to take action $a = 1$, and (iii) is not compelling to the pessimist. To see this, for the message to be compelling to the optimist but not the pessimist we would need

$$\frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{optimist}}(g)\pi_m(h|g)}{1 - \mu_0^{\text{optimist}}(g)} < \pi_m(h|b) \leq \frac{\Pr^{\text{pessimist}}(h|d) - \mu_0^{\text{pessimist}}(g)\pi_m(h|g)}{1 - \mu_0^{\text{pessimist}}(g)}.$$

But the existence of a message that satisfies this condition when Eq (9) fails to hold further implies

that

$$\frac{\mu_0^{\text{pessimist}}(g)}{1 - \mu_0^{\text{pessimist}}(g)} < \frac{\Pr^{\text{pessimist}}(h|d) - \mu_0^{\text{pessimist}}(g)}{1 - \mu_0^{\text{pessimist}}(g)},$$

which contradicts the pessimist being persuadable. By the same argument, whenever the sender cannot send a single message that gets both receivers to take $a = 1$ she cannot send a menu of messages that gets both receivers to take action $a = 1$.

Finally, when the optimist takes action $a = 1$ under their default interpretation then there is necessarily a message that gets both receivers to take action $a = 1$. Either (i) a message that gets the pessimist to take action $a = 1$ (which exists under the assumption that the pessimist is persuadable) is not compelling to the optimist; or (ii) such a message is compelling to the optimist. Under (i), the optimist continues to take action $a = 1$. Under (ii), the optimist will also take action $a = 1$ since any message that gets the pessimist to take action $a = 1$ will also get the optimist to take action $a = 1$ if it is compelling to the optimist. □

Proof of Corollary 4. If receivers share the same prior, then Eq (9) boils down to

$$\frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{optimist}}(g)}{1 - \mu_0^{\text{optimist}}(g)} < \frac{\mu_0^{\text{optimist}}(g)}{1 - \mu_0^{\text{optimist}}(g)},$$

which holds by the assumption that the optimist is individually persuadable.

If receivers share the same default interpretation, then

$$\begin{aligned} \frac{\mu_0^{\text{pessimist}}(g)}{1 - \mu_0^{\text{pessimist}}(g)} &> \frac{\Pr^{\text{pessimist}}(h|d) - \mu_0^{\text{pessimist}}(g)}{1 - \mu_0^{\text{pessimist}}(g)} \\ &= \frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{pessimist}}(g)}{1 - \mu_0^{\text{pessimist}}(g)} \\ &\geq \frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{optimist}}(g)}{1 - \mu_0^{\text{optimist}}(g)}, \end{aligned}$$

which means that Eq (9) holds. The first line follows from the pessimist being individually persuadable, the second from the optimists and pessimists sharing a default interpretation, and the third from the optimist having a weakly larger prior on g than the pessimist. □

Proof of Lemma 2. By the law of total expectation,

$$\mathbb{E}[\theta(x)|m, h] = \theta(x) \cdot p(x|c_m(x)) + \sum_{x' \in c_m(x) \setminus \{x\}} \theta(x') \cdot p(x'|c_m(x)).$$

The persuader will choose the $m \in M$ that maximizes this expectation. □

Proof of Proposition 7. For the first part, let

$$\begin{aligned}\hat{\theta}_\delta(x) &= \max_{m \in M} \theta(x) \cdot p_\delta(x|c_m(x)) + \sum_{x' \in c_m(x) \setminus \{x\}} \theta(x') \cdot p_\delta(x'|c_m(x)). \\ \hat{\theta}_\delta(x|m) &= \theta(x) \cdot p_\delta(x|c_m(x)) + \sum_{x' \in c_m(x) \setminus \{x\}} \theta(x') \cdot p_\delta(x'|c_m(x)).\end{aligned}$$

