NBER WORKING PAPER SERIES

USING MODELS TO PERSUADE

Joshua Schwartzstein
Adi Sunderam

Working Paper 26109
http://www.nber.org/papers/w26109

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2019

We thank Ned Augenblick, Francesca Bastianello, Pedro Bordalo, Stefano DellaVigna, Erik Eyster, Tristan Gagnon-Bartsch, Nicola Gennaioli, Matt Gentzkow, Robert Gibbons, Brian Hall, Sam Hanson, Botond Koszegi, Deepak Malhotra, Ulrike Malmendier, Sendhil Mullainathan, Ramana Nanda, Matthew Rabin, Jesse Shapiro, Andrei Shleifer, Kelly Shue, Erik Stafford, Jeremy Stein, Dmitry Taubinsky, and seminar participants at UC Berkeley, MIT, Harvard, BEAM 2019, and EBEM 2019 for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Using Models to Persuade
Joshua Schwartzstein and Adi Sunderam
NBER Working Paper No. 26109
July 2019
JEL No. G11,G4,L15,M3

ABSTRACT

We present a framework for analyzing "model persuasion." Persuaders influence receivers' beliefs by proposing models (likelihood functions) that specify 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 data, fixing receivers' prior beliefs over states of the world. A key tradeoff facing persuaders is that models that better fit the data induce less movement in receivers' beliefs. Model persuasion sometimes makes the receiver worse off than he would be in the absence of persuasion. Even when the receiver is exposed to the true model, the wrong model often wins because it better fits the past. The receiver is most misled by persuasion when there is a lot of data that is open to interpretation and exhibits randomness, as this gives the persuader "wiggle room" to highlight false patterns. With multiple persuaders, competition pushes towards models that provide overly good fits, which tend to neutralize the data by leading receivers to view it as unsurprising. With multiple receivers, a persuader is more effective when he can send tailored, private messages than a menu of public messages. We illustrate with examples from finance, business, politics, and law.

Joshua Schwartzstein
Baker Library 461
Harvard Business School
Soldiers Field
Boston, MA 02163
jschwartzstein@hbs.edu

Adi Sunderam
Harvard Business School
Baker Library 359
Soldiers Field
Boston, MA 02163
and NBER
asunderam@hbs.edu

1 Introduction

Persuasion frequently involves an expert providing a "model" of the world, an interpretation of known data. 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 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," suggesting 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 data 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 long-term averages, 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 debating climate change, one side might argue that extreme weather events provide evidence of global warming, while the other might argue that they reflect "noise" in an inherently unpredictable process. In politics, there are "spin rooms" where campaigns 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, firms propose frames that positively highlight known aspects of their products. The car-rental company Avis, 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 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 before taking an action needs to interpret a history of outcomes that may be informative about a payoff-relevant state of nature. Persuaders propose models for interpreting the history to the receiver. A model is a likelihood function that maps the history to posterior beliefs for the receiver, in turn leading the receiver to take certain actions.

[1]The 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.

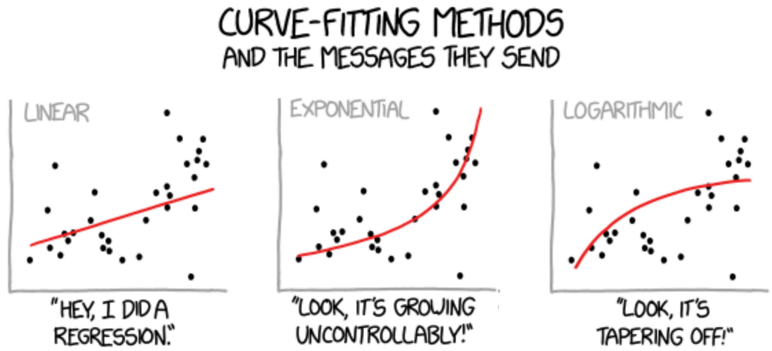


Figure 1: Stylized Example of Model Persuasion from xkcd.com
Source: https://xkcd.com/2048/

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 endogenize.

A key ingredient of our framework is that we assume a proposed model is more compelling than an alternative if it fits the receiver's knowledge—the data plus the receiver's prior—better than the alternative. Essentially, we assume that the receiver performs a "Bayesian hypothesis test": from the set of models he is exposed to, he picks 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' (with likelihood function $\pi_{m'}$) given data h and prior μ_0 over states ω if:

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

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 (iii) feature the most "determinism" as documented in work on developmental and cognitive psychology (Schulz and Sommerville 2006; Gershman 2018).[2]

[2]There 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 our basic insights, consider a simple example, which we will return to throughout the paper. An investor is 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 that startup's probability of success, 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 all of the entrepreneur's startups. Also assume for the purpose of the example that this is the true model.

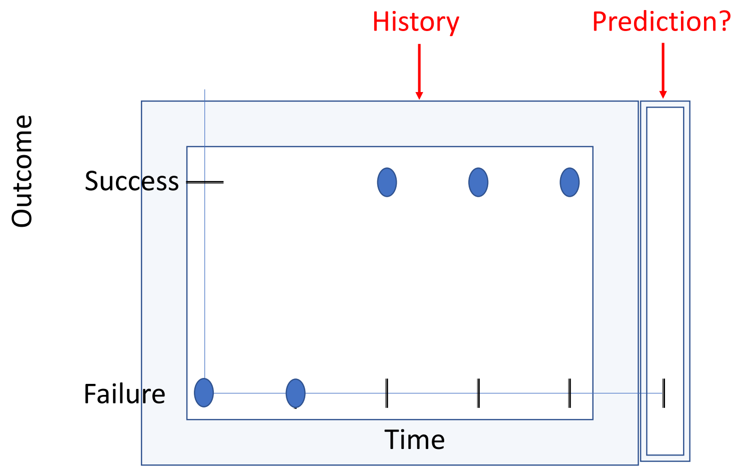
The persuader wants the investor to invest, and thus wishes to propose models that maximize the investor's posterior expectation of θ. Suppose the receiver is willing to entertain the possibility that "this time is different". That is, the receiver will entertain models suggesting that the entrepreneur's success probability was re-drawn from the uniform distribution on $[0, 1]$ 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.[3]

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

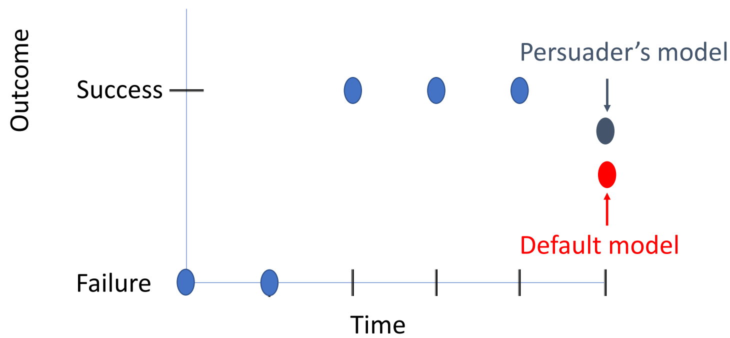
A few important assumptions drive the results. First, persuaders are more able than receivers to come up with models to make sense of data. Household investors rely on financial advisers to help interpret mutual fund performance data; voters rely on pundits to interpret polling data; jurors rely on experts and lawyers to interpret evidence at a trial; patients rely on doctors to 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

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.

[3]Section 2 presents this example more formally. Later sections also establish what the persuader is able to convince the receiver of if the latter is willing to entertain models beyond "this time is different".



(a) Setup



(b) Persuader vs. Default Model

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

a default model. But, crucially, receivers do not generate new stories themselves. Rather, they need experts to supply them.[4] 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 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 our setting. However, our results are qualitatively robust to simply requiring models proposed by more biased experts to satisfy stricter goodness-of-fit tests. 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

[4]This 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 non-trivial to correct for them: methods in machine learning and economics are still being developed to deal with these issues.[5]

Section 2 sets up our general framework and provides some basic properties. We show that the impact of model persuasion combines information and framing components. The information component—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. Thus, model persuasion may reduce receivers' welfare relative to if they used 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 the data is unsurprising, which means beliefs should not move much in response to it. The constraint that a persuader's model be more compelling than 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 single 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.[6] 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

[5] See, e.g., Barberis et al. (2015), and Harvey (2017).

[6] Receivers 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 truthteller (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 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 the case when the 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 scope 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.

Section 5 then considers the impact of competition between persuaders. Competition pushes persuaders to propose models that overfit the data, given the receiver's prior over states. 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. 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 fit better.

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, competing persuaders often neutralize 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 to heterogeneous receivers. We provide examples where the persuader can get two receivers to each take a desired action (e.g., make an investment desired by the persuader) with tailored, private messages, but cannot do so when constrained to send a common public message (or menu of messages). The key factor is the similarity in priors and default interpretations across receivers: it is harder to simultaneously persuade dissimilar audience members.

Section 7 considers examples in finance, law, and business. We apply our framework to shed light on what makes technical analysis in financial markets so compelling and why people follow biased advice in finance and business, even when exposed to better advice. These applications illustrate how to take our key ideas to the data. In a more theoretical application, we examine a canonical example in the information-persuasion literature—persuading a jury—and show how incorporating model persuasion modifies the analysis.

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.[7] In this paper, we provide a more portable and systematic treatment of this idea, which 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 Thysen (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 certain messages are persuasive. 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

[7] At 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; DeMarzo, Vayanos, and Zwiebel 2003; 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 and the cheap-talk persuasion literature (Crawford and Sobel 1982), assumes that the receiver is dogmatic that they are using the right model. By contrast, 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 1940; 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 that imply only certain facts are relevant. However, it frequently operates quite differently, by proposing models that invite the receiver to consider all the evidence, but interpret it differently than they would under the true model. For instance, a bullish stock market analyst who is 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-pickers would naturally combine the information provided by each, 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 wish to influence the beliefs of receivers, which depend on both the past history of outcomes, as well as the model used to interpret this history. 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 as well as multiple receivers.

Broadly, the setup is as follows. The persuader proposes a model to the receiver. If the receiver finds the proposed model more compelling than his default model, meaning that the proposed model better explains available data, then the receiver adopts it. The receiver then 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 $U^R(a,\omega)$.[8] 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.[9]

The receiver 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 (we suppress the dependence of d on h when it does not cause confusion) 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 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

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

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.

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:

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

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, solving

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

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

[8]In applications, we will sometimes relax the assumption that Ω is finite.

[9]It many of our applications, $U^S(\cdot)$ is independent of ω, meaning that the persuader's optimal action is independent of the true likelihood π. For example, an advertiser always wants to sell their product. In these situations, it is without loss of generality for our analysis to assume the persuader knows the data-generating process.

A few points about the default model merit discussion. First, we allow the default to be history dependent to capture the idea that a receiver might only come up with a default explanation for the data after seeing it. Second, while many of our results hold for all defaults, we sometimes analyze two special cases. In one (extreme) case, the default renders the data uninformative, so the receiver sticks with his prior in the absence of persuasion (i.e., $\mu(h, d(h)) = \mu_0 \ \forall \ h$) and finds any model in M more compelling than the default.[10] This captures 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. The second case, which we analyze in Section 4, is that the default is the true model. This captures situations where the true model is readily accessible, perhaps because there are academics or watchdogs actively pushing it. Moreover, it is a natural assumption in applications where the default model is not obvious.

