Using Models to Persuade

Joshua Schwartzstein
Harvard Business School

Adi Sunderam*
Harvard Business School

August 31, 2020

Abstract

We present a framework where "model persuaders" influence receivers' beliefs by proposing models that organize past data to make predictions. Receivers are assumed to find models more compelling when they better explain the data, fixing receivers' prior beliefs. Model persuaders face a tradeoff: better-fitting models induce less movement in receivers' beliefs. Consequently, a receiver exposed to the true model can be most misled by persuasion when that model fits poorly, competition between persuaders tends to neutralize the data by pushing towards better-fitting models, and a persuader facing multiple receivers is more effective when he can send tailored, private messages.

Persuasion often 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 high because of the schools," they are supplying a model: home buyers should pay attention to local schools, which are an important determinant of house prices. 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.

---

*E-mails: jschwartzstein@hbs.edu and asunderam@hbs.edu. We thank Ned Augenblick, Francesca Bastianello, Max Bazerman, Roland Benabou, Pedro Bordalo, Stefano DellaVigna, Erik Eyster, Xavier Gabaix, Tristan Gagnon-Bartsch, Nicola Gennaioli, Matt Gentzkow, Robert Gibbons, Russell Golman, Brian Hall, Sam Hanson, Botond Koszegi, George Loewenstein, Deepak Malhotra, Ulrike Malmendier, Kathleen McGinn, Sendhil Mullainathan, Ramana Nanda, Matthew Rabin, Jesse Shapiro, Andrei Shleifer, Kelly Shue, Erik Stafford, Jeremy Stein, Dmitry Taubinsky, Tim Vogelsang, three anonymous referees, and seminar participants at UC Berkeley, MIT, Harvard, Princeton, CMU, BEAM 2019, and EBEM 2019 for helpful comments.

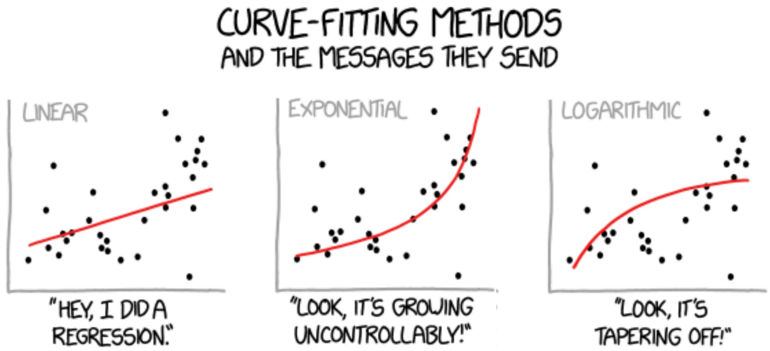


Figure 1: Stylized Example of Model Persuasion from xkcd.com
Source: https://xkcd.com/2048/

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.

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:

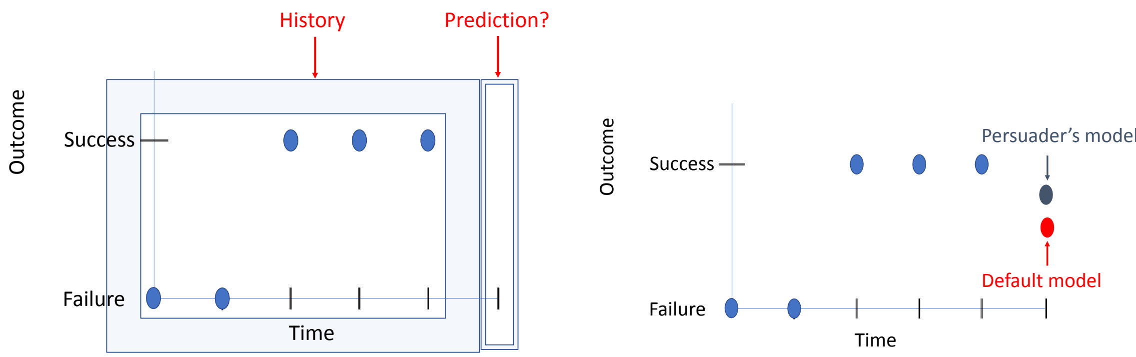
$$\Pr(h|m,\mu_0) = \int \pi_m(h|\omega)d\mu_0(\omega) \geq \int \pi_{m'}(h|\omega)d\mu_0(\omega) = \Pr(h|m',\mu_0).$$

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]

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

[2]While our emphasis on models that satisfy fit constraints is (to our knowledge) novel in the context of persuasion, there is work in decision theory that uses similar criteria in other settings. Most closely related, Levy and Razin (2020) contemporaneously analyzes how people combine expert forecasts, assuming they favor explanations that maximize the likelihood of the data. There is also work that draws out implications of related assumptions, including Epstein and Schneider (2007) which 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 errors.



(a) Setup

(b) Persuader vs. Default Model

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

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%.[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.[4]

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

[3]Section I presents this example formally. The probability of observing the data under the true model is $\int_0^1 (1-\theta)^2\theta^3 d\theta = .017$, while the probability under the model the persuader proposes is $\int_0^1 (1-\theta)^2 d\theta \times \int_0^1 \theta^3 d\theta = .083$. Below we establish what the receiver can be persuaded of if she is willing to entertain a broader set of models.

[4]A Bayes Factor comparing two models M_1 and M_2 is the ratio of the likelihood of the data under M_1 to its likelihood under M_2. While a traditional likelihood ratio test fixes model parameters, a Bayes Factor integrates over them. See, e.g., Kass and Raftery (1995) for a fuller discussion.

may discard certain stories because they "do not hang together"—in our framework, receivers may be unwilling to consider some models. And they may interpret data through the lens of a default model. But, crucially, receivers do not generate new stories themselves. They need experts to supply them.[5] Second, because receivers need persuaders to supply models, they do not have a prior over models. Instead, a receiver judges models only by how well they fit 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. However, as we discuss below, 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 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.[6]

Section I sets up our general framework. We make the basic observation that model persuasion may make receivers worse off on average than they would be if they interpreted data through the lens of their default model, e.g., if their default is accurate to begin with. The idea that persuasion can be harmful to receivers on average 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 II 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 imply 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

[5] 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.

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

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.[7] 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 on other data that is open to interpretation is limited.

Section III 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. Model persuasion is particularly effective 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 imply the data are consistent with their prior more compelling than a true model that implies the data contradicts their prior.

Section IV 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 strategy 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

[7] 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).

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 are more compelling when they lead to conclusions that receivers are predisposed to believe.

Section V 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 VI 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.

Finally, Section VII briefly considers three types of extensions before we conclude in Section VIII. First, we consider a simple dynamic problem, in which persuaders can propose new models after new data is revealed. We show that while dynamic considerations sometimes constrain persuaders, they still benefit from persuasion in the environments we consider. In some situations, dynamic considerations in fact free persuaders, making them better off than in the static case. Second, we relax our receiver naivete assumptions, allowing receivers to be more skeptical of persuaders with misaligned incentives. Our main results remain qualitatively unchanged. Third, we consider alternative assumptions about receivers' prior knowledge. We demonstrate how our assumptions that receivers have prior knowledge that the persuader cannot directly influence are crucial for deriving the fit versus movement tradeoff underpinning our analysis. But we also show that natural further refinements of receivers' knowledge, e.g., capturing knowledge about the distribution over observables, do not per se further constrain persuaders. Taken together, these extensions show that the core intuitions that emerge from our baseline setup are likely to persist in more complex environments, and they suggest avenues for future work.