We can re-express $\hat{\theta}_\delta(x|m)$ as a weighted average of $\theta(x)$ and $\sum_{x' \in c_m(x) \setminus \{x\}} \theta(x') \cdot p_\delta(x'|c_m(x)) / (1 - p_\delta(x|c_m(x)))$, with respective weights $p_\delta(x|c_m(x))$ and $1 - p_\delta(x|c_m(x))$. Since

$$\frac{p_\delta(x'|c_m(x))}{1 - p_\delta(x|c_m(x))}$$

is independent of δ for each x' , δ only impacts $\hat{\theta}_\delta(x|m)$ through these weights.³⁵ Consequently, $\hat{\theta}_\delta(x|m)$ is strictly decreasing in δ for all m such that $\hat{\theta}_1(x|m) > \theta(x)$. This then implies that $\hat{\theta}_\delta(x)$ is strictly decreasing in δ whenever there exists an $m \in M$ such that $\hat{\theta}_1(x|m) > \theta(x)$ and is constant in δ otherwise (since the persuader would always propose a model with $c_m(x) = \{x\}$ in the latter case).

For the second part, if $\tilde{\psi}$ is a mean-preserving spread of ψ , then

$$\mathbb{E}_{\tilde{\psi}} \left[\max_{m \in M} \theta(x) \cdot p(x|c_m(x)) + \sum_{x' \in c_m(x) \setminus \{x\}} \theta(x') \cdot p(x'|c_m(x)) \right] > \mathbb{E}_{\psi} \left[\max_{m \in M} \theta(x) \cdot p(x|c_m(x)) + \sum_{x' \in c_m(x) \setminus \{x\}} \theta(x') \cdot p(x'|c_m(x)) \right]$$

by Jensen's inequality. □

Proof of Proposition 8. If $m(x) = m(x') \forall x, x'$, then the common model is the finest partition of X (since this is the optimal model for the x that maximizes $\theta(x)$). Under the finest partition, $\hat{\theta}(x) = \theta(x)$ for all $x \in X$.

If there exists x, x' such that $m(x) \neq m(x')$, then the persuader is not always proposing the finest partition. By the persuader's revealed preference, there exists an x such that $\hat{\theta}(x) > \theta(x)$. □

Proof of Proposition 9. Follows the logic of the proof of Proposition 1. □

³⁵Indeed, it is easy to check that

$$\frac{p_\delta(x'|c_m(x))}{1 - p_\delta(x|c_m(x))} = \frac{p(x'|c_m(x))}{1 - p(x|c_m(x))}.$$

References

- Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz**, “Fragility of Asymptotic Agreement Under Bayesian Learning,” *Theoretical Economics*, 2016, 11 (1), 187–225.
- Akerlof, George A and Robert J Shiller**, *Phishing for Phools: The Economics of Manipulation and Deception*, Princeton University Press, 2015.
- Alonso, Ricardo and Odilon Câmara**, “Persuading Voters,” *The American Economic Review*, 2016, 106 (11), 3590–3605.
- Andreassen, Paul B**, “Judgmental Extrapolation and Market Overreaction: On the Use and Disuse of News,” *Journal of Behavioral Decision Making*, 1990, 3 (3), 153–174.
- Anwar, Shamena, Patrick Bayer, and Randi Hjalmarrsson**, “The Impact of Jury Race in Criminal Trials,” *The Quarterly Journal of Economics*, 2012, 127 (2), 1017–1055.
- , —, and —, “The Role of Age in Jury Selection and Trial Outcomes,” *The Journal of Law and Economics*, 2014, 57 (4), 1001–1030.
- Aragones, Enriqueta, Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler**, “Fact-Free Learning,” *American Economic Review*, 2005, 95 (5), 1355–1368.
- Arnold, David, Will Dobbie, and Crystal S Yang**, “Racial Bias in Bail Decisions,” *The Quarterly Journal of Economics*, 2018, 133 (4), 1885–1932.
- Bajgrowicz, Pierre and Olivier Scaillet**, “Technical Trading Revisited: False Discoveries, Persistence Tests, and Transaction Costs,” *Journal of Financial Economics*, 2012, 106 (3), 473–491.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny**, “A Model of Investor Sentiment1,” *Journal of financial economics*, 1998, 49 (3), 307–343.
- Barron, Daniel and Michael Powell**, “Markets for Rhetorical Services,” 2018.
- Bénabou, Roland, Armin Falk, and Jean Tirole**, “Narratives, Imperatives and Moral Reasoning,” 2018.
- Bergstresser, Daniel, John MR Chalmers, and Peter Tufano**, “Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry,” *The Review of Financial Studies*, 2008, 22 (10), 4129–4156.