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 and business. 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 technology had caused a structural shift to very rapid growth was popularized in part by financial analysts with incentives that rewarded high stock prices (Shiller 2015). We will analyze another example of highlighting strips of data in Section 7.

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

[10]Under 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.

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. If the persuader proposes the K-model, where S of those K flips came up heads, and the receiver adopts it, then the receiver believes 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). Appendix B analyzes this example in detail.

Example 2 (Highlighting characteristics)**.** While some of the examples in the paper fit within the highlighting strips framework, many other real world examples involve highlighting characteristics. For instance, Tesla is both a technology company and an automotive company. Suppose the receiver is an investor with knowledge of these characteristics who is trying to forecast Tesla's future performance. Persuaders may propose models in which one or both characteristics help forecast success. For instance, a persuader aiming to inflate Tesla's value could propose a partition grouping Tesla with technology companies.

The broad intuition here is also 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.'"[12] 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.

In Appendix B, we show how to formalize such examples in our framework.

[11] Suppose 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

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

[12] http://archive.fortune.com/magazines/fortune/fortune_archive/2001/05/14/302981/index.htm

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. 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 beneficial to 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. More broadly, the prior captures any knowledge of the receiver that the persuader cannot influence without referencing the data. In settings where there are existing models of information persuasion with Bayesian receivers, we can take the state space Ω and prior μ_0 to be the same as assumed in those models. We perform such an exercise in Section 7. In settings where no such models exist, analyst judgment is required to specify Ω and μ_0.

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

That said, receivers are unlikely to take everything persuaders say at face value. Section 4 considers the case where persuaders have to compete with truthtellers, such as consumer watchdogs, and thus must propose models that are more compelling than the truth. 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. As we show formally in Appendix D, this will tend to reduce receivers' sensitivity to all data. Thus, receiver skepticism may in fact backfire, while still leaving room for misleading persuasion.

Fourth, we assume that receivers select rather than average models. This is consistent with psychological evidence on "thinking through categories", for example as discussed in Mullainathan (2002). It is also natural given we assume that receivers do not have a prior over models. And it allows us to crisply communicate the role of model persuasion. However, in some settings, the idea that receivers average over models they are exposed to has appeal as well. In Appendix D, we draw out implications of model averaging. We show that it shares key qualitative implications with our model selection setup. We also show how averaging models makes receivers more persuadable than selecting models in some situations and less persuadable in others.[13]

Finally, our assumption that receivers adopt models that fit the data well, embodied in Eq. (1), drives many of our results. It 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. As is well known from the literature on Bayesian Model Selection (e.g., Kass and Raftery 1995), there is a sense in which our formulation then does not mechanically favor "more complex" models or ones with more degrees of freedom.[14] As we discuss further below, our formulation favors models under which the history is unsurprising in hindsight. Such models typically do not include unspecified degrees of freedom, but rather plug in values that best explain the history. In addition, our 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 D.[15]

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 people *should* find a good model fit persuasive (e.g., Roberts and Pashler 2000), but as far as we can tell little debate that they *do* find a good fit persuasive.

[13]For instance, if the receiver's default perfectly fits the data, then Eq. (1) implies that there is no scope for persuasion if the receiver is a model selecter because the persuader cannot propose a better fitting model. However, if the receiver is a model averager, the persuader can still influence him with a model that fits worse than his default. If the receiver's default fits poorly, on the other hand, then there is less scope for persuasion when the receiver is a model averager. In this case, there is much scope for persuasion if the receiver is a model selecter because the persuader can easily propose a better fitting model. However, if the receiver is a model averager, he still puts some weight on his default, regardless of how well the model the persuader proposes fits.

[14]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: $\Pr(2 \text{ failures}|2 \text{ parameter model}) = 1/4 < 1/3 = \Pr(2 \text{ failures}|1 \text{ parameter model})$. Bayes Factors are approximated by the Bayesian Information Criterion, which penalizes degrees of freedom, again suggesting our assumption does not mechanically favor more complex models.

[15]This alternative favors models that prescribe actions that have greater utility consequences when those models are true. For example, consider a persuader who attempts to frame evidence on the efficacy of a low-cost herbal treatment to prevent cancer. With a value-adjusted notion of fit, the receiver may find a model compelling if it supports taking the supplement, even if it does not fit the data as well as a model that says the treatment is ineffective. Failing to take the treatment when the model is true has far greater utility consequences than getting the treatment when the model is not true.

2.4 Basic Framework Properties

Model persuasion has two effects. First, it enables receivers to act on more information. Second, it frames that information. To formalize these two effects let

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

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 $V^S(h,m) \equiv \mathbb{E}_{\mu_h}\left[U^S(a(h,m),\omega)\right]$ denote the sender's payoff as a function of the history and adopted model.

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). We can decompose the impact of persuasion on the receiver as:

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

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 the default renders the data uninformative, 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 A.)

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 where default models fit better than the true model (e.g., receivers overfit the data on their own).

As illustrated in, e.g., Appendix B.1, the two components of persuasion may also 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 is incentivized to persuade 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,[16] and he will invest only if his belief moves to at least 50%.

The broker tries to convince Pat to invest by framing available data. For simplicity, suppose the only data the broker is able to frame is the active fund's returns (high or low) last year. Formally, this is a restriction on the set of models M; Pat is unwilling to entertain models implying that other data (e.g., the fund manager's educational background) is relevant. In general, specifying what data is frameable is a key modeling choice.[17] Finally, assume that Pat's default model is that past

[16] 24% of all actively managed have outperformed their passive benchmarks over the last 10 years (Morningstar Active/Passive Barometer, 2019).

[17] In some applications, a natural constraint is to only allow the persuader to frame data that are relevant under the true model. In other cases, however, analyst judgment is required to specify h. For example, in cases where no data

returns are somewhat informative. He believes that good funds have a higher probability of high past returns than bad funds: $\pi_d(\text{high returns}|\text{good}) = \pi_d(\text{low returns}|\text{bad}) = 75\%$.

Suppose that the active fund Pat and the broker are considering has high past returns. Under his default model, Pat believes it is moderately surprising to observe a fund with high past returns:

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

Under his default, Pat will 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, to dictate investing. What can the broker convince Pat of?

First, note that broker cannot get Pat to believe anything she wants. For instance, she cannot simply assert that Pat's prior is wrong and the fraction of good funds is higher than 20% without referencing the data. The prior captures all of Pat's knowledge that is not conditional on the data. Pat's beliefs only change in response to data, framed by the broker.

Further, even though the broker has great flexibility to frame the data, 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 active fund's high past returns mean it is good for sure: $\pi_m(\text{high returns}|\text{good}) = 100\%, \pi_m(\text{high returns}|\text{bad}) = 0$. Under this model, high past returns are even more surprising than under Pat's default:

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

That is, if the broker tries to tell Pat that high past returns are strongly associated with a relatively rare event (a good fund), Pat thinks the story is too good to be true. He finds his default model—that high past returns are not that rare but also not perfectly informative about the quality of the fund—a better explanation of the data.

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 $\pi_m(\text{high returns}|\text{bad})$

is relevant under the true model (i.e., situations where outcomes are completely random), there may still be data that persuaders can frame because receivers believe such data to be potentially relevant.

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. This model avoids the too-good-to-be-true problem by acknowledging that high returns do not imply that the fund is good for sure. But it does imply that high returns are what Pat should expect to see if the fund is good.

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, any model saying high past returns are informative enough that he should invest has the too-good-to-be-true problem.

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 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.[18] 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 *movement* the model causes 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

[18]Formally, 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. 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: $\Pr(\text{very high returns}|m) = 20\% > \Pr(\text{very high returns}|d) = 7\%$.

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 (when $\tilde{\mu} = \mu_0$) and ∞ (when $\tilde{\mu}$ places positive probability on a state the prior μ_0 says is zero probability).

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

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. Thus, 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 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 (once the data h that the receiver is willing to consider is specified).

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 in Appendix B—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*

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

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 and generalizes the limits on persuasion we found in the example of Pat and the active mutual fund. 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 that the state is one that the receiver's prior puts low probability on. 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.

3.2 When Are Receivers Persuadable?

Proposition 1 also has implications for *when* receivers are persuadable. Returning to the case of Pat, since the target belief to induce investment is $\mu(\text{good}) = 50\%$ and the prior is $\mu_0(\text{good}) = 20\%$, movement to the target equals $50\%/20\% = 5/2$. This implies that the maximum fit of any model that gets Pat to invest is $2/5 = 40\%$. The broker will only persuade Pat to invest if Pat's default has a worse fit.

This means that even when the broker uses the best possible argument and Pat is willing to entertain any model, Proposition 1 implies that he can only be persuaded to invest under certain conditions. Specifically, 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. Second, the probability of the particular realization of this data observed must also be sufficiently low under his default.

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.

The first three points are fairly straightforward. How well the default fits the data determines the tightness of constraint Eq. (1). For instance, 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. Following this logic, persuadability is affected by the expected (ex ante, prior to h being realized) difficulty receivers will have explaining the data under their default interpretation. Receivers are persuadable that this time is different when interpreting financial market data because they are often puzzled by what they see; they are not when considering whether the sun will rise tomorrow because they always have a ready explanation for what they see. Finally, openness to interpretation, captured by the size of the model space M naturally impacts persuadability. For example, a persuader would have a hard time convincing an audience that, all else equal, being older reduces mortality risk. Audiences are likely only willing to entertain models that suggest that mortality risk rises with age. On the other hand, the 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. We provide formal details on these three points in Appendix C.

The fourth point is less obvious. 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 always ineffective. In contrast, if we fix unambiguous data that does not rule out any state (i.e., $\pi_{m^T}(h_1|\omega) > 0 \; \forall \; \omega \in \Omega$), then for "enough" ambiguous data (i.e., where $\Pr(h_2|d)$ is sufficiently low) the persuader is able to induce any target belief provided the model space M is sufficiently rich.[19] For instance, for a US presidential candidate, long track records are both a blessing and a curse, providing lots of ambiguous material for both the candidate and opponents to frame. In contrast, information that a potential candidate is 29 years old is unambiguous in this context—it pins down that the candidate

[19]The 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.

cannot be President.

To formalize this intuition, define "more data" as follows. Consider sequences of histories $(h_1^i)_{i=1,\ldots,\infty}$ and $(h_2^i)_{i=1,\ldots,\infty}$ and think of higher i as representing more data. Assume that the likelihood of a history falls asymptotically as the length of the history increases: $\pi_{m^T}(h_x^i|\omega) \to 0$ as $i \to \infty$ for $x = 1, 2$. In addition, assume that the true state is identified asymptotically as the length of the history increases: there is a $\omega^T \in \Omega$ such that $\pi_{m^T}(h_x^i|\omega)/\pi_{m^T}(h_x^i|\omega^T) \to 0$ as $i \to \infty$ for all $\omega \neq \omega^T$ and $x = 1, 2$. Both of these properties hold for almost all sequences generated by π_{m^T} under standard 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 2. *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 $\tilde{\mu} \in \Delta(\Omega)$ 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 that does not rule out any state in Ω and any target belief $\tilde{\mu} \in \Delta(\Omega)$, 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 *un*able 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. This finding contrasts with an intuition from "information persuasion" that more publicly available data if anything limits the scope for 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

$$\Pr(h|m) > \Pr(h|m^T), \tag{3}$$

where m is the model proposed by the persuader. The persuader takes this constraint, which we refer to as the truthteller constraint, into account in choosing what model to propose.[20]

When the true model is 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*

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

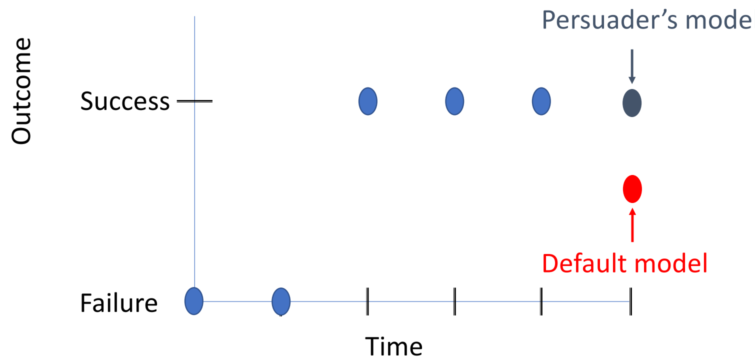
and is not able to induce beliefs that do not satisfy this inequality.