Related Literature

Our paper is related to several strands of the economics and psychology literatures. Many of the logical stories, narratives, analogies, and metaphors people use are models to make sense of the data (e.g., Lakoff and Johnson 1980; Bruner 1991; Chong and Druckman 2007; Shiller 2017). While people engage in such sensemaking even absent persuasion, persuasion impacts their interpretations (e.g., Andreassen 1990; DiFonzo and Bordia 1997). 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), where one of the ways that persuasion works is through providing advantageous "frames" of known aspects of a product.[8] 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.[9] While we study what makes messages persuasive, these papers assume 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 differently from persuaders in KG and generalizations of their framework such as Alonso and Camara (2016), Galperti (2019), and Ely (2017). KG's persuaders influence by providing information, fixing the models receivers use to interpret information; ours influence by providing models, fixing the information receivers have. 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.

[8] 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.

[9] 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. Olea et al. (2019) study a related question: if agents observe the same data and use heterogeneous models in a prediction problem, who is the most confident in their predictive ability? Glazer and Rubinstein (2004, 2006) conceptualize persuasion and debate as a mechanism design problem, and consider optimal persuasion rules from the point of view of the receiver.

I Model Persuasion

I.A 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 $U^R(a, \omega)$.[10] 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.[11]

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

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

[11] In many 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.

posterior distribution over Ω given h and model $\tilde{m} \in M \cup \{d\}$, as derived by Bayes' rule.[12] We assume the receiver adopts the persuader's model m and hence posterior $\mu(h, m)$ if

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

and adopts the default model and hence posterior $\mu(h, d)$ if the inequality is reversed. To ensure the existence of persuader-optimal models, we assume that in the case of a tie the receiver goes with the persuader's 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.

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.[13] 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 III, 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.

Model persuasion has two effects, spelled out in Appendix B. First, it potentially enables re-

[12] That is, for all $\omega \in \Omega$, $h \in H$, and $\tilde{m} \in M \cup \{d\}$,

$$\mu(\omega|h, \tilde{m}) = \frac{\pi_{\tilde{m}}(h|\omega) \cdot \mu_0(\omega)}{\int \pi_{\tilde{m}}(h|\omega')d\mu_0(\omega')}.$$

[13] 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.

ceivers to act on more information, e.g., when the receiver uses an uninformative default. Second, it frames information, which can make the receiver worse off on average. For example, if the receiver correctly interprets data in the absence of persuasion ($d = m^T$), the receiver is led astray on average by being persuaded. However, as illustrated in Appendix C.1, 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.

I.B 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 entrepreneur 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:

> 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 rapid growth was popularized in part by financial analysts with incentives that rewarded high stock prices (Shiller 2015).

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 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 probability S/K.[14] For example, if $\psi = \text{Uniform}[0,1]$ and m is the K-model, then application of the beta-binomial Bayesian updating formula implies that 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 C analyzes this example in detail.

Example 2 (Highlighting characteristics)**.** While some applications fit within the highlighting strips framework, many other real world examples involve highlighting characteristics. For instance, consider 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.'"[15] 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." 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 C, we show how to formalize such examples in our framework.

I.C Additional Discussion and Interpretation 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. Politicians find it advantageous to stump for measures that are not beneficial to their constituents.

Second, while the receiver does not know how to interpret data, he does have prior knowledge.[16] Even a casual investor may understand that a mutual fund cannot be expected to outperform the market 100% of the time. Similarly, a voter may understand that a third-party candidate is unlikely to win a presidential election no matter the interpretation of polls the candidate proposes. More broadly, the prior captures any knowledge of the receiver that the persuader cannot influence with-

[14] 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

$$\Pr(h|K \text{ model},\mu_0) = \left(\int_0^1 \theta^S(1-\theta)^{K-S}\psi(\theta)d\theta\right) \cdot \left(\int_0^1 \theta^{l-S}(1-\theta)^{t-l-K+S}\psi(\theta)d\theta\right).$$

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

[16] Appendix E.1.1 illustrates the importance of prior knowledge. Under an alternative formulation where the persuader is able to propose any prior over states and any likelihood function, the persuader can implement any beliefs she wants, so long as the receiver is willing to entertain any such model. The assumption that the receiver has prior knowledge underlies our results below that persuaders face fundamental tradeoffs in proposing models.

out 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 VI.

Third, in some applications it is more natural to think of the receiver's utility as being over actions and outcomes Y, rather than over actions and latent states of the world.[17] In such cases, letting $f(y|\omega)$ capture the distribution over outcomes given states and for simplicity taking the outcome space as finite, we think of $U^R(a, \omega)$ as being the reduced form for underlying expected utility $\sum_{y' \in Y} \tilde{U}^R(a, y') f(y'|\omega)$. This makes two features of our setup more explicit:

1. The receiver's prior knowledge includes both beliefs over the likelihood of different states (captured by $\mu_0(\omega)$) and also how these states translate into payoff-relevant outcomes (captured by $f(y|\omega)$). Just like the persuader is unable to directly change μ_0, she is unable to directly influence f.

2. There may be inherent uncertainty that the persuader is unable to resolve. Even if the receiver can be persuaded to believe with certainty in some state $\bar{\omega}$, he is still uncertain about the outcome whenever $f(\cdot|\bar{\omega})$ places positive weight on more than one outcome.

For instance, take the entrepreneur example from the introduction. The receiver ultimately cares about a binary outcome: whether or not the next startup will be successful. However, persuasion changes the receiver's beliefs about the entrepreneur's quality. The receiver knows the mapping between quality and the likelihood of success (quality θ corresponds to success probability θ), which the persuader cannot alter. This knowledge, combined with the receiver's prior that quality is uniformly distributed over $[0, 1]$, means that the receiver knows it is impossible to guarantee the entrepreneur's next startup is a success, no matter her quality. Intuitively, though the receiver is open to models that help make sense of past data, he recognizes that there is always some uncertainty involved when it comes to the success of a start-up.[18]

Fourth, the receiver does not take the persuader's incentives into account when reacting to proposed models. We make this naivete assumption—a common building block in the behavioral-IO literature (e.g., Heidhues and Koszegi, 2018; Eyster, 2019)—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.

[17] We are abusing terminology to use the term "states of the world" interchangeably with the parameters that index distributions over observables.

[18] There may be situations where the receiver does not have such prior knowledge, which could be captured by taking the state space to be the same as the outcome space, with $f(y = \tilde{y}|\omega = \tilde{y}) = 1$ for all $\tilde{y} \in Y = \Omega$. In this case, the persuader might be able to convince the receiver in the inevitability of an outcome.

That said, receivers are unlikely to take everything persuaders say at face value. We explore forms of receiver sophistication in Section VII and Appendix E.

Fifth, as we discuss a bit more in Section VII, 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.

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.[19] 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.[20] 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.

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.

[19] Receivers may have an intuitive sense that some models are less plausible than others, regardless of how well they fit receivers' broader knowledge. One could extend our framework by allowing receivers to assign prior weights to models after they are proposed. For example, receivers could have a "background prior" $\nu(m)$ over models and after hearing model m assign it prior weight $\nu(m)/\sum_{m' \in \text{Models Exposed To}} \nu(m')$ ("Models Exposed To" includes the default and models proposed by persuaders). With this formulation, receivers would not consider models they are not exposed to, but they would view some models they are exposed to more skeptically than others. For example, $\nu(\cdot)$ could penalize "more complex" models.