- Bhattacharya, Utpal, Andreas Hackethal, Simon Kaesler, Benjamin Loos, and Steffen Meyer**, “Is Unbiased Financial Advice to Retail Investors Sufficient? Answers From a Large Field Study,” *The Review of Financial Studies*, 2012, 25 (4), 975–1032.
- Billingsley, Randall S and Don M Chance**, “Benefits and Limitations of Diversification Among Commodity Trading Advisors,” *Journal of Portfolio Management*, 1996, 23 (1), 65.
- Broockman, David and Joshua Kalla**, “Durably Reducing Transphobia: A Field Experiment on Door-to-Door Canvassing,” *Science*, 2016, 352 (6282), 220–224.
- Bruner, Jerome**, “The Narrative Construction of Reality,” *Critical inquiry*, 1991, 18 (1), 1–21.
- Busby, Ethan, DJ Flynn, James N Druckman, and Scott Hall**, “Studying Framing Effects: Existing Research and Lingering Questions.”
- Cain, Daylian M, George Loewenstein, and Don A Moore**, “The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest,” *The Journal of Legal Studies*, 2005, 34 (1), 1–25.
- Campbell, John Y**, “Household Finance,” *The journal of finance*, 2006, 61 (4), 1553–1604.
- Carhart, Mark M**, “On Persistence in Mutual Fund Performance,” *The Journal of finance*, 1997, 52 (1), 57–82.
- Chalmers, John and Jonathan Reuter**, “Is Conflicted Investment Advice Better Than No Advice?,” Technical Report, National Bureau of Economic Research 2012.
- Chater, Nick and George Loewenstein**, “The Under-Appreciated Drive for Sense-Making,” *Journal of Economic Behavior & Organization*, 2016, 126, 137–154.
- Cheung, Yin-Wong and Menzie David Chinn**, “Currency Traders and Exchange Rate Dynamics: A Survey of the US Market,” *Journal of International Money and Finance*, 2001, 20 (4), 439–471.
- Chong, Dennis and James N Druckman**, “Framing Theory,” *Annu. Rev. Polit. Sci.*, 2007, 10, 103–126.
- Crawford, Vincent P and Joel Sobel**, “Strategic Information Transmission,” *Econometrica: Journal of the Econometric Society*, 1982, pp. 1431–1451.
- DellaVigna, Stefano and Ethan Kaplan**, “The Fox News Effect: Media Bias and Voting,” *The Quarterly Journal of Economics*, 2007, 122 (3), 1187–1234.