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, we return to the entrepreneur example from the introduction. There we showed that the persuader can 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 higher future success probability, 99%. Since the true model does not fit the data all that well, the persuader is able to move the receiver's beliefs a lot in response to the data.

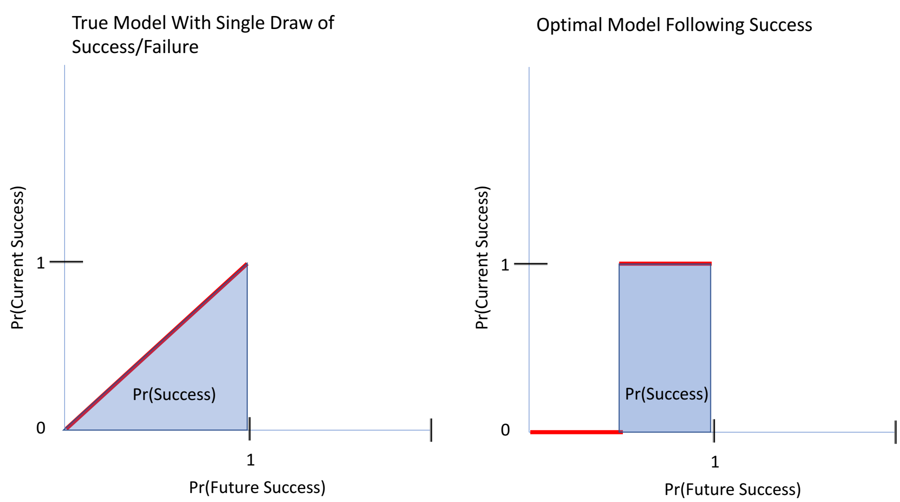
What do models look like when the receiver is maximally open to persuasion? To illustrate we simplify the entrepreneur example so that there is only a single previous startup, which was successful. The left panel of Figure 3b shows the true model relating the probability of success of the entrepreneur's first startup to the probability of success of the second. Since a common success probability θ governs the success of each startup, the curve relating $\Pr$(Current Success) to $\Pr$(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 to the probability of success of the 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

[20] 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. Of course, in 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.



(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

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 truthteller constraint: the largest $\tilde{\theta}$ such that an initial success is as well explained by persuader's model as 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 best the persuader is able to do is to get the investor to estimate that the entrepreneur's next startup will be successful with probability $3/4$. Again, 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.

When the default model is the true model, the comparative statics we derived in Section 3 on when receivers are persuadable 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 truthteller constraint is relatively weak for such events. This result suggests that we should see a lot of persuasive activity creating narratives surrounding "tail events", such as particularly good or bad performance of a company or worker.

Continuing to apply the comparative statics from Section 3 to the case where the true model is the default, the framework predicts 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 truthteller 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. This point comes alive in the technical analysis illustration in Section 7.

Taken together, these results suggest that a truthteller does not constrain the persuader much, particularly when the data are random. Appendix B.1 analyzes 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. We show how the qualitative results derived assuming that receivers are maximally open to persuasion carry over to this case.

5 Competition Between Persuaders

In the previous section we considered a single persuader who had to compete with the truth. This section considers competition between persuaders more broadly. 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). With more than two persuaders, the receiver goes with the proposed model, including the default, that maximizes $\Pr(h|\cdot)$. In the case of ties involving the default model, we assume that the receiver goes with the default model. Otherwise, we assume that the equilibrium determines the receiver's tie-breaking procedure.

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

Proposition 3. *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 $d(h)$ 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, the proposition indicates that competition may 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. 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.

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 set of equilibrium beliefs will be a function of the persuaders' incentives, the data, and the receiver's 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 4. *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) Fit$(\mu; h, \mu_0) > \Pr(h|d)$ and (ii) for all persuaders $j = 1, 2, \ldots, J$*

$$V^j(\mu', h) > V^j(\mu, h) \Rightarrow \textit{Movement}(\mu'; \mu_0) \geq \textit{Movement}(\mu; \mu_0), \tag{6}$$

recalling that 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. The result implies 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 is that competition need not lead to more accurate beliefs. While competition 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) often pushes towards the truth, with model persuasion 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 pushes receivers towards adopting models that overfit the past, thus rendering it uninformative about the state. A model that 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.[21] 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.

The tendency to associate market movements with narratives, noted among popular observers of financial markets, illustrates the idea that models that overfit the past emerge in equilibrium:[22]

> 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, and the other's is strictly decreasing in the posterior probability. Then Proposition 4 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 4. Competition between persuaders with opposing interests pushes receivers to adopt models that view the data as uninformative. That is, competition *neutralizes the data.*

[21] However, 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.

[22] Vermont Royster (Wall Street Journal, "Thinking Things Over Abaft of the Market," January 15, 1986).

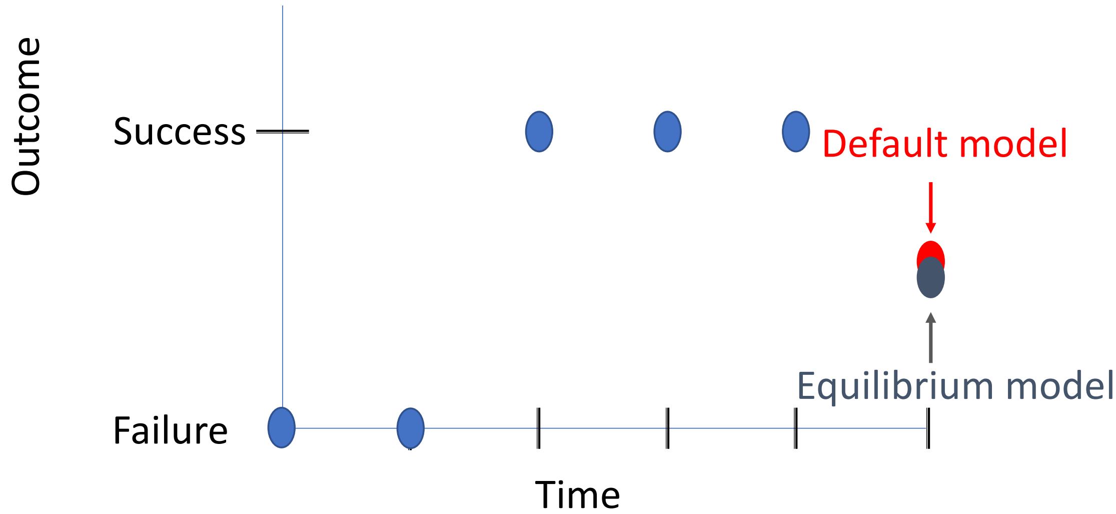


Figure 4: Competition between persuaders. One wants an investor to believe the entrepreneur's next startup will be successful and the other that it will be unsuccessful.

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 persuaders who have an interest in moving beliefs. This stubbornness seems particularly puzzling in light of recent work showing that short conversations are surprisingly effective at changing minds about political issues (Broockman and Kalla 2016, Pons 2018). Our results suggest that when all persuaders have identical incentives they will indeed have success in getting receivers to adopt models that lead them to overreact to data. However, when they have 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 4 is that a *strategic* truthteller—who wants the receiver to hold correct beliefs but is not constrained to propose the true model—is more effective than a non-strategic truthteller. Specifically, assume the strategic truthteller's payoff equals $v > 0$ if the receiver ends up with correct beliefs μ_h and equals 0 otherwise. Whenever the true model cannot perfectly explain the data, the strategic truthteller constrains equilibrium beliefs by more than the non-strategic truthteller.

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

1. *Suppose* $\max_{\omega\in\Omega}\pi(h|\omega) < 1$. *If* $\mu \neq \mu_h$ *is an equilibrium belief then there is a model inducing that belief that also satisfies the (non-strategic) truthteller constraint. However, there is a belief* μ *that is induced by a model that satisfies the (non-strategic) truthteller constraint but is not an equilibrium belief.*

2. *Suppose* $\max_{\omega\in\Omega}\pi(h|\omega) = 1$ *and the default is the true model. In this case* μ *is an equilib-*

rium belief if and only if there is a model inducing that belief that satisfies the (non-strategic) 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 29 times better than the true model. A persuader who wants to induce optimistic beliefs is much more constrained competing with a strategic truthteller who promotes this model than with a non-strategic truthteller promoting the true model. Indeed, the best the persuader can do against a strategic truthteller is to induce the belief that the entrepreneur's success probability is 76% for her next project—well below the 99% forecast the persuader is able to induce if competing with a non-strategic truthteller.

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.

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 the case he where he can 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 $\pi_m(\text{high returns}|\text{bad}) = (64\% - 1 \times 40\%)/60\% = 40\%$. For Pat, this model implies the probability of observing high returns is

$$\Pr{}^{\text{Pat}}(\text{high returns}|m) = \pi_m(\text{high returns}|\text{good}) \times 20\% + \pi_m(\text{high returns}|\text{bad}) \times 80\% = 52\%. \tag{7}$$

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

$$\Pr{}^{\text{Pat}}(\text{good}|m, \text{high returns}) = \pi_m(\text{high returns}|\text{good}) \frac{\mu_0^{\text{Pat}}(\text{good})}{\Pr^{\text{Pat}}(\text{high returns})} = 38\%. \tag{8}$$

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 situations 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 the analog of (7) and will now not invest by the analog of (8).[23]

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 such problems to the persuader arise, we generalize the example slightly. Retain a binary state space $\Omega = \{b, g\}$ (e.g., "bad", "good"), 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 $\pi_d^{\text{pessimist}}$. We assume that the optimist and pessimist are both individually persuadable but would not invest in the absence of persuasion. By Proposition 1 this means $1/2 > \mu_0^j(g) > \Pr^j(h|d)/2$ for $j =$ optimist, pessimist.

Proposition 5. *Suppose in the absence of data (i.e., under their priors), neither the optimist nor the pessimist would choose $a = 1$, but both are persuadable to choose $a = 1$ at history h. 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 $a = 1$:*

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

Corollary 4. *Under the assumptions of Proposition 5, the persuader is able to send a message that gets both receivers to take action $a = 1$ if they share the same prior, $\mu_0^{\textit{optimist}} = \mu_0^{\textit{pessimist}}$, or the same default interpretation, $\pi_d^{\textit{optimist}} = \pi_d^{\textit{pessimist}}$.*

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

[23] Intuitively, 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.

to send even a *menu* of public messages that simultaneously persuades 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 communications textbooks like Severin and Tankard (2001) emphasize, 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. The first and third concern real-world examples of model persuasion from finance and business on persuading investors and business managers. In between, we return to a motivating example from the Bayesian Persuasion literature, persuading jurors, and show how incorporating model persuasion alters conclusions from the analysis.