[20] 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: $\Pr(\text{2 failures}|\text{2 parameter model}) = (\int_0^1 (1-\theta)d\theta)^2 = 1/4 < 1/3 = \Pr(\text{2 failures}|\text{1 parameter model}) = \int_0^1 (1-\theta)^2 d\theta$. 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.

II Scope and Limits of Model Persuasion

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 net of fees 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,[21] 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. Finally, assume that Pat's default model is that past returns are somewhat informative. He believes that good funds have a higher probability of high past returns than bad funds: $\pi_d(\text{high returns}|\text{good}) = \pi_d(\text{low returns}|\text{bad}) = 75\%$.

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:

$$\Pr(\text{high returns}|d, \mu_0) = \pi_d(\text{high returns}|\text{good}) \times 20\% + \pi_d(\text{high returns}|\text{bad}) \times 80\% = 35\%.$$

Under his default, Pat will not invest in the fund because

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

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. 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

[21] 24% of all actively managed have outperformed their passive benchmarks net of fees over the last 10 years (Morningstar Active/Passive Barometer, 2019).

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:

$$\Pr(\text{high returns}|m, \mu_0) = \pi_m(\text{high returns}|\text{good}) \times 20\% + \pi_m(\text{high returns}|\text{bad}) \times 80\% = 20\%.$$

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

$$\Pr(\text{high returns}|m, \mu_0) = \pi_m(\text{high returns}|\text{good}) \times 20\% + \pi_m(\text{high returns}|\text{bad}) \times 80\% \geq 35\%.$$

If she proposes a model with $\pi_m(\text{high returns}|\text{good}) = 100\%$, this means she must set $\pi_m(\text{high returns}|\text{bad})$ above 18.75%. Under the most favorable such model to the broker, $\Pr(\text{good}|\text{high returns}) = 57\%$, and Pat will invest in the active fund. 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.[22] 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.

[22] Formally, suppose Pat's default is given by: $\pi_d(\text{very high returns}|\text{good}) = 15\%$, $\pi_d(\text{high returns}|\text{good}) = 60\%$, $\pi_d(\text{low returns}|\text{good}) = 25\%$, $\pi_d(\text{very high returns}|\text{bad}) = 5\%$, $\pi_d(\text{high returns}|\text{bad}) = 20\%$, $\pi_d(\text{low returns}|\text{bad}) = 75\%$. The alternative model that good funds always reveal themselves with very high 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: $\Pr(\text{very high returns}|m, \mu_0) = 20\% > \Pr(\text{very high returns}|d, \mu_0) = 7\%$.

II.A What Can Receivers Be Persuaded Of?

The above example shows 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 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 much. 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 C—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*

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

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.

II.B 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, as formalized in Appendix D, there are at least 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 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.

The fourth point is somewhat 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 $\pi_m(h|\omega) = \pi_{m^T}(h_1|\omega)\pi_m(h_2|\omega)$. 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 in $\tilde{\Omega} \subset \Omega$ (i.e., $\pi_{m^T}(h_1|\omega) > 0 \;\forall\; \omega \in \tilde{\Omega}$), then with "enough" ambiguous data the persuader is able to induce any target belief with support on $\tilde{\Omega}$, provided the model space M is sufficiently rich. For instance, for a US presidential candidate, long track records are both a blessing and a curse, providing a lot of ambiguous data 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 cannot be President.

To formalize this intuition, define "more data" as follows. Consider sequences of histories $(h_1^i)_{i=1,\dots,\infty}$ and $(h_2^i)_{i=1,\dots,\infty}$, with higher i 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 $\tilde{\Omega} \subset \Omega$ and any target belief $\tilde{\mu} \in \Delta(\tilde{\Omega})$, then there exists a $\bar{j}$ such that for any h_2^j with $j \geq \bar{j}$ the persuader is able to propose a model that induces $\tilde{\mu}$.*

This proposition provides a sense in which more unambiguous data constrains the persuader, while more ambiguous data liberates the persuader. This finding contrasts with an intuition from "information persuasion" that more data if anything limits the scope for persuasion.

III 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 m be more compelling than the default (Eq. (1)) becomes $\Pr(h|m, \mu_0) \geq \Pr(h|m^T, \mu_0)$.[23] The persuader takes

[23] 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.

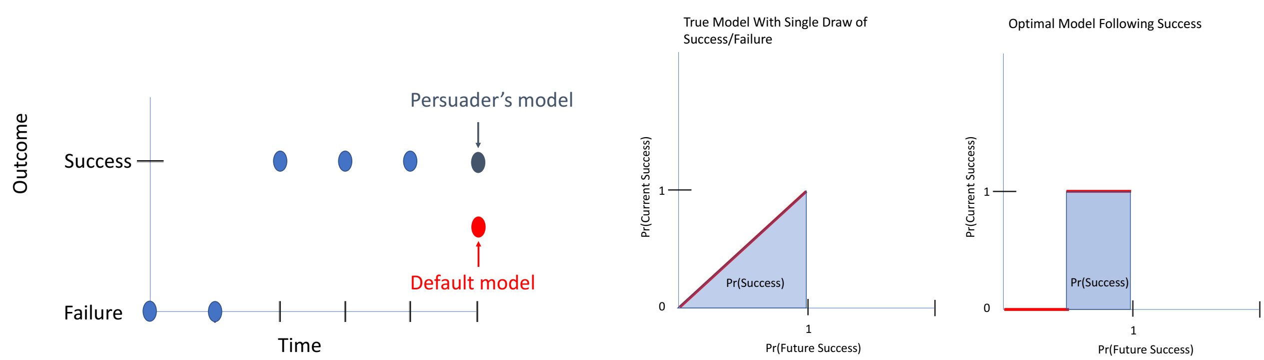
this constraint, which we refer to as the truthteller constraint, into account in choosing which model to propose. When the true model is the default model, a trivial corollary of Proposition 1 (just plugging in m^T for d) characterizes the beliefs the persuader is able to induce.[24]

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*

$$\mu(\omega|h, m) \leq \frac{\mu_0(\omega)}{\Pr(h|m^T, \mu_0)} \; \forall \omega \in \Omega \tag{3}$$

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.



(a) Predicting success

(b) Optimal model when maximally open

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

What do models look like when the receiver is maximally open 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

[24] In the Bayesian Network approach of Eliaz and Spiegler (2019)—which specifies an equilibrium based on infinite data—the true model cannot be beaten. As a result, that approach cannot address many of the points we investigate.

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 always reveal themselves by being successful early".[25] 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 true model is the default, the comparative statics in Section II on when receivers are persuadable become statements about the true process. For instance, the less likely the history under the true model, the more space there is for misleading persuasion. Events that are truly surprising are ripe to be framed because the truthteller constraint is relatively weak for such events. Thus, we should see a lot of persuasive activity creating narratives surrounding "tail events", such as particularly good or bad performance by a company. Similarly, when the true model is the default, 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 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 VI.

[25] To see that such a model is optimal, suppose the persuader instead proposes a model m that satisfies $\Pr(h|m,\mu_0) \geq \Pr(h|m^T,\mu_0)$, but does not follow such a cutoff rule over a positive measure of θ. Consider instead model m' with $\pi_{m'}(h|\theta) = 1$ for all $\theta \geq 1 - \Pr(h|m,\mu_0)$ and $\pi_{m'}(h|\theta) = 0$ for all $\theta < 1 - \Pr(h|m,\mu_0)$. Model m' satisfies $\Pr(h|m',\mu_0) = \Pr(h|m,\mu_0)$ and

$$\mathbb{E}[\theta|h,m'] = \int_{1-\Pr(h|m,\mu_0)}^{1} \frac{\theta}{\Pr(h|m,\mu_0)} d\theta > \int_0^1 \theta \frac{\pi_m(h|\theta)}{\Pr(h|m,\mu_0)} d\theta = \mathbb{E}[\theta|h,m],$$

where the inequality follows from simple calculations. This contradicts m being an optimal model.