- and **Matthew Gentzkow**, “Persuasion: Empirical Evidence,” *Annu. Rev. Econ.*, 2010, 2 (1), 643–669.
- Diaconis, Persi and David Freedman**, “On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities,” *The Annals of Statistics*, 1990, pp. 1317–1327.
- DiFonzo, Nicholas and Prashant Bordia**, “Rumor and Prediction: Making Sense (but Losing Dollars) in the Stock Market,” *Organizational Behavior and Human Decision Processes*, 1997, 71 (3), 329–353.
- Egan, Mark L, Gregor Matvos, and Amit Seru**, “Arbitration With Uninformed Consumers,” Technical Report, National Bureau of Economic Research 2018.
- Eliaz, Kfir and Ran Spiegel**, “A Model of Competing Narratives,” 2019.
- , — , and **H Thysen**, “Strategic Interpretations,” Technical Report, mimeo 2019.
- Epstein, Larry G and Martin Schneider**, “Learning Under Ambiguity,” *The Review of Economic Studies*, 2007, 74 (4), 1275–1303.
- Esponda, Ignacio and Demian Pouzo**, “Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models,” *Econometrica*, 2016, 84 (3), 1093–1130.
- Eyster, Erik and Michele Piccione**, “An Approach to Asset Pricing Under Incomplete and Diverse Perceptions,” *Econometrica*, 2013, 81 (4), 1483–1506.
- Fama, Eugene F and Kenneth R French**, “Luck Versus Skill in the Cross-Section of Mutual Fund Returns,” *The Journal of Finance*, 2010, 65 (5), 1915–1947.
- Fischhoff, Baruch**, “Hindsight is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty.,” *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 1 (3), 288.
- Fisher, Walter R**, “The Narrative Paradigm: In the Beginning,” *Journal of communication*, 1985, 35 (4), 74–89.
- Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein**, “Channeled Attention and Stable Errors,” 2018.
- Galbraith, John Kenneth**, *The New Industrial State* 1967.
- Galperti, Simone**, “Conceal, Surprise, and Change Worldview,” 2016.

- Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny**, “Money Doctors,” *The Journal of Finance*, 2015, 70 (1), 91–114.
- Gentzkow, Matthew and Emir Kamenica**, “Competition in Persuasion,” 2012.
- and — , “Competition in Persuasion,” *The Review of Economic Studies*, 2017, 84 (1), 300–322.
- and **Jesse M Shapiro**, “Media Bias and Reputation,” *Journal of Political Economy*, 2006, 114 (2), 280–316.
- and — , “Competition and Truth in the Market for News,” *Journal of Economic Perspectives*, 2008, 22 (2), 133–154.
- Gershman, Samuel J**, “How to Never be Wrong,” *Psychonomic Bulletin & Review*, 2018, pp. 1–16.
- Goffman, Erving**, *Frame Analysis: An Essay on the Organization of Experience.*, Harvard University Press, 1974.
- Hackethal, Andreas, Michael Haliassos, and Tullio Jappelli**, “Financial Advisors: A Case of Babysitters?,” *Journal of Banking & Finance*, 2012, 36 (2), 509–524.
- Harrison, J Michael and David M Kreps**, “Speculative Investor Behavior in a Stock Market With Heterogeneous Expectations,” *The Quarterly Journal of Economics*, 1978, 92 (2), 323–336.
- Heidhues, Paul, Botond Kőszegi, and Philipp Strack**, “Unrealistic Expectations and Misguided Learning,” *Econometrica*, 2018, 86 (4), 1159–1214.
- Kamenica, Emir and Matthew Gentzkow**, “Bayesian Persuasion,” *The American Economic Review*, 2011, 101 (6), 2590–2615.
- Kass, Robert E and Adrian E Raftery**, “Bayes Factors,” *Journal of the American Statistical Association*, 1995, 90 (430), 773–795.
- Kindleberger, Charles P and Robert Aliber**, *Manias, Panics, and Crashes*, 6 ed., Palgrave Macmillan, 2010.
- Koehler, Jonathan J and Molly Mercer**, “Selection Neglect in Mutual Fund Advertisements,” *Management Science*, 2009, 55 (7), 1107–1121.
- Lakoff, George**, *Don’t Think of an Elephant!: Know Your Values and Frame the Debate*, Chelsea Green Publishing, 2004.