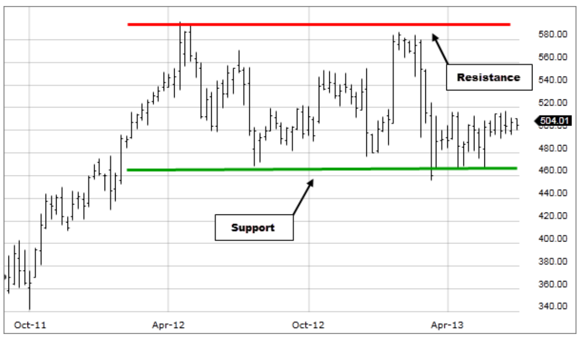
7.1 Persuading an Investor: Technical Analysis

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 5a 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 5a is an illustrative example from the brokerage center Fidelity's "Learning Center" for investors, Figure 5b shows a real world example of technical analysis from TradingView.com. The analysis was done by a brokerage firm that sells investment advice and services to clients. Using data on Amazon's stock price in 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.

Technical analysis is also a situation where there are many competing narratives. Figure 5c shows another real world example of technical analysis using the same data as Figure 5b on Amazon's stock price in 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.

This kind of analysis is extremely common in financial markets. Major brokerage services catering to individual investors, including Fidelity, E-Trade, Charles Schwab, Merrill Lynch, and TD Ameritrade, offer their clients tools for technical analysis. 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.[24] 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).



(a) Illustration of Support and Resistance.

(b) Technical Analysis: Long Amazon.

(c) Technical Analysis: Short Amazon.

Figure 5: Technical Analysis

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. Indeed, as one example, the support and resistance model looks so compelling in Figures 5a and 5b because the historical data are used to determine the support and resistance levels after seeing the data.

[24]For 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).

In Appendix E.1, we formally show that in our framework the support/resistance model describing Amazon's stock price in Figure 5b is more compelling than the default model that Amazon's stock price follows a random walk.[25] We assume the underlying state the investor is trying to learn about is the probability that Amazon's stock price rises on January 29. The investor's prior is that this probability is either 25%, 50%, or 75%, and her prior puts equal weight on all three possibilities. Under the default model, Amazon's stock price is a random walk so the history always implies that the probability Amazon's stock price rises is 50%. The support/resistance model proposed by the persuader says that Amazon's stock price follows a random walk until it hits either the support or the resistance. If it hits the resistance, then the probability the stock price rises is 25%. If it hits the support, then the probability it rises is 75%. The key flexibilities available to the persuader are in (i) picking the support and resistance levels after seeing the data and (ii) selecting the sample period over which the model applies. In Appendix E.1, we formalize these models in the notation of our framework and show that the data are four times more likely under the support/resistance model than the default model.

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). Kamenica and Gentzkow focus on 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. For example, closing arguments are used not to introduce new evidence, but rather to push narratives for interpreting the evidence. The view that juror decision-making is influenced by narratives that 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 a juror finds most compelling frames such evidence as reinforcing his prior beliefs.

The primitives of this applied model are taken directly from the Kamenica and Gentzkow (2011) setup. 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 $v > 0$ if he convicts a guilty defendant or acquits an innocent defendant; payoff $-a < 0$ if he convicts an innocent defendant; and payoff $-b < 0$ if he acquits a guilty defendant. The juror will then optimally follow a cutoff rule where he convicts a defendant if and only if his posterior

[25]We analyze the support/resistance model rather than, say, a head and shoulders model because it is simpler to formalize. We conjecture that the head and shoulders model will also be more compelling than a random walk.

beliefs about guilt $\mu(g)$ are above a certain threshold. 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 (2011) emphasize that each side may tailor an investigative strategy (e.g., interviewing certain witnesses) to benefit from Bayesian persuasion because payoffs are naturally non-linear in beliefs. And when attorneys compete, there is full revelation (Kamenica and Gentzkow 2012). What happens with model persuasion?

Suppose everyone sees the same evidence h and the representative juror is maximally open to persuasion. Then Proposition 4 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 all 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 the defense would benefit from proposing a model that confirms the juror's prior, which contradicts μ being the juror's equilibrium beliefs.[26] While the impact of Bayesian persuasion is for the juror to make the correct decision in equilibrium, the impact of model persuasion is to neutralize the evidence. For evidence that is open to interpretation, model persuasion keeps the juror's beliefs in a range where he would make the same decision as in the absence of such evidence. 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.

The intuition is that models resonate with the juror if they frame the evidence to fit what the juror already believes to be true. A juror who thinks that the defendant is very likely to be innocent will find alternative explanations for damning evidence compelling; a juror who thinks that the defendant is likely to be guilty will find arguments that the same evidence is diagnostic of guilt compelling. This basic force carries through to situations where jurors are not maximally open to persuasion: model persuasion then mitigates, but does not eliminate, the impact of evidence on juror decisions. To illustrate, 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 (with skilled attorneys) will be for jurors to arrive at the same conclusion that they would in the absence of the arguments and data those arguments frame.[27]

[26] Formally, condition (6) is violated.

[27] This is broadly consistent with the literature on competitive framing more generally (Busby, Flynn, and Druckman 2018), which finds that equally strong competing frames "cancel out".

Relaxing the assumption of a common prior, these results 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 the fact that trials and arbitrations reveal evidence that Bayesians should agree on. Our results also suggest that parties will take juror and judge characteristics into account when proposing narratives on how to interpret the facts.

7.3 Persuading a Client: Advice in Individual Investing and Business

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 the recommendations of advisors, who have incentives to give biased advice. 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 the broader idea that people make mistakes by following biased advice is incomplete. They are likely exposed to advice from multiple sources, including advice that would lead to better decisions if followed. This raises a question: Why do individuals follow the biased advice? Our model offers a particular answer to the question: they find biased advice more compelling than the truth.

The key intuition was presented in the simple example of investment advice from Section 3. We develop a more elaborate formulation in Appendix E.2. Following Proposition 3, we show in the appendix that investors will tend to follow biased advice when unbiased advice comes from persuaders whose incentives are to push correct models, rather than compelling models. This may help explain empirical results showing that investors sometimes choose not to follow unbiased investment advice that would improve their portfolio performance even if they obtain it (e.g., Bhattacharya et al. 2012).

The idea that misleading advice is followed because it looks compelling in the data is not limited to finance. It may play an important role in business advice books that conduct *ex post* analyses to uncover factors that make businesses successful. For instance, consider the well-known book "Good to Great" by Jim Collins (Collins 2001), consistently ranked one of the ten most influential and best selling management books of all time. The book provides management advice arrived at by the following procedure:

> We identified companies that made the leap from good results to great results and sustained those results for at least fifteen years ... we then compared the good-to-great companies to comparison companies to discover the essential and distinguishing

factors at work. (page 3)

In particular, the author selected 11 firms that previously had 15 years of exceptional stock market performance. He then identified factors that made those 11 firms unique ex post and proposed that if other firms followed the example of the 11 firms he studied, they too could become great.[28] This design was explicit. As the author writes:

> We developed all of the concepts in this book by making empirical deductions directly from the data. We did not begin this project with a theory to test or prove. We sought to build a theory from the ground up. (page 10)

Advice generated by this procedure sounds compelling in part because the story seems compelling in the data.

Figure 6 shows the cumulative stock market performance of the 11 firms selected relative to the aggregate stock market, reproducing Figure 2 in the book. Year 0 on the horizontal axis corresponds to the year that Collins argues the companies made the leap from good to great; Year 15 corresponds to the last year that Collins includes in his analysis; Years 15-30 follow the book's publication. Collins's selected firms did vastly outperform the market in years 0-15—that is why Collins chose to study them. Thus, the argument that there is something special about these firms looks compelling in the data.

In Appendix E.3, we formalize this argument by extending the "this time is different" model we used in the introduction in the context of entrepreneurship. In this case, we assume that the mean (log) stock return for the 11 good-to-great firms is drawn from a normal distribution. Realized annual returns are equal to the mean return plus normally distributed noise. We compare the default model—that the mean return for the firms is draw once and is constant across the 30-year sample Collins studied—with a "this time is different model"—that the mean return was drawn once at the beginning of the sample and again after 15 years (at Year 0 in the figure). We find that the "this time is different model" is 8 times more likely to explain the data than the default model. At the time the book was written, Collins's argument that the 11 companies he focused on "made the leap" from good to great when he says they did is much more compelling than the argument that they were just lucky.

But perhaps the companies were just lucky. Since the book was published in 2001, we can now extend the sample by nearly 20 years. As shown in Figure 6, over these intervening years,

[28] The firms were Abbott Laboratories, Circuit City Stores, Fannie Mae, Gillette Company, Kimberly-Clark, Kroger, Nucor, Philip Morris, Pitney Bowes, Walgreens, and Wells Fargo. The identified factors include humility of the management team, having the right people, willingness to confront unpleasant facts, faith that obstacles can be overcome, focusing on simple strategic plans, a culture of discipline, and adoption of carefully selected technologies.

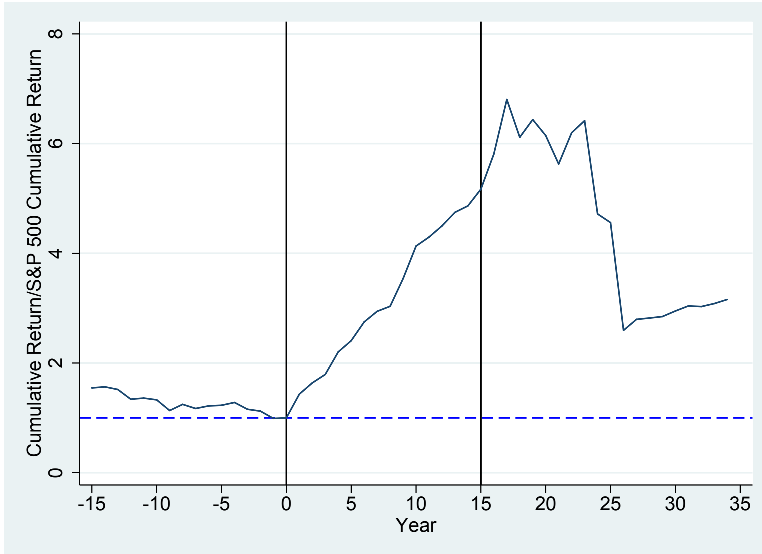


Figure 6: Performance of "Good-to-Great Firms" Relative to the Aggregate Stock Market

the firms studied have had slightly below average performance.[29] In the extended sample, the "this time is different model" is 25% less likely to explain the data than the default model. However, in a remarkable feat of model persuasion, Collins's book remains popular: it was a top-5 bestselling business book in 2016-2017, 15 years after publication.[30]

8 Discussion

This paper presents a framework for analyzing model persuasion—persuasion that 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 a lot of data that is open to interpretation and when outcomes are close to random. The presence of truthtellers does not eliminate the impact of misleading persuasion because there are wrong models that better fit the past than the

[29]Others have noted that following the book's advice has not been a recipe for success (e.g., Rosenzweig 2007; Levitt 2008; Niendorf and Beck 2008) for other businesses. In particular, it did not go unnoticed in the late 2000s that the "good to great" companies included Circuit City and Fannie Mae, which failed. But previous critiques did not hone in on our explanation for why the book is so compelling. In response to these earlier critics of the book, Collins is summarized by the New York Times as writing (https://www.nytimes.com/2009/05/24/business/24collins.html):

> [T]he merits of analyzing the reasons for a company's long winning streak—or, for that matter, a sport's team's—are just as valid even if the company or team can't maintain the winning formula. If people eat right and exercise, then stop doing so, it doesn't make those habits any less valid ...