Taken together, these results suggest that a truthteller does not constrain the persuader much. Appendix C.1 analyzes the highlighting strips example from Section I.B using simulations to better understand the intuitions and magnitudes involved when receivers are open only to a limited set of models. We show how our qualitative results derived assuming that receivers are maximally open to persuasion carry over to this case.

IV Competition Between Persuaders

In the previous section we considered a single persuader competing with the truth. This section considers competition between persuaders more broadly. To incorporate such competition, 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

$$\Pr(h|m, \mu_0) > \Pr(h|m', \mu_0) \tag{4}$$

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 $\Pr(h|\cdot, \mu_0)$. In the case of ties, 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 observation: Fix history h and suppose there are at least two persuaders. If $\tilde{m}$ solves

$$\max_{m \in M \cup \{d(h)\}} \Pr(h|m, \mu_0),$$

then there is an equilibrium where the receiver holds beliefs $\mu(h, \tilde{m})$. If a model maximizes the probability of the data, then there is an equilibrium where receivers interpret the data through the lens of that model, though it may not be the only equilibrium, or the most natural. This observation 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 3. *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$(\mu; h, \mu_0) \geq \Pr(h|d, \mu_0)$ and (ii) for all persuaders $j = 1, 2, \ldots, J$*

$$V^j(\mu', h) > V^j(\mu, h) \Rightarrow \textit{Fit}(\mu'; h, \mu_0) \leq \textit{Fit}(\mu; h, \mu_0), \tag{5}$$

recalling that Fit$(\tilde{\mu}; h, \mu_0) = \max_m \Pr(h|m, \mu_0)$ such that $\mu(h, m) = \tilde{\mu}$ is a measure of the maximum fit of any model that induces posterior belief $\tilde{\mu}$ given data h.

This result provides a necessary and sufficient condition for checking whether a belief is an equilibrium belief. Invoking Lemma 1, we can rewrite (5) as the following necessary condition for μ to be an equilibrium belief: for all persuaders j, $V^j(\mu', h) > V^j(\mu, h) \Rightarrow$ Movement$(\mu'; \mu_0) \geq$ Movement$(\mu; \mu_0)$. Written this way, the result implies that a persuader is at an advantage when she wants 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 accurate interpretations. Rather, as Section VI illustrates, 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. And such a model promotes underreaction to data since it frames the data as 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:[26]

> 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 entrepreneur example, modifying it to consider two persuaders: one who wants the investor to invest, the other 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 that posterior probability. Then Proposition 3 implies that, in equilibrium, the investor will not react to the entrepreneur's history 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.*[27]

To formalize this argument, consider situations where there is a natural ordering of states, and the persuader is always better off (benefits from positive beliefs) or worse off (benefits from negative beliefs) if the receiver's expectation of the state is higher.

Definition 2. Suppose $\Omega \subset \mathbb{R}$. The persuader *benefits from positive beliefs* when $V^S(\mu, h) \leq V^S(\mu', h)$ whenever $\mathbb{E}_\mu[\omega] \leq \mathbb{E}_{\mu'}[\omega]$, with strict inequality if and only if $\mathbb{E}_\mu[\omega] < \mathbb{E}_{\mu'}[\omega]$. The persuader *benefits from negative beliefs* when $V^S(\mu, h) \geq V^S(\mu', h)$ whenever $\mathbb{E}_\mu[\omega] \leq \mathbb{E}_{\mu'}[\omega]$, with strict inequality if and only if $\mathbb{E}_\mu[\omega] < \mathbb{E}_{\mu'}[\omega]$.

Proposition 4. *Suppose there are multiple persuaders, the receiver is maximally open to persuasion, and $\Omega \subset \mathbb{R}$. If at least one persuader benefits from positive beliefs and at least one persuader benefits from negative beliefs, then μ_0 is the only equilibrium belief.*

This result generalizes the above examples and 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

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

[27] In the appendix, we illustrate how this intuition carries over to a restricted model space with an example. In the example, competition pushes beliefs towards the receiver's prior, but not all the way there. In other words, competition partially neutralizes the data. Thus, restrictions on the model space could help explain why we sometimes see underreaction in financial markets.

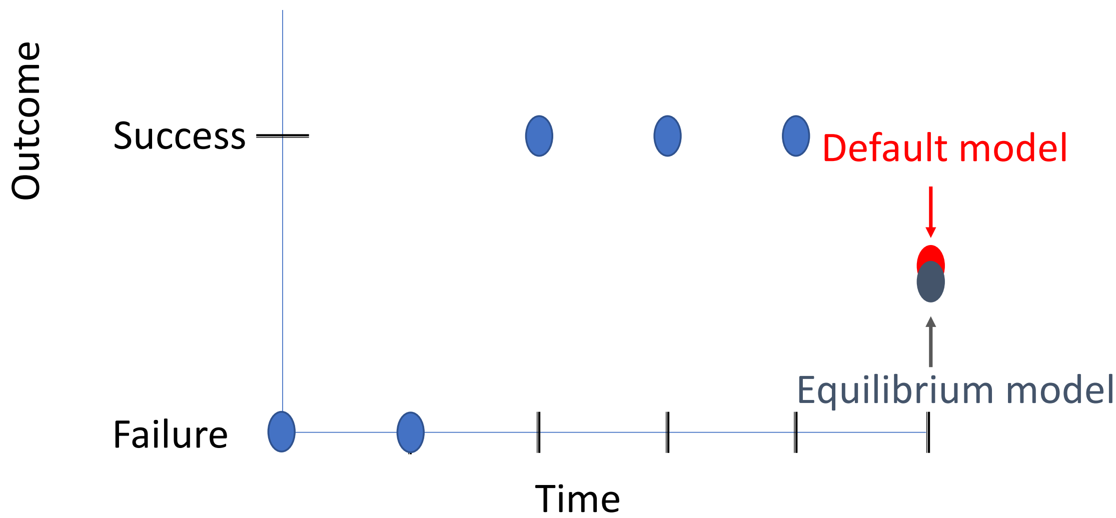


Figure 4: Competition between persuaders. One wants an investor to believe the entrepreneur's next startup will be successful and the other that it will be unsuccessful.

persuaders who have an interest in moving beliefs.[28] 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.

Turning back to Proposition 3, a third implication 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.*

[28]Under the conditions of Proposition 4, there is an equilibrium that implements μ_0 in undominated strategies. (Under more general conditions, it is not always true that there is an equilibrium implementing μ_0 in undominated strategies.) To see this, suppose the persuader benefits from positive beliefs and $\mathbb{E}_{\mu_0}[\omega] > \min_{\omega' \in \Omega} \omega'$. Consider the strategy of proposing the model $\pi_m(h|\omega) = 1$ for all $\omega \in \Omega$, which induces belief μ_0. This strategy is undominated: for every alternative strategy, proposing m yields a strictly higher payoff for the persuader given some profile of competitor strategies. Compare proposing m with proposing a model m' that, if adopted, induces a μ with $\mathbb{E}_{\mu}[\omega] > \mathbb{E}_{\mu_0}[\omega]$. Proposing m leads to strictly higher payoff for the persuader than proposing m' whenever a competitor proposes a model that fits better than m' and induces a belief μ' with $\mathbb{E}_{\mu'}[\omega] < \mathbb{E}_{\mu_0}[\omega]$. We can similarly analyze the other cases.