- and **Mark Johnson**, *Metaphors We Live By*, University of Chicago press, 1980.
- Levy, Gilat and Ronny Razin**, “An Explanation-Based Approach to Combining Forecasts,” Technical Report, mimeo 2019.
- Li, Sheng**, “The Influence of Financial Advisors on Return Chasing,” 2015.
- Lo, Andrew W, Harry Mamaysky, and Jiang Wang**, “Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation,” *The Journal of Finance*, 2000, 55 (4), 1705–1765.
- Malmendier, Ulrike and Devin Shanthikumar**, “Are Small Investors Naive About Incentives?,” *Journal of Financial Economics*, 2007, 85 (2), 457–489.
- Milgrom, Paul and John Roberts**, “Relying on the Information of Interested Parties,” *The RAND Journal of Economics*, 1986, pp. 18–32.
- Milgrom, Paul R**, “Good News and Bad News: Representation Theorems and Applications,” *The Bell Journal of Economics*, 1981, pp. 380–391.
- Morris, Stephen**, “Speculative Investor Behavior and Learning,” *The Quarterly Journal of Economics*, 1996, 111 (4), 1111–1133.
- Mullainathan, Sendhil**, “Thinking Through Categories,” 2002.
- and **Andrei Shleifer**, “The Market for News,” *American Economic Review*, 2005, 95 (4), 1031–1053.
- , **Joshua Schwartzstein, and Andrei Shleifer**, “Coarse Thinking and Persuasion,” *The Quarterly journal of economics*, 2008, 123 (2), 577–619.
- Olszewski, Wojciech**, “Self-Esteem Preferences and Imaginary Effects in Information Processing,” 2018.
- Ortoleva, Pietro**, “Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News,” *American Economic Review*, 2012, 102 (6), 2410–36.
- Pennington, Nancy and Reid Hastie**, “Evidence Evaluation in Complex Decision Making,” *Journal of personality and social psychology*, 1986, 51 (2), 242.
- and —, “Explanation-Based Decision Making: Effects of Memory Structure on Judgment,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1988, 14 (3), 521.

- and — , “Explaining the Evidence: Tests of the Story Model for Juror Decision Making.” *Journal of personality and social psychology*, 1992, 62 (2), 189.
- Phillips, Blake, Kantura Pukthuanthong, and P Raghavendra Rau**, “Past Performance May Be an Illusion: Performance, Flows, and Fees in Mutual Funds,” *Critical Finance Review*, 2016, 5 (2), 351–398.
- Pons, Vincent**, “Will a Five-Minute Discussion Change Your Mind? A Countrywide Experiment on Voter Choice in France,” *American Economic Review*, 2018, 108 (6), 1322–63.
- Roberts, Seth and Harold Pashler**, “How Persuasive is a Good Fit? A Comment on Theory Testing,” *Psychological review*, 2000, 107 (2), 358.
- Schulz, Laura E and Jessica Sommerville**, “God Does Not Play Dice: Causal Determinism and Preschoolers’ Causal Inferences,” *Child development*, 2006, 77 (2), 427–442.
- Schwartzstein, Joshua**, “Selective Attention and Learning,” *Journal of the European Economic Association*, 2014, 12 (6), 1423–1452.
- Severin, J Werner and James W Tankard**, “Communication Theory: Origin, Methods, and Uses in The Mass Media,” 2001.
- Shaton, Maya**, “The Display of Information and Household Investment Behavior,” 2017.
- Shiller, Robert J**, *Irrational Exuberance*, Princeton university press, 2015.
- , “Narrative Economics,” *American Economic Review*, 2017, 107 (4), 967–1004.
- Smith, David M, Christophe Faugère, and Ying Wang**, “Head and Shoulders Above the Rest? The Performance of Institutional Portfolio Managers who use Technical Analysis,” in “Research in Finance,” Emerald Group Publishing Limited, 2014, pp. 167–189.
- Spiegler, Ran**, “Bayesian Networks and Boundedly Rational Expectations,” *The Quarterly Journal of Economics*, 2016, 131 (3), 1243–1290.
- Taylor, Mark P and Helen Allen**, “The Use of Technical Analysis in the Foreign Exchange Market,” *Journal of International Money and Finance*, 1992, 11 (3), 304–314.
- Weick, Karl E**, *Sensemaking in Organizations*, Vol. 3, Sage, 1995.
- Zhou, Dexin**, “Good News in Numbers,” 2017.