One interpretation is that Collins is now promoting a new model: the companies made the leap from good to great when he says they did, but subsequently fell from great back to good, or worse.

[30]https://www.forbes.com/sites/jeffkauflin/2017/06/20/the-years-5-bestselling-leadership-books-and-why-theyre-so-great/#2685927e3ac0

correct model. Similarly, rather than promoting the truth, competition favors models that overfit the past, leading beliefs to underreact to evidence.

The framework is amenable to a number of applications and extensions. Some extensions, e.g., on receiver sophistication and model averaging, are explored in Appendix D. But many others are possible. Under suitable assumptions, for example, our framework could be used to study self persuasion. Endogenizing default models and, in particular, better understanding when receivers "see patterns" and overfit the data on their own could shed light on the contexts in which it is particularly difficult to persuade others.

Another interesting set of applications considers the problem from the perspective of a policymaker trying to effectively regulate model persuasion. Our analysis here suggests that simply requiring that consumers be exposed to the right model (e.g., "past performance is not indicative of future results") is often not enough to guarantee that consumers are not misled. An open question is whether there are effective regulations beyond heavy-handed methods that directly limit the messages persuaders are allowed to send.

Our static model has 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 a commit to a model in advance of some data being released? Proposition 2 suggests one intuition: Interpreting data that is closed to interpretation as arising from previous narratives supplied by a persuader, it suggests that fresh information liberates the persuader from his previous statements. But this result does not address dynamic considerations the persuader might face. For instance, supplying a myopically optimal model might be constraining in the future. Having a static model also abstracts from phenomena such as "preemptive framing", where a party benefits from proposing the first narrative surrounding new data. Similarly, fixing the data abstracts from potential interactions between Bayesian and model persuasion. For example, our results showing that more information may benefit the persuader (even in the presence of a truthteller) suggest that model persuaders prefer to collect and reveal information ex ante if they can then frame it in beneficial ways ex post. We provide an illustrative example in Appendix D.4. But analyzing 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 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. Let $h^{i,j} = (h_1^i, h_2^j)$. We have

$$\begin{aligned} \mu(h^{i,j}, m)[\omega] &= \frac{\pi_{m^T}(h_1^i|\omega)\pi_m(h_2^j|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi_{m^T}(h_1^i|\omega')\pi_m(h_2^j|\omega')\mu_0(\omega')} \\ &< \frac{\pi_{m^T}(h_1^i|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi_{m^T}(h_1^i|\omega')\pi_d(h_2^j|\omega')\mu_0(\omega')}. \end{aligned} \tag{10}$$

The inequality follows from $\pi_m(h_2^j|\omega) \le 1$ 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 $i \to \infty$ by the fact that, for such ω, $[\pi_{m^T}(h_1^i|\omega)/\pi_{m^T}(h_1^i|\omega^T)] \to 0$ as $i \to \infty$. The result then follows.

To establish the second part of the result, first note that any belief satisfying (10) for all ω is implementable with model

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

Second note that that, for fixed i, the right hand side of (10) tends to ∞ as $j \to \infty$, since $\pi_d(h_2^j|\omega') \to 0$ as $j \to \infty$. The result then follows from these two facts.

□

Proof of Corollary 1. Follows directly from Proposition 1.

□

Proof of Proposition 3. Consider a candidate equilibrium where all persuaders propose $\tilde{m}$ and receivers follow a tie-breaking rule maximizing $\Pr(h|\cdot)$ 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 4. Suppose conditions (i) and (ii) hold. Then there is an equilibrium where all persuaders propose the best-fitting m that induces μ and receivers follow a tie-breaking procedure where they favor m over any model that fits equally well: 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 $\Pr(h|m') > \Pr(h|m)$.

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 $(m^1, \ldots, m^J)$. Some persuader j would have an incentive to deviate to proposing the best-fitting model $\tilde{m}^j$ that induces a μ' satisfying $V^j(\mu', h) > V^j(\mu, h)$ and Movement$(\mu'; \mu_0) <$ Movement$(\mu; \mu_0)$: by the first inequality the induced beliefs would be profitable for the persuader and by the second it would in fact would be adopted by the receiver (by Lemma 1). 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 between at least two persuaders, 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 4. By Equation (4), there is a model inducing μ which satisfies the (non-strategic) truthteller 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 truthteller only if $\max_{\omega\in\Omega}\mu(\omega)/\mu_0(\omega) \leq \max_{\omega\in\Omega}\mu_h(\omega)/\mu_0(\omega)$. When the default model is the true model, $\mu \neq \mu_h$ is an equilibrium belief if and only if the latter condition holds and $\text{Fit}(\mu; h, \mu_0) > \Pr(h|m^T)$ (by Proposition 4), which is equivalent to $\max_{\omega\in\Omega}\mu(\omega)/\mu_0(\omega) < 1/\Pr(h|m^T)$ (by Lemma 1).

It 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, after re-arranging and using Bayes' rule, 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 5. 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,

$$\pi_m(h|b)(1-\mu_0^{\text{optimist}}(g)) + \pi_m(h|g)\mu_0^{\text{optimist}}(g) > \Pr^{\text{optimist}}(h|d) \iff \tag{11}$$

$$\pi_m(h|b) > \frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{optimist}}(g)\pi_m(h|g)}{1-\mu_0^{\text{optimist}}(g)}. \tag{12}$$

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,

$$\frac{\pi_m(h|g)\mu_0^{\text{pessimist}}(g)}{\pi_m(h|g)\mu_0^{\text{pessimist}}(g) + \pi_m(h|b)(1-\mu_0^{\text{pessimist}}(g))} \geq 1/2 \tag{13}$$

$$\frac{\mu_0^{\text{pessimist}}(g)\pi_m(h|g)}{1-\mu_0^{\text{pessimist}}(g)} \geq \pi_m(h|b). \tag{14}$$

Since the right hand side of Eq (12) is decreasing in $\pi_m(h|g)$ and the left hand side of Eq (14) 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 (14) 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.

□

Appendices for Online Publication

B Highlighting Strips and Characteristics

B.1 Highlighting Strips of Data

This section analyzes the highlighting strips of data example. 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 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. We use the "tc" superscript is short-hand for "truthteller constrained". 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 7 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., $(0 \ldots, 0, 1, \ldots 1)$. As an example, 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 truthteller 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 $-(\hat{\theta} - \theta)^2 + (1/2 - \theta)^2$.[31] Adding a truthteller benefits

[31] Applying our general formula for the impact of persuasion with a true-model default would, for given h, yield

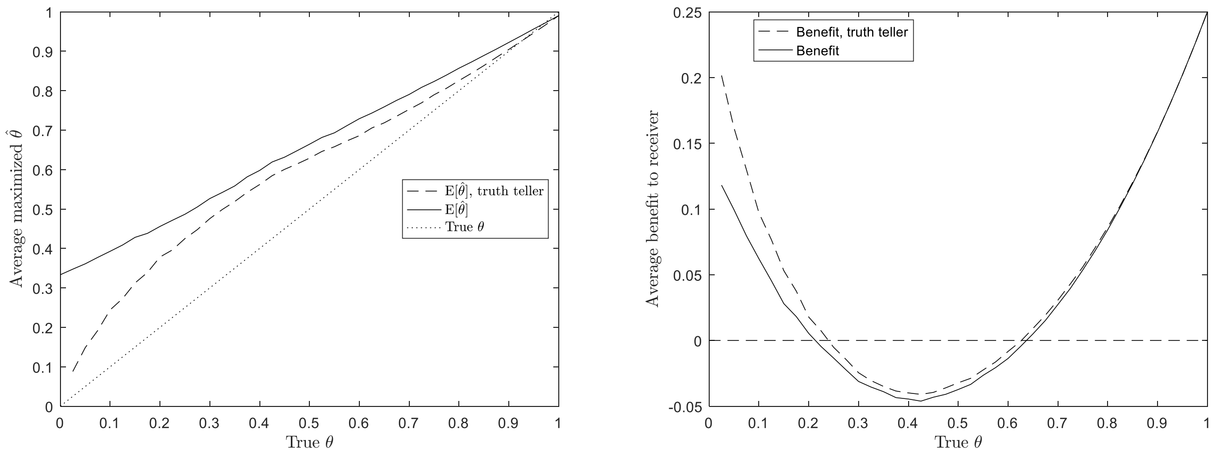


Figure 7: Simulated Impact of Persuasion on Beliefs and Welfare When The Receiver's Default Model is the Truth
This figure presents results on the impact of persuasion from simulations of the coin-flipping example for the case where $\psi \sim U[0,1]$ and $\underline{K} = 1$. For each of 40 values of 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.

receivers when θ is low, but has little benefit for intermediate or high values of θ.[32]

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.

$-(\hat{\theta}-\theta)^2 + (\mathbb{E}_\psi[\theta|h]-\theta)^2$. This would obviously be negative in expectation for large enough t since $\mathbb{E}_\psi[\theta|h]$ converges to θ.

[32]For 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.

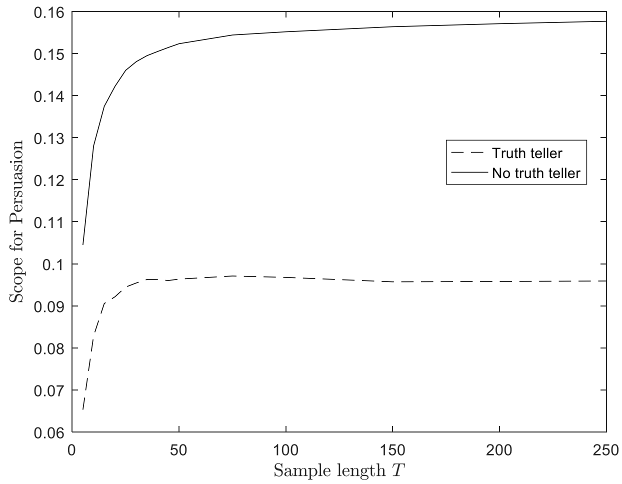


Figure 8: Simulated Impact of the Truthteller Constraint on the Scope for Persuasion
This figure presents results on the impact of persuasion from simulations of the coin-flipping example for the case where $\psi \sim U[0,1]$ and $\underline{K} = 1$. For each sample length, we compute the average over θ of the difference between the receiver's average post-persuasion beliefs and the econometrician's beliefs.

We next study how the impact of persuasion varies with sample size. Figure 8 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.[33] For instance, 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.

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

[33]On 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 ($\underline{K} = t - 3$). 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.

trouble. Similarly, the venture capitalist Marc Andreesen 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.

B.2 Highlighting Characteristics

Here, we show how to formalize the highlighting characteristics example in our framework. 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 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.