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 equilibrium belief if and only if there is a model inducing that belief that satisfies the (non-strategic) truthteller constraint.*

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—that is, proposing the model m where $\pi_m(h|\omega) = 1$ for all $\omega \in \Omega$.

We can illustrate this point in the entrepreneur example. There is a model that leads to the true-model posterior that fits the historical data around 29 times better than the true model.[29] A persuader who wants to induce optimistic beliefs is much more constrained competing with a strategic truthteller than with a non-strategic truthteller promoting the true model. Indeed, following the logic that the optimal model takes a cutoff form as in Figure 3b, we derive that the best the persuader can do against a strategic truthteller is to induce the belief that the entrepreneur's success probability is near 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.

V 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

[29] Let $\pi_m(h|\theta) = 3125/108 \cdot \theta^3(1-\theta)^2$ for all $\theta \in [0,1]$ and the history of 2 failures followed by 3 successes. Since π_m is proportional to the true likelihood of 2 failures followed by 3 successes, it preserves posterior beliefs following that history. And it is indeed a likelihood because $\max_{\theta\in[0,1]}\theta^3(1-\theta)^2 = 108/3125$.

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 II. As before, there is a broker who is incentivized to get investors to invest in an actively managed mutual fund, which is either good—meaning that it will have high net-of-fee returns in the future—or bad—meaning that it will have low net-of-fee returns in the future. 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 about future returns:

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

Oscar has an uninformative default, believing past returns are not informative about future returns:

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

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

$$\Pr{}^{\text{Pat}}(\text{high returns}|m) = \pi_m(\text{high returns}|\text{good}) \times 20\% + \pi_m(\text{high returns}|\text{bad}) \times 80\% = 52\%. \tag{6}$$

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

$$\Pr{}^{\text{Pat}}(\text{good}|m, \text{high returns}) = \pi_m(\text{high returns}|\text{good}) \frac{\mu_0^{\text{Pat}}(\text{good})}{\Pr^{\text{Pat}}(\text{high returns}|m)} = 38\%. \tag{7}$$

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 (6) and will now not invest by the analog of (7).[30]

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 $1/2 > \mu_0^j(g) \geq \Pr^j(h|d)/2$ for $j =$ optimist, pessimist.

Proposition 5. *Suppose that 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 $a = 1$:*

$$\text{(8)} \qquad \frac{\Pr^{\textit{optimist}}(h|d) - \mu_0^{\textit{optimist}}(g)}{1 - \mu_0^{\textit{optimist}}(g)} \leq \frac{\mu_0^{\textit{pessimist}}(g)}{1 - \mu_0^{\textit{pessimist}}(g)}.$$

[30] 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.

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, $\mu_0^{optimist} = \mu_0^{pessimist}$, or the same default interpretation, $\pi_d^{optimist} = \pi_d^{pessimist}$.*

The proposition and corollary imply that when there are multiple receivers, persuasion is more effective when the receivers share similar priors and default interpretations. The proposition strengthens the above examples by characterizing instances where the persuader is unable to send even a *menu* of public messages that simultaneously persuades 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.

Condition (8) of the proposition also has the implication that the fit of the optimist's default model plays an important role in determining whether there is a single message that is compelling to both receivers. This adds nuance to our previous result that with a single receiver tail events are ripe for persuasion. With multiple receivers, such events (e.g., crises) leave receivers open to being jointly persuaded by the same message.

VI Applications

Our framework lends itself to at least two types of applications. The first highlights a key contribution of our approach—it makes predictions about the content of persuasion. In this section, we give two examples of such applications, shedding light on why investors and business managers appear influenced by messages that invite them to reach misleading conclusions from the data. Applying our framework in this way requires some analyst judgment in specifying how a given setting maps into our framework, which we discuss in more detail at the end of this section.

Second, our model makes testable predictions on the impact of persuasion in specific settings. We return to a motivating example from the Bayesian Persuasion literature, persuading jurors, and show how incorporating model persuasion alters conclusions from the analysis.We conclude this section by collecting predictions to explore in future applications.

VI.A 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 at which there is posited to be high latent demand, which prevents prices from falling further. Resistance is a price point 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 reminiscient of ones shown by brokers to educate clients,[31] Figure 5b shows a real world example of technical analysis from TradingView.com. The analysis was done by a firm that sells investment advice 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 setting with 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 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.[32] The ubiquity of technical analysis is puzzling given that it is arguably ineffective in producing trading profits (Lo, Mamaysky, and Wang 2000; Bajgrowicz and Scaillet 2012).

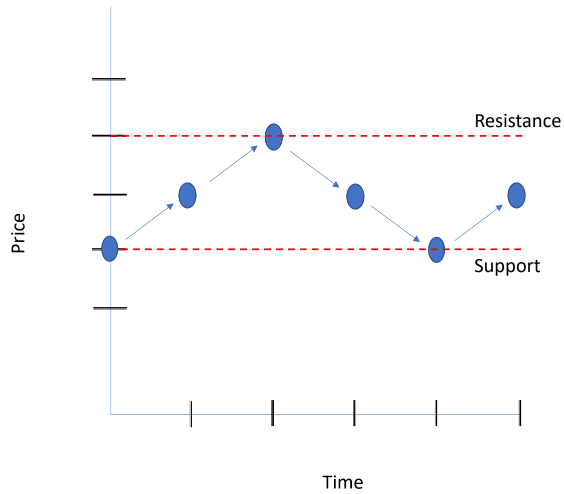
Why is technical analysis so common if it does not reliably generate profits? Basic lessons of our framework may shed light on this question: a key advantage of any model persuader is that he can tailor models to the data. For instance, the support and resistance model looks so compelling in Figures 5a and 5b because the support and resistance levels are chosen after seeing the data.

In Appendix G.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.[33] 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 al-

[31]For example, see the entry from Fidelity's "Learning Center" for investors on support/resistance models: https://www.fidelity.com/learning-center/trading-investing/technical-analysis/introduction-technical-analysis/support-and-resistance

[32]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).

[33]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.



(a) Illustration of Support and Resistance.

(b) Technical Analysis: Long Amazon.

(c) Technical Analysis: Short Amazon.

Figure 5: Technical Analysis

ways 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 G.1, we formalize this model 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. It is worth noting that we are showing that the proposed models beat the default, not that they are optimal. In other words, we are explaining why a model like support/resistance is compelling, not exactly which model becomes most popular in the world. To explain exactly which model becomes popular, we would need to add more structure to the problem, specifying the persuader's incentives, the model space, the receiver's action space, the competitive structure, etc.

VI.B 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 broad idea that mistakes are driven by biased advice is incomplete. People 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: they find biased advice more compelling than the truth.