We write $c_m(x)$ to denote the element of the partition that contains x under model m. We assume the persuader can always propose the finest partition, where each x is in its own cell. To illustrate, if each x is described as a vector of attributes, $x = (x_1, x_2, \ldots, x_J)$, 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. Formally, let $k_m(x, h)$ denote the number of times an element in $c_m(x)$ appears in the history h, and $s_m(x, h)$ denote the number of times an element in $c_m(x)$ appears in h as a success rather than a failure. If $\psi = \text{Uniform}[0, 1]$, the probability the receiver attaches to the actor being successful given model m is $\hat{\theta}(x) \equiv \mathbb{E}[\theta(x)|m, h] = (s_m(x, h) + 1)/(k_m(x, h) + 2)$.

C When Are Receivers Persuadable? Details

Recall from Section 3 that 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.

We illustrated the fourth point in Section 3. To illustrate each of the first three 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. When receivers' defaults fit the data well, they are hard to persuade. Formally, 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'. 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: there is less scope for persuasion under histories that fit the default better. Under conditions we will make precise below, if persuasion is ineffective at history h given default d, then it is also ineffective at any history $\tilde{h}$ given default $\tilde{d}$ that induces the same beliefs and fits better: $\mu(\tilde{h},\tilde{d}) = \mu(h,d)$ and $\Pr(\tilde{h}|\tilde{d}) > \Pr(h|d)$. 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.

Proposition 6. *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: $\Pr(\tilde{h}|\tilde{d}, \mu_0) > \Pr(h|d, \mu_0)$.*

Proof. Proofs of all appendix propositions are collected in Appendix F. □

This proposition formalizes the two ways that increasing the fit of the receivers' default 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 bad gaffe by the candidate is meaningless if the voter ex ante believed the candidate to be competent than if the voter thought the candidate was incompetent.[34]

A 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.[35] 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.

Receivers are also less persuadable when the data is less open to interpretation. If persuasion is ineffective at history h and default d 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.

D Further Robustness, Extensions, and Examples

D.1 Model Averaging

In the main text, we assume that the receiver adopts the model he is exposed to that is most compelling given the data plus his prior. What if instead of "selecting" a model, he "averages" models

[34]As 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.

[35]In 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.

according to how compelling they are? This section characterizes the set of beliefs a single persuader is able to induce in this case, and compares this to the set of beliefs he is able to induce when the receiver is a model selector. One finding is that these sets are not nested: model averaging is sometimes more constraining to the persuader than model selecting, but in many situations is actually less constraining. Another finding is that key qualitative insights on when receivers are persuadable do not hinge on the assumption that receivers select rather than average models.

Suppose a receiver who as an *ex post model averager* exposed to models $\tilde{M}$ forms beliefs

$$\mu(h, \tilde{M})[\omega] = \sum_{m' \in \tilde{M}} \Pr(m'|h, \mu_0, \tilde{M}) \mu(h, m')[\omega].$$

Here, $\Pr(m|h, \mu_0, \tilde{M})$ is the receiver's "posterior" over models which 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 following result provides a simple characterization of beliefs a single persuader is able to induce when receivers are ex post model averagers.

Proposition 7. *Suppose the receiver is an ex post model averager who is maximally open to persuasion and fix d, μ_0, and h. For any $\mu' \in \Delta(\Omega)$, the persuader is able to induce any target belief*

$$a \cdot \mu' + (1 - a)\mu(h, d)$$

with $a \in [0, \mathit{Fit}(\mu'; h, \mu_0)/(\mathit{Fit}(\mu'; h, \mu_0) + \Pr(h|d))]$. The persuader is unable to induce any target belief of this form with $a > \mathit{Fit}(\mu'; h, \mu_0)/(\mathit{Fit}(\mu'; h, \mu_0) + \Pr(h|d))$.

This result just comes from, for every μ', deriving the range of posterior weights on m vs. d attainable when $\mu(h, m) = \mu'$. This, in turn, is a simple application of Bayes' Rule and the fact that for any $0 \leq p \leq \mathrm{Fit}(\mu'; h, \mu_0)$, the persuader is able to induce μ' with an $m(p)$ satisfying $\Pr(h|m(p), \mu_0) = p$. This result, together with a simple lemma, implies that the set of beliefs the persuader is able to induce is convex.

Lemma 2. *The function $\mathit{Fit}(\mu; h, \mu_0)$ is concave in μ.*

Proposition 8. *Suppose the receiver is an ex post model averager who is maximally open to persuasion and fix d, μ_0, and h. The set of beliefs the persuader is able to induce is convex. That is, if the persuader is able to induce belief $\mu^1 \in \Delta(\Omega)$ and belief $\mu^2 \in \Delta(\Omega)$, then he is also able to induce belief $\mu^3 = \alpha\mu^1 + (1 - \alpha)\mu^2$ for any $\alpha \in [0, 1]$.*

Armed with these results, we are able to supply more revealing characterizations of the set of beliefs the persuader is able to induce when receivers average models ex post.

Proposition 9. *Suppose the receiver is an ex post model averager who 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*

$$\tilde{\mu} \in \textit{Convex Hull}\left(\left\{\bar{\mu}^{\omega}, \underline{\mu}^{\omega}\right\}_{\omega\in\Omega}\right),$$

where

$$\bar{\mu}^{\omega}(\omega') = \begin{cases} \frac{\mu_0(\omega')[1+\pi_d(h|\omega')]}{\mu_0(\omega)+\Pr(h|d)} & \textit{if } \omega' = \omega \\ \frac{\pi_d(h|\omega')\mu_0(\omega')}{\mu_0(\omega)+\Pr(h|d)} & \textit{if } \omega' \neq \omega \end{cases}$$

and

$$\underline{\mu}^{\omega}(\omega') = \begin{cases} \frac{\pi_d(h|\omega')\mu_0(\omega')}{1-\mu_0(\omega)+\Pr(h|d)} & \textit{if } \omega' = \omega \\ \frac{[1+\pi_d(h|\omega')]\mu_0(\omega')}{1-\mu_0(\omega)+\Pr(h|d)} & \textit{if } \omega' \neq \omega. \end{cases}$$

The persuader is unable to induce any belief $\tilde{\mu} \in \Delta(\Omega)$ with $\tilde{\mu}(\omega) > \bar{\mu}^{\omega}(\omega)$ or $\tilde{\mu}(\omega) < \underline{\mu}^{\omega}(\omega)$ for any $\omega \in \Omega$.

To interpret this result, among beliefs that are implementable, $\bar{\mu}^{\omega}$ involves the largest possible belief in ω and $\underline{\mu}^{\omega}$ involves the lowest possible belief in ω. What this result says is that any convex combination of such beliefs is implementable by the persuader. In the case where there are only two states, this result reduces to a simple characterization of all implementable beliefs.

Corollary 5. *Assume $|\Omega| = 2$. Further suppose the receiver is an ex post model averager who 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*

$$\tilde{\mu}(\omega) \leq \frac{\mu_0(\omega) + \Pr(h|d)\mu(h,d)[\omega]}{\mu_0(\omega) + \Pr(h|d)} = \frac{\mu_0(\omega)[1 + \pi_d(h|\omega)]}{\mu_0(\omega) + \Pr(h|d)} \;\forall\, \omega \in \Omega.$$

Corollary 5 makes it easy to compare the set of beliefs that are implementable when receivers average models to the set of beliefs that are implementable when receivers select models (characterized in Proposition 1). To see this simply, let's stack the two conditions:

$$\text{Model Selection: } \tilde{\mu}(\omega) \leq \frac{\mu_0(\omega)}{\Pr(h|d)} \equiv \bar{\mu}^{\text{selection}}(\omega) \;\forall\, \omega \in \{\omega_1, \omega_2\}$$
$$\text{Model Averaging: } \tilde{\mu}(\omega) \leq \frac{\mu_0(\omega)[1 + \pi_d(h|\omega)]}{\mu_0(\omega) + \Pr(h|d)} \equiv \bar{\mu}^{\text{averaging}}(\omega) \;\forall\, \omega \in \{\omega_1, \omega_2\}.$$

Key comparative statics on when receivers are persuadable hold whether receivers select or average models. For example, under either assumption, receivers are more persuadable when they have difficulty explaining the data under their default interpretation or when there is a lot of open-to-interpretation data available: Both $\bar{\mu}^{\text{selection}}(\omega)$ and $\bar{\mu}^{\text{averaging}}(\omega)$ increase as $\Pr(h|d)$ goes down (all else equal) and tend to limiting values weakly above 1 as $\Pr(h|d) \to 0$.

Averaging rather than selecting models makes receivers more persuadable in some situations and less persuadable in others. In particular, when receivers are able to explain data well under their default interpretation then they are *more* persuadable when they average models: $\lim_{\Pr(h|d)\to 1} \bar{\mu}^{\text{selection}}(\omega) = \mu_0(\omega)$ while $\lim_{\Pr(h|d)\to 1} \bar{\mu}^{\text{averaging}}(\omega) = 2\mu_0(\omega)/(1+\mu_0(\omega))$. The idea is that when the default fits the data really well it leads to beliefs close to the receiver's prior and the persuader can only beat the default with a model that implies beliefs even closer to the prior. But, with model averaging, the persuader is able to propose a model that receivers will place non-trivial weight on even if it implies beliefs far from the prior. Conversely, when receivers have difficulty explaining data under their default interpretation then they are *less* persuadable when they average models: for $\Pr(h|d)$ sufficiently close to 0, $\bar{\mu}^{\text{selection}}(\omega) > 1$ and $\bar{\mu}^{\text{averaging}}(\omega) < 1$. When the data fits the default poorly, the persuader can easily beat the default even by proposing a model that implies beliefs far from the prior. With model averaging, receivers will continue placing non-trivial weight on the default. So there is not a nested relationship between the set of beliefs that are implementable when receivers average models compared to the set of beliefs that are implementable when receivers select models.

D.2 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.

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) $\omega \in \{0,1\}$, $U^R(a,\omega) = \text{Indicator}(a=\omega)$.[36]

When does the model that maximizes VFit meaningfully differ from the model that maximizes Fit? One class of examples involve situations where the receiver cares more about taking the correct action in some states than in others. For example, consider the example of 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 ($a(h,m') = 0$, $\mu(h,m')[1] < 1/101$). 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.

D.3 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

[36]When $|\tilde{M}| > 2$, 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.

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 10. *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*

$$\tilde{\mu}(\omega) < \frac{\mu_0(\omega)}{\Pr(h|d)} \forall \omega \in \Omega \text{ and } V^S(m(\tilde{\mu}), h) < V^S(d(h), h) \tag{15}$$

or

$$\tilde{\mu}(\omega) < \frac{\mu_0(\omega)}{\Pr(h|d)} \cdot (1-\sigma) \; \forall \omega \in \Omega \text{ and } V^S(m(\tilde{\mu}), h) \geq V^S(d(h), h), \tag{16}$$

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 (16) 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, but 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.

D.4 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

$$f\left(|H| \cdot \mu_0(1)\right), \tag{17}$$

which is implemented by an ambiguous data process π that features $\pi(h|\omega) = 1/|H|$ for all h, ω.[37] *So the persuader maximizes his payoff by collecting ambiguous data that is completely uninformative.* Noting that (17) 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 (17) 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.[38] 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 that this maximizes the wiggle room the persuader has to frame the information.

E Examples: Details

This appendix fleshes out arguments behind claims made in Section 7.

E.1 Persuading an Investor: Technical Analysis

This section shows how our framework predicts that the support resistance model from Figure 5b is more compelling than a random walk model.

At date t, the state is $\theta_t \in \{0.25, 0.5, 0.75\}$, where θ_t is the probability AMZN stock rises at date $t+1$. The persuader frames the history h of returns from 1/8/2019 to 1/28/2019 to influence

[37] This 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.

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

the receiver's posterior on $\theta_{1/29}$. Suppose the receiver's prior is evenly distributed across the three possible states: $\mu(\theta_{1/29} = 0.25) = \mu(\theta_{1/29} = 0.5) = \mu(\theta_{1/29} = 0.75) = 1/3$.

The default model is a history-dependent version of the random walk model—at each date, AMZN is equally likely to rise or fall:[39]

$$\pi^d(h|\theta) = \begin{cases} 0.5^{length(h)} & \text{if } \theta = 0.5 \\ 0 & \text{otherwise} \end{cases}.$$

The persuader proposes a support-resistance model. The model says that AMZN follows a random walk ($\theta_t = 0.5$) until it hits either the support or the resistance. If it hits the resistance, then it is likely to fall, i.e., $\theta_t = 0.25$ after hitting the resistance. If it hits the support, then it is likely to rise, i.e., i.e., $\theta_t = 0.75$ after hitting the support. The key flexibilities available to the persuader are in (i) picking the support and resistance levels after seeing the data and (ii) selecting the window of returns over which the model applies. Formally, let U^S and D^S be the number of up and down moves respectively after the support has been hit (and the resistance has not since been hit). Let U^R and D^R be the number of up and down moves respectively after the resistance has been hit (and the support has not since been hit). The model implies that the probability of the history is $\chi = (0.75)^{U^S+D^R}(0.25)^{U^R+D^S}(0.5)^{length(h)-(U^S+U^R+D^S+D^R)}$: up moves are likely after hitting the support and down moves are likely after hitting the resistance; conversely, up moves are unlikely after hitting the resistance and down moves are unlikely after hitting the support.

The model is formally:[40]

$$\pi^{RS}(h|\theta) = \begin{cases} \chi & \text{if last hit support and } \theta = 0.75 \\ \chi & \text{if last hit resistance and } \theta = 0.25 \\ \chi & \text{if never hit either and } \theta = 0.5 \\ 0 & \text{otherwise} \end{cases}.$$

Figure 9 shows a simple example of how the model applies to a stylized price path. The support is at 2 and the resistance is at 4. The price starts at 3, and neither the support nor the resistance has yet been hit. Thus, the probability of an up move is 50%. The price then rises to 4, hitting the resistance. Now the probability of an up move is 25%, and the probability of a down move is 75%. The next two price moves are down, and the support is hit. At this point, the probability of an up

[39]This model is history dependent because $\pi^d(h|\theta = .25) = \pi^d(h|\theta = .75) = 0$ for the particular h that is realized, though clearly $\sum_{\tilde{h}} \pi^d(\tilde{h}|\theta = .25) = 1 = \sum_{\tilde{h}} \pi^d(\tilde{h}|\theta = .75)$. The idea is that the investor as a default views the data as being diagnostic of a random walk. It would not change the conclusions of our analysis to instead specify the default as saying that returns data is uninformative about θ; i.e., $\pi^d(h|\theta = .25) = \pi^d(h|\theta = .5) = \pi^d(h|\theta = .75) = .5^{length(h)}$.

[40]Again, this model is history dependent and closed by specifying probabilities for un-realized histories that sum to 1.

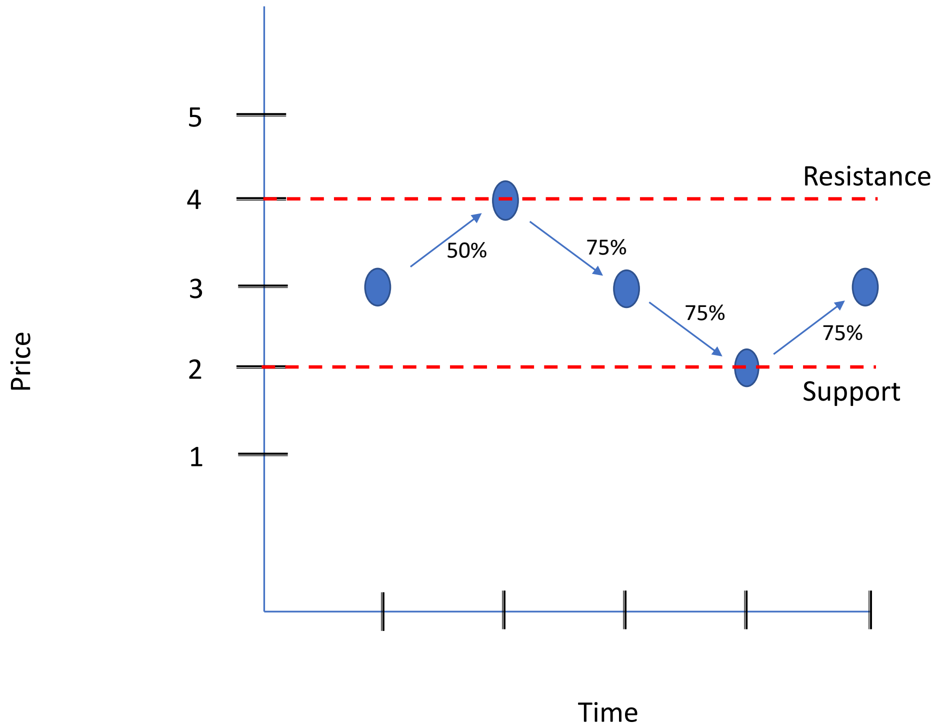


Figure 9: Applying the Support and Resistance Model to a Stylized Price Path

move is 75%. The last price move is up. Since the support was last hit, the probability of an up move next is 75%. Thus, according to the support-resistance model the probability of the history is

$$\Pr{}^{RS}(h) = \pi^{RS}(h|\theta = 0.75)\mu(\theta = 0.75) = (0.50)(0.75)^3(1/3) = 0.07.$$

Under the random walk model, the probability of the history is

$$\Pr{}^{d}(h) = \pi^{d}(h|\theta = 0.5)\mu(\theta = 0.5) = (0.50)^4(1/3) = 0.02.$$

Performing the analogous calculations on the actual AMZN price path from 1/8/2019 to 1/28/2019, the support-resistance model implies that AMZN is likely to rise, as AMZN had most recently hit its support. The probability of the history is more than four times higher under the support-resistance model than the random walk model. Note that this means that even if we modified the setup so that the prior strongly favored the random walk model (e.g., $\mu(\theta = 0.5) = 2 \times \mu(\theta = 0.75)$), the receiver would still find the support-resistance model more compelling.

E.2 Persuading a Client: Expert Advice in Individual Investing

This section provides a somewhat more elaborate formulation of individual investment advice, relative to the one presented in Section 3. Suppose there are N investments with investment j having characteristics $(x_1^j, \ldots, x_K^j)$. 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 $t \in \{1, \ldots, T\}$ 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 truthteller 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 truthteller. In this case, a simple application of Proposition 4 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.

E.3 Persuading a Client: Good to Great

This section shows how "Good to Great" advice from Collins (2001) is compelling according to our framework.

The setup is very similar to the entrepreneur problem. The underlying state is the expected (log) return for the portfolio of 11 good-to-great companies highlighted by Jim Collins. We assume the receiver's prior is that expected log returns are distributed $\mu \sim N(\bar{\mu}, \sigma_\mu^2)$. We observe realized returns, which are expected returns plus noise: $r_t = \mu + \varepsilon_{it}$ where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Let $\tau_\mu = 1/\sigma_\mu^2$ and $\tau_\varepsilon = 1/\sigma_\varepsilon^2$.

The default model is that μ is drawn once at the beginning of time. The posterior mean is $\frac{\tau_\mu}{\tau_\mu + T\tau_\varepsilon}\bar{\mu} + \frac{T\tau_\varepsilon}{\tau_\mu + T\tau_\varepsilon}\bar{r}$ where $\bar{r} = \frac{1}{T}\sum_t r_t$. Further let $s^2 = \frac{1}{T}\sum_t (r_t - \bar{r})^2$. How compelling is the default relative to alternatives? According to the default model, the likelihood of a return sequence $\mathbf{r}_t$ for a given μ is

$$\begin{aligned}
\Pr(\mathbf{r}_t|\mu) &= \prod_t \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{(r_{it}-\mu)^2}{2\sigma_\epsilon^2}\right\}. \\
&= \frac{1}{(2\pi)^{T/2}} \frac{1}{\sigma_\epsilon^T} \exp\left\{-\frac{T}{2\sigma_\epsilon^2}s^2\right\} \exp\left\{-\frac{T}{2\sigma_\epsilon^2}(\bar{r}-\mu)^2\right\}.
\end{aligned}$$

The prior for μ is given by

$$\Pr(\mu) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left\{-\frac{(\mu-\bar{\mu})^2}{2\sigma_\mu^2}\right\}.$$

Thus, the probability of return sequence $\mathbf{r}_t$ is

$$\Pr(\mathbf{r}_t) = \frac{1}{(2\pi)^{(T+1)/2}} \frac{1}{\sigma_\epsilon^T} \frac{1}{\sigma_\mu} \exp\left\{-\frac{T}{2\sigma_\epsilon^2}s^2\right\} \int \left[\exp\left\{-\frac{T}{2\sigma_\epsilon^2}(\bar{r}-\mu)^2 - \frac{(\mu-\bar{\mu})^2}{2\sigma_\mu^2}\right\}\right] d\mu.$$

This simplifies to (line 42 of https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf):

$$\Pr(\mathbf{r}_t) = \frac{\sigma_\varepsilon}{(\sqrt{2\pi}\sigma_\varepsilon)^T \sqrt{T\sigma_\mu^2 + \sigma_\varepsilon^2}} \exp\left(\frac{-\sum_t r_t^2}{2\sigma_\varepsilon^2} - \frac{\bar{\mu}^2}{2\sigma_\mu^2}\right) \exp\left(\frac{\frac{\sigma_\mu^2 T^2 \bar{r}^2}{\sigma_\varepsilon^2} + \frac{\sigma_\varepsilon^2 \bar{\mu}^2}{\sigma_\mu^2} + 2T\bar{r}\bar{\mu}}{2(T\sigma_\mu^2 + \sigma_\varepsilon^2)}\right). \quad (18)$$

The "this time is different" model is that μ is redrawn at the transition point selected by Jim Collins. Denote the transition point by L (for "leap"). Further, let $r_{[j,k]} \equiv \frac{1}{k-j+1}\sum_{t=j}^{k} r_t$. Then the likelihood of the data under the "this time is different model" is

$$\frac{\sigma_\varepsilon}{(\sqrt{2\pi}\sigma_\varepsilon)^L\sqrt{L\sigma_\mu^2+\sigma_\varepsilon^2}}\exp\left(\frac{-\sum_{t=1}^L r_t^2}{2\sigma_\varepsilon^2}-\frac{\bar{\mu}^2}{2\sigma_\mu^2}\right)\exp\left(\frac{\frac{\sigma_\mu^2 L^2\bar{r}_{[1,L]}^2}{\sigma_\varepsilon^2}+\frac{\sigma_\varepsilon^2\bar{\mu}^2}{\sigma_\mu^2}+2L\bar{r}_{[1,L]}\bar{\mu}}{2(L\sigma_\mu^2+\sigma_\varepsilon^2)}\right)\times$$

$$\frac{\sigma_\varepsilon}{(\sqrt{2\pi}\sigma_\varepsilon)^{T-L}\sqrt{(T-L)\sigma_\mu^2+\sigma_\varepsilon^2}}\exp\left(\frac{-\sum_{t=L+1}^T r_t^2}{2\sigma_\varepsilon^2}-\frac{\bar{\mu}^2}{2\sigma_\mu^2}\right)\times$$

$$\exp\left(\frac{\frac{\sigma_\mu^2(T-L)^2\bar{r}_{[L+1,T]}^2}{\sigma_\varepsilon^2}+\frac{\sigma_\varepsilon^2\bar{\mu}^2}{\sigma_\mu^2}+2(T-L)\bar{r}_{[L+1,T]}\bar{\mu}}{2((T-L)\sigma_\mu^2+\sigma_\varepsilon^2)}\right).$$

Applying these formulas to the actual annual return path of 11 firms selected by Collins, we find that the "this time is different model" is 8 times more likely to explain the data than the default model.[41] If we extend the data to the present day (an additional 20 years of data), then the "this time is different model" is 25% less likely to explain the data than the default model.