The key intuition was presented in the simple example of investment advice from Section II. We develop a more elaborate formulation in Appendix G.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.[34] This design was explicit. As the author writes:

[34]The firms ("leap years") were Abbott Laboratories (1974), Circuit City Stores (1982), Fannie Mae (1984), Gillette

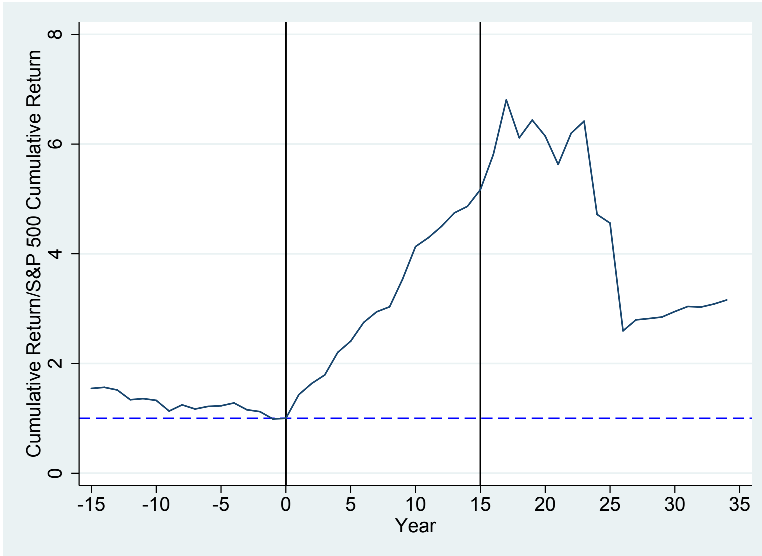


Figure 6: Performance of "Good-to-Great Firms" Relative to the Aggregate Stock Market

> 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.[35] Year 0 on the horizontal axis corresponds to the year that Collins argues the companies made the leap from good to great (each firm made the leap in a different calendar year); Year 15 corresponds to the last year that Collins includes in his analysis; Years 15-35 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 G.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

Company (1980), Kimberly-Clark (1972), Kroger (1973), Nucor (1975), Philip Morris (1964), Pitney Bowes (1973), Walgreens (1975), and Wells Fargo (1983). 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.

[35] Calculated based on data from the Monthly Stock File ©2020 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business, accessed on June 19, 2019. Wharton Research Data Services (WRDS) was used in preparing this paper. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

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 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, the firms studied have had slightly below average performance. 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 feat of model persuasion, Collins's book remains popular: it was a top-5 bestselling business book in 2016-2017, 15 years after publication.[36]

VI.C 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, model persuasion works to neutralize evidence that is open to interpretation but informative under the true model: 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, convicting a defendant if and only if his posterior 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 defense's payoff is v if the defendant is acquitted and 0

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

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 the representative juror is maximally open to persuasion and all evidence is open to interpretation. Then Proposition 3 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—that is, condition (5) is violated. 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 reduces, 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 (e.g., the defendant had blood on his hands at the scene of the crime) and softer or more circumstantial evidence (e.g., the defendant does not have a great alibi). 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 a similar conclusion to what they would in the absence of the arguments and data those arguments frame.[37] Relaxing the assumption of a common prior, these results may shed light on the importance of juror, judge, or arbitrator

[37]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". If jurors evaluate the defense and prosecution's arguments sequentially rather than at the same time, certain dynamic extensions of our model will modify this conclusion by suggesting a first-mover advantage in framing the evidence. We will return to this issue in Section VII.

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.

VI.D Some Guidance for Further Applications

The above examples illustrate the steps necessary to apply our model to understand the content of model persuasion in empirical settings. The key ingredients of our framework can enumerated by writing out the receiver's posterior expected utility over actions, as in Section I.C:

$$\sum_{\omega' \in \Omega} \sum_{y' \in Y} \tilde{U}^R(a, y') f(y'|\omega') \underbrace{\frac{\pi_m(h|\omega')}{\Pr(h|m, \mu_0)} \mu_0(\omega')}_{\mu(\omega'|h,m)},$$

where a denotes actions, y denotes outcomes, and ω denotes states. Many of these ingredients, including actions, outcomes, and utility functions, follow directly from the setting. There are five areas where analyst judgment may be necessary: (i) what to include as part of the history h, (ii) how to distinguish between outcomes y and states ω, (iii) how to distinguish between models m and states ω, (iv) what to make the default model d, and (v) how to specify the prior over states μ_0.

Broadly speaking there are two sources of guidance on how to specify these elements: (i) existing models of the setting and (ii) what persuaders themselves say. In many cases, like our persuading jurors application, almost all of the elements can be taken from a rational Bayesian model of the setting. In other cases, like our technical analysis example above, the data persuaders draw on in practice can help inform analyst judgment. For instance, under the true model, little or no information predicts whether a stock will rise on a given day, but persuaders often point to the stock's recent price movements. This suggests recent price movements should be contained in the history h and that receivers are willing to entertain models framing those price movements.

What persuaders say can also inform the distinction between outcomes and states. As we discuss in Section I.C, some situations appear to involve irreducible uncertainty, which can be captured by a nondegenerate distribution of outcomes given states, $f(y|\omega)$. For instance, a persuader may be able to perfectly convince a receiver of an entrepreneur's quality (i.e., success probability), but not convince the receiver that the entrepreneur's next startup will for sure be a success. Similarly, the distinction between states and models can also be informed by what persuaders say. Roughly speaking, states are about the conclusions that persuaders want receivers to reach, while models are the explanations for why the data implies those conclusions. In our technical analysis example, the persuader wants the receiver to conclude that the stock will rise (a state) and explains that the previous price history implies that conclusion (a model).

In many cases, the choice of the receiver's default model and prior can be guided by "off the shelf" models of the setting. For instance, in the technical analysis example, we use the random walk hypothesis (Fama 1965) as the default model. The true model is another natural default and can serve as a useful rational benchmark. Similarly, priors can be informed by existing models of the setting. For instance, in our Good-to-Great example, we specify the receiver's prior over expected stock returns using a standard finance model, the Capital Asset Pricing Model. In other cases, the analyst can specify the receiver's prior over states as in empirical Bayesian analysis. Roughly speaking, this involves using data from analogous problems to discipline receivers' priors in a sensible way. For instance, while a receiver may only be willing to entertain models framing the price history of a single stock in deciding whether to buy that stock, it may be sensible for an analyst to discipline the receiver's prior based on the price histories of a greater universe of stocks.

Note that both existing models and what persuaders say need only be used to specify building-block elements of our framework (e.g., what data should be included in the history). Once these decisions are made, our framework provides testable predictions about how the history will be framed. These predictions suggest directions for future work. While we analyze popular messages and try to shed light on why they are persuasive, the model predicts messages that do not fit receivers' knowledge well will not be compelling. Are such messages indeed less common or popular? We also make predictions about settings where model persuasion is likely to be effective, e.g., when there is a lot of randomness, when receivers do not have ready explanations for what they see, etc. Do we indeed tend to see a lot of persuasive activity in such settings? We predict that in settings where model persuasion is effective, exposure to the true model, e.g., from watchdogs, consumer advocates, or disclosure regulations, is unlikely to guide receivers to the best choice. Is such truthtelling ineffective, as we predict? In our framework, competition between persuaders with opposing interests neutralizes the audience's reaction to data. What is the impact of such competition in practice? We show that model persuasion is easier when audience members are more similar. In reality, do persuaders expend more resources on targeted messaging when facing heterogeneous audiences?

VII Extensions and Robustness

This section briefly considers three classes of extensions and modifications to our analysis: the first to dynamics, the second to receiver sophistication, and the third to receiver knowledge.

VII.A Dynamics

We first consider dynamics. Take the application from the previous section on the book "Good to Great" (Collins 2001) as a motivating example. 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), highlighting that the "good to great" companies included Circuit City and Fannie Mae, which both failed. 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 greatness. One of his later books is in fact called "How the Mighty Fall" (Collins 2009).

How do receivers respond to such shifting models over time? To answer this question, we need to extend our static framework to consider dynamics. A simple possibility is that our static model applies in dynamic situations: receivers may not take into account persuaders' previous statements in evaluating models they are currently proposing. However, as in the Collins example, sometimes persuaders do have to confront their previous statements. Proposition 2 suggests one intuition for such cases. Suppose that data closed to interpretation arises from previous models supplied by a persuader. Then the proposition suggests that new data liberates the persuader from his previous statements. But this result does not fully address dynamic considerations the persuader might face. For instance, supplying a myopically optimal model might be constraining in the future.

While a full analysis of such dynamic considerations is beyond the scope of this paper, in Appendix F we extend our framework to a simple dynamic environment, where sequentially (1) there is a signal h_1, (2) the persuader proposes a model m_1, (3) there is a signal h_2, (4) the persuader proposes another model m_2, and (5) the receiver makes a decision. Note that the receiver only makes a single decision.

We consider three ways to extend our framework to this setting. The specifications differ in the way the first model proposed by the persuader influences the receiver. In the first case, which we call *prior dynamics*, the model influences the receiver's knowledge going forward. In particular, the persuader's first model, m_1, influences the receiver's prior going into the second period. Essentially, the receiver holds the persuader to implications of the first model proposed, but not the model itself. In the second and third cases, the receiver holds the persuader to the first model itself, but comes into the second period with the prior he started with. In the second case, which we call *consistency dynamics*, the persuader always has to propose models that are consistent with previous statements.[38] One interpretation is that the receiver remembers that the persuader proposed the first model and harshly penalizes the persuader for contradicting herself. In the third case, which we call *default dynamics*, the persuader's first model, m_1, becomes the receiver's default model going forward. One interpretation is that the receiver allows the persuader to contradict herself, but only by proposing a more compelling model.

In all three specifications, model persuasion remains effective. While dynamics may constrain persuaders relative to the static specification for some realized histories, under general conditions model persuaders strictly benefit from persuasion. Moreover, the consistency dynamics specification does not constrain model persuaders at all if persuaders have the foresight to supply a model today that will be consistent with whatever model they want to supply tomorrow.[39] In a sense, this generalizes the Collins example above and shows how persuaders may benefit from tailoring models to the data over time without receivers being able to easily recognize they are doing so.[40]

In Appendix F, we also consider an example of sequential competition. Returning to the jury application, we illustrate how the framework accommodates phenomena such as "preemptive framing" under our prior dynamics specification. We assume that the jury starts with a prior that favors acquittal. We show that if the prosecutor goes first, she may be able to frame the evidence so persuasively that (i) the jury's posterior will be for conviction and (ii) no reframing of the evidence by the defense can convince the jury to acquit. This analysis modifies the conclusions of the static example where competition neutralizes the evidence entirely. Whether a static or dynamic analysis of competition is more relevant to a given situation is a topic for future research, but it likely depends on factors such as whether the spacing between arguments is long enough that one side's argument has time to "sink in" before the competitor's argument.

[38] Formally, model m_2 must satisfy $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega) = \pi_{m_1}(h_1|\omega)$ for all $\omega \in \Omega$.That is, the probability of the first signal, h_1, must be the same under model m_2, integrating over all realizations of the second signal h_2, as it was under m_1.

[39] We show in Appendix F that the following strategy gives persuaders the same flexibility as they have in the static setting: until the period in which a receiver makes a decision, the persuader says "we cannot learn anything from the data so far because it was inevitable". In the period in which the receiver makes a decision, the persuader says "with the most recent data, we have learned [statically optimal model]".

[40] Dynamics could be constraining for model persuaders in a way that is not fully reflected in the above discussion: it could in effect give receivers a holdout sample to evaluate the persuaders' arguments. This would be the case if, unlike the dynamics described above, the persuader cannot propose a second model after more data is released. While we leave a full analysis for future work, note that this assumption is somewhat similar to the case of multiple receivers above. In both cases, the persuader does not quite know the lens that receivers will bring to the data, which likely mitigates, but does not eliminate, the impact of model persuasion.

VII.B Receiver Sophistication

Our analysis embeds a form of receiver naivete beyond the crucial assumption that receivers do not take into account persuaders' flexibility in proposing models after seeing the data: Receivers do not take persuaders' incentives into account in reacting to their proposed models.

Our results are qualitatively robust to relaxing this assumption in straightforward ways. We show in Appendix E that we obtain qualitatively similar results modifying Eq. (1) so that persuaders whose incentives are known to be misaligned with receivers' must satisfy tighter constraints. This will tend to reduce receivers' sensitivity to all data. Thus, receiver skepticism may in fact backfire, while still leaving room for misleading persuasion.[41]

VII.C Receiver Knowledge

In our analysis, persuaders are unable to directly alter a receiver's prior over states or her knowledge of the utility consequences of those states. Appendix E.1 broadens the conception of a model to be a joint distribution over payoff-relevant outcomes, observable histories, and states to help make these assumptions more explicit by representing them as restrictions to allowable models in this broader space. The appendix also shows how these restrictions are crucial for our analysis.

The broader conception also hints at ways to make principled refinements to the model space to reflect the receiver's knowledge. We briefly consider three classes of refinements in Appendix E.1: refinements that (i) capture knowledge about the distribution over observables, (ii) reflect views about internal consistency, and (iii) reflect knowledge about the true model.

We show that refinements along the lines of (i) do not meaningfully constrain the beliefs a single persuader is able to induce. The reason is that a persuader who is able to induce some belief with a model that fits better than the receiver's default is also able to induce the belief with a model that fits as well as the receiver's default. However, refinements along the lines of (ii) and (iii) would be constraining to the persuader. We leave an analysis of such refinements to future work.

VIII Discussion

This paper presents a framework for analyzing model persuasion—persuasion that operates by providing receivers with models for interpreting data they already have ready access to. Such

[41] Our analysis embeds another form of receiver naivete: receivers select models they are exposed to instead of averaging across them. The appendix to our earlier working paper, Schwartzstein and Sunderam (2020), performs a preliminary analysis assuming that receivers average over their default and the model the persuader proposes. We show that this form of model averaging is sometimes more constraining to the persuader than model selecting (e.g., if the receiver's default fits poorly), but in many situations is less constraining (e.g., if the receiver's default fits well). We also show that key qualitative insights on when receivers are persuadable do not seem to hinge on the assumption that receivers select rather than average models.

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

While model persuasion is an important kind of persuasion, it is not the only type. For instance, model persuasion is related to, but distinct from, cherry picking and slanting, which are about selective reporting of facts (Hayakawa 1940; Milgrom 1981; Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006). Model persuasion can sometimes operate similarly to cherry picking. However, it often operates quite differently, with proposed 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 1995-1999 look from all previous years. Similarly, model persuasion does not capture the social (Cialdini 1993) or emotional elements key to many kinds of advertising. Nor does it capture situations where persuaders provide incorrect data to receivers.

Our framework is amenable to a number of applications and extensions. Some extensions, e.g., on receiver sophistication and dynamics are briefly explored in Appendices E and F. 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 true 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 assumption that the data is fixed abstracts from potential interactions between Bayesian and model persuasion. For example, we provide an illustrative example in Appendix E.3 showing that model persuaders may prefer to collect and reveal information ex ante if they can frame it in beneficial ways ex post. But analyzing the case where data is publicly available and exogenous 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 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 $\min_{\omega\in\Omega} \mu_0(\omega)/\tilde{\mu}(\omega)$. $\square$

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

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

by Bayes' Rule. Since $\pi_m(h|\omega) \leq 1$ and, under the constraint Eq (1), $\Pr(h|m,\mu_0) \geq \Pr(h|d,\mu_0)$, 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

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

$\square$

Proof of Proposition 2. Let $h^{i,j} = (h_1^i, h_2^j)$. We have

$$\begin{aligned} \mu(\omega|h^{i,j},m) &= \frac{\pi_{m^T}(h_1^i|\omega)\pi_m(h_2^j|\omega)\mu_0(\omega)}{\sum_{\omega'\in\Omega} \pi_{m^T}(h_1^i|\omega')\pi_m(h_2^j|\omega')\mu_0(\omega')} \\ &\leq \frac{\pi_{m^T}(h_1^i|\omega)\mu_0(\omega)}{\sum_{\omega'\in\Omega} \pi_{m^T}(h_1^i|\omega')\pi_d(h_2^j|\omega')\mu_0(\omega')}. \end{aligned} \tag{A.9}$$

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

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

To establish the first part of the result, rewrite the inequality above as

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

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 there is a $\bar{j}$ such that for all $j \geq \bar{j}$,

$$\frac{\mu(\omega|h^{i,j},m)}{\mu_0(\omega)\pi_{m^T}(h_1^i|\omega)} \cdot \Pr(h^{i,j}|d,\mu_0) \leq 1 \; \forall \; \omega \in \tilde{\Omega},$$

since $\pi_{m^T}(h_1^i|\omega) > 0$ for all $\omega \in \tilde{\Omega}$ and $\pi_d(h_2^j|\omega') \to 0$ as $j \to \infty$ for all $\omega' \in \Omega$. Taking $j \geq \bar{j}$, any belief μ with support on $\tilde{\Omega}$ is then implementable with a model that sets $\pi_m(h_2^j|\omega) = 0$ for $\omega \in \Omega \setminus \tilde{\Omega}$ and

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

□

Proof of Corollary 1. Follows directly from Proposition 1.

□

Proof of Proposition 3. 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 (5) implies that it is impossible for any persuader to unilaterally deviate to a model m' that benefits them for which $\Pr(h|m', \mu_0) > \Pr(h|m, \mu_0)$.

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 (5) 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 Fit$(\mu'; h, \mu_0) >$ Fit$(\mu; h, \mu_0)$: 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. 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 (5) given Lemma 1.

The second part follows from the fact that adding persuaders just adds more constraints that need to hold in order to satisfy (5).

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 (5).

□

Proof of Proposition 4. Suppose $\mu \neq \mu_0$ is an equilibrium belief. There exists a $\tilde{\mu}$ such that Movement$(\tilde{\mu}; \mu_0) <$ Movement$(\mu; \mu_0)$ and either (i) $\mathbb{E}_{\tilde{\mu}}[\omega] > \mathbb{E}_{\mu}[\omega]$ or (ii) $\mathbb{E}_{\tilde{\mu}}[\omega] < \mathbb{E}_{\mu}[\omega]$. To see this, first note that it is trivially true when $\mathbb{E}_{\mu}[\omega] \neq \mathbb{E}_{\mu_0}[\omega]$ because we can then just take $\tilde{\mu} = \mu_0$. So it remains to check the case where $\mu \neq \mu_0$ but $\mathbb{E}_{\mu}[\omega] = \mathbb{E}_{\mu_0}[\omega]$. Let $\bar{\Omega} = \arg\max_{\omega \in \Omega} \mu(\omega)/\mu_0(\omega)$.

If $\max_{\omega\in\Omega}\omega \notin \bar{\Omega}$, then take

$$\tilde{\mu}(\omega) = \begin{cases} \mu(\omega) & \text{if } \omega \notin \bar{\Omega} \cup \{\max_{\omega'\in\Omega}\omega'\} \\ \mu(\omega) - \varepsilon & \text{if } \omega \in \bar{\Omega} \\ \mu(\omega) + |\bar{\Omega}| \cdot \varepsilon & \text{if } \omega = \max_{\omega'\in\Omega}\omega' \end{cases}$$

for $\varepsilon > 0$ sufficiently small. For this $\tilde{\mu}$, $\mathbb{E}_{\tilde{\mu}}[\omega] > \mathbb{E}_{\mu}[\omega]$ and $\text{Movement}(\tilde{\mu};\mu_0) < \text{Movement}(\mu;\mu_0)$. We can similarly analyze the situation where $\min_{\omega\in\Omega}\omega \notin \bar{\Omega}$. The remaining situation is where both $\max_{\omega\in\Omega}\omega \in \bar{\Omega}$ and $\min_{\omega\in\Omega}\omega \in \bar{\Omega}$. In this case, take

$$\tilde{\mu}(\omega) = \begin{cases} \mu(\omega) - \varepsilon_1 & \text{if } \omega = \min_{\omega'\in\Omega}\omega' \\ \mu(\omega) - \varepsilon_2 & \text{if } \omega = \max_{\omega'\in\Omega}\omega' \\ \mu(\omega) - \varepsilon_3 & \text{for } \omega \in \bar{\Omega} \setminus \{\min_{\omega'\in\Omega}\omega', \max_{\omega'\in\Omega}\omega'\} \\ \mu(\omega) + \varepsilon_1 + \varepsilon_2 + \left|\bar{\Omega} \setminus \{\min_{\omega'\in\Omega}\omega', \max_{\omega'\in\Omega}\omega'\}\right| \cdot \varepsilon_3 & \text{for some } \omega \notin \bar{\Omega} \\ \mu(\omega) & \text{otherwise.} \end{cases}$$

There exist strictly positive $\varepsilon_1, \varepsilon_2, \varepsilon_3$ such that $\tilde{\mu}$ remains a probability measure and $\mathbb{E}_{\tilde{\mu}}[\omega] \neq \mathbb{E}_{\mu}[\omega]$. Further, by construction, $\text{Movement}(\tilde{\mu};\mu_0) < \text{Movement}(\mu;\mu_0)$.

Finally, return to cases (i) and (ii) above and recall Lemma 1. In case (i), μ is not in fact an equilibrium belief because the persuader who benefits from positive beliefs is able to profitably deviate to a model that induces $\tilde{\mu}$. In case (ii), μ is not in fact an equilibrium belief because the persuader who benefits from negative beliefs is able to profitably deviate to a model that induces $\tilde{\mu}$.

□

Proof of Corollary 3. This is a corollary of Propositions 1 and 3. By Eq (3), there is a model inducing μ which satisfies the (non-strategic) truthteller constraint if and only if $\max_{\omega\in\Omega}\mu(\omega)/\mu_0(\omega) \leq 1/\Pr(h|m^T,\mu_0)$. By Eq (5), $\mu \neq \mu_h$ 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 $\text{Fit}(\mu;h,\mu_0) \geq \Pr(h|m^T,\mu_0)$ (by Proposition 3), which is equivalent to $\max_{\omega\in\Omega}\mu(\omega)/\mu_0(\omega) \leq 1/\Pr(h|m^T,\mu_0)$ (by Lemma 1).

It suffices to show that $\max_{\omega\in\Omega}\mu_h(\omega)/\mu_0(\omega) \leq 1/\Pr(h|m^T,\mu_0)$ with equality if and only if $\max_{\omega\in\Omega}\pi(h|\omega) = 1$. Note that, after rearranging 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 (8) 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 $\Pr^{\text{optimist}}(h|m) \geq \Pr^{\text{optimist}}(h|d)$, or, equivalently,

$$\text{(A.10)} \quad \pi_m(h|b)(1-\mu_0^{\text{optimist}}(g)) + \