A naive regression approach gives similar results, presented in Table 1. The constant says the firms average (log) returns of 6.2% per year before their leap and 6.2%+16.6% = 22.8% after the leap. The difference between pre- and post-leap is enormously statistically significant. If we extend the sample to the present day (an additional 20 years of data), after their leap firms average returns of 6.2%+7.0% = 13.2%, and the difference between pre- and post-leap is only barely statistically significant at the 10%. The last regression shows that if we split the post-leap period into the part Collins covered (post1) and the following years (post2), firms did worse in the years Collins did not cover than they did in the pre-leap period, but the difference is not statistically significant. It is as if the great firms just went back to being average. In untabulated results, we find similar patterns when we examine earnings, as opposed to stock returns.

F Appendix Proofs

Proof of Proposition 6. 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 $V^S(m,h) > V^S(d,h)$.

[41] These calculations assume $\bar{\mu} = 15\%$ per year (the average return across all stocks in CRSP over this period), $\sigma_\mu = 6\%$ (the standard deviation of expected returns under the assumption the CAPM holds), and $\sigma_\varepsilon = 17\%$ (the in-sample standard deviation of the portfolio of 11 stocks). Our results are robust to perturbing these numbers somewhat.

Table 1: Stock Return Performance of Good-to-Great Firms

Dependent Variable	Average Log Return$_t$		
	Original Sample	Extended Sample	
$\text{Post}_{[0,15]}$	16.58**		16.58**
	(3.26)		(3.25)
$\text{Post}_{[0,35]}$		7.00*	
		(3.96)	
$\text{Post}_{[16,35]}$			-1.06
			(-5.09)
Constant	6.21**	6.21**	6.21**
	(2.77)	(2.74)	(2.76)
Adj. R^2	0.46	0.02	0.24
N	31	50	50

This table regresses the average log stock return of the 11 good-to-great firms in event time on dummy variables selecting different periods. The variable $\text{Post}_{[x,y]}$ is equal to one for years x to y in event time and zero otherwise. Data is annual. In column (1), the sample is the original sample in the book. In columns (2) and (3) the sample is extended through 2018. Robust standard errors are reported.

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 $\tilde{h}$, $\tilde{d}$, μ_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 7. Fix a $\mu' \in \Delta(\Omega)$ and recall that $\text{Fit}(\mu'; h, \mu_0) = \max_m \Pr(h|m, \mu_0)$ such that $\mu(h,m) = \mu'$. For any $0 \leq p \leq \text{Fit}(\mu'; h, \mu_0)$, the persuader is able to induce μ' with an $m(p)$ satisfying $\Pr(h|m(p), \mu_0) = p$. To see this, let $\pi_{m(p)}(h|\omega) = (\mu'(\omega)/\mu_0(\omega)) \cdot p \ \forall\ \omega$. So for any $0 \leq p \leq \text{Fit}(\mu'; h, \mu_0)$, the persuader is able to induce any

$$\mu(h, m(p)) \cdot \Pr(m(p)|h, \mu_0) + \mu(h,d) \cdot \Pr(d|h, \mu_0) =$$
$$\mu' \cdot \left(\frac{p}{p + \Pr(h|d)}\right) + \mu(h,d) \cdot \left(1 - \frac{p}{p + \Pr(h|d)}\right).$$

Since the persuader is also *not* able to induce $\mu(h,m) = \mu'$ with $\Pr(h|m) > \text{Fit}(\mu'; h, \mu_0)$, the result follows.

□

Proof of Lemma 2. We know from Lemma 1 that $\text{Fit}(\mu; h, \mu_0) = 1/\text{Movement}(\mu; \mu_0)$. $\text{Movement}(\mu; \mu_0) =$

$\max_{\omega\in\Omega} \mu(\omega)/\mu_0(\omega)$ is convex in μ: for any μ', $\mu'' \in \Delta(\Omega)$ and $\alpha \in [0,1]$,

$$\max_{\omega\in\Omega}[\alpha\mu'(\omega) + (1-\alpha)\mu''(\omega)]/\mu_0(\omega) \leq \alpha \max_{\omega\in\Omega} \mu'(\omega)/\mu_0(\omega) + (1-\alpha) \max_{\omega\in\Omega} \mu''(\omega)/\mu_0(\omega).$$

As a result, $\text{Fit}(\mu; h, \mu_0)$ is concave in μ.

□

Proof of Proposition 8. Write

$$\begin{aligned}
\mu^1 &= a_1\mu_1 + (1-a_1)\mu(h,d) \\
\mu^2 &= a_2\mu_2 + (1-a_2)\mu(h,d) \\
\mu^3 &= \alpha\mu^1 + (1-\alpha)\mu^2
\end{aligned}$$

with a_1, a_2, $\alpha \in [0,1]$. We can re-write

$$\mu^3 = a_3[\tilde{\alpha}\mu_1 + (1-\tilde{\alpha})\mu_2] + (1-a_3)\mu(h,d),$$

where

$$\begin{aligned}
\tilde{\alpha} &= \frac{\alpha a_1}{\alpha a_1 + (1-\alpha)a_2} \\
a_3 &= \alpha a_1 + (1-\alpha)a_2.
\end{aligned}$$

By Proposition 7, we know that μ^3 is implementable if

$$a_3 \leq \frac{\text{Fit}(\mu^3; h, \mu_0)}{\text{Fit}(\mu^3; h, \mu_0) + \Pr(h|d)}. \tag{19}$$

To establish this inequality, note that μ^1 and μ^2 being implementable implies, respectively, that $a_1 \leq \text{Fit}(\mu^1; h, \mu_0)/(\text{Fit}(\mu^1; h, \mu_0) + \Pr(h|d))$ and $a_2 \leq \text{Fit}(\mu^2; h, \mu_0)/(\text{Fit}(\mu^2; h, \mu_0) + \Pr(h|d))$ by applications of Proposition 7. This further implies that

$$\begin{aligned}
a_3 = \alpha a_1 + (1-\alpha)a_2 &\leq \alpha\frac{\text{Fit}(\mu^1; h, \mu_0)}{\text{Fit}(\mu^1; h, \mu_0) + \Pr(h|d)} + (1-\alpha)\frac{\text{Fit}(\mu^2; h, \mu_0)}{\text{Fit}(\mu^2; h, \mu_0) + \Pr(h|d)} \\
&\leq \frac{\text{Fit}(\mu^3; h, \mu_0)}{\text{Fit}(\mu^3; h, \mu_0) + \Pr(h|d)}.
\end{aligned}$$

The last inequality follows from $\frac{\text{Fit}(\mu;h,\mu_0)}{\text{Fit}(\mu;h,\mu_0)+\Pr(h|d)}$ being concave in μ by virtue of $\text{Fit}(\mu; h, \mu_0)$ being concave in μ (Lemma 2) and $x/(x+\Pr(h|d))$ being increasing and concave in x. So this establishes inequality (19) and the result follows.

□

Proof of Proposition 9. Note that, among beliefs that are implementable, $\bar{\mu}^\omega$ involves the largest possible belief in ω and $\underline{\mu}^\omega$ involves the lowest possible belief in ω. Belief $\bar{\mu}^\omega$ is implemented with a model m that yields $\mu(h,m)[\omega] = 1$ with maximal possible fit $\mu_0(\omega)$. Belief $\underline{\mu}^\omega$ is implemented with a model m that yields $\mu(h,m)[\omega] = 0$ and $\mu(h,m)[\omega'] = \mu_0(\omega')/(1-\mu_0(\omega))$ $\forall$ $\omega' \neq \omega$ with maximal possible fit $(1-\mu_0(\omega))$.

Any belief in Convex Hull $\left(\{\bar{\mu}^{\omega}, \underline{\mu}^{\omega}\}_{\omega\in\Omega}\right)$ is implementable since the set of implementable beliefs is convex (Lemma 8).

By construction, the persuader is unable to induce any belief $\tilde{\mu} \in \Delta(\Omega)$ with $\tilde{\mu}(\omega) > \bar{\mu}^{\omega}(\omega)$ or $\tilde{\mu}(\omega) < \underline{\mu}^{\omega}(\omega)$ for any $\omega \in \Omega$.

□

Proof of Corollary 5. Specializing to $|\Omega| = 2$, this follows immediately from Proposition 9: the upper bound is just $\bar{\mu}^{\omega}(\omega)$ and the lower bound is redundant with binary states.

□

Proof of Proposition 10. Follows the logic of the proof of Proposition 1.

□

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 Hjalmarsson, "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 Sentiment," *Journal of financial economics*, 1998, *49* (3), 307–343.

Barberis, Nick, Campbell Harvey, and Neil Shephard, "Overfitting and Its Impact on the Investor," *Man AHL Academic Advisory Board*, 2015.

Barron, Daniel and Michael Powell, "Markets for Rhetorical Services," 2018.

Becker, Gary and Kevin Murphy, "A Simple Theory of Advertising as a Good or Bad," *The Quarterly Journal of Economics*, 1993, *108* (4), 941–964.

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.

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.

Collins, Jim, *Good to Great: Why Some Companies Make the Leap... and Others Don't*, Random House, 2001.

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.

DeMarzo, Peter, Dimitri Vayanos, and Jeffrey Zwiebel, "Persuasion Bias, Social Influence, and Unidimensional Opinions," 2003, *118* (3), 909–968.

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 Spiegler, "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.

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.

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.

Harvey, Campbell, "Presidential Address: The Scientific Outlook in Financial Economics," *The Journal of Finance*, 2017, *72* (4), 1399–1440.

Hayakawa, Samuel I, *Language in Thought and Action*, Harcourt, 1940.

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.

Levitt, Steven D, "From Good to Great... to Below Average," *Freakonomics Blog. Retrieved*, 2008.

Levy, Gilat and Ronny Razin, "An Explanation-Based Approach to Combining Forecasts," Technical Report, mimeo 2019.

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.

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.

Niendorf, Bruce and Kristine Beck, "Good to Great, or Just Good?," *The Academy of Management Perspectives*, 2008, pp. 13–20.

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.

Rosenzweig, Phil, *The Halo Effect:... and the Eight Other Business Delusions That Deceive Managers*, Simon and Schuster, 2007.

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.

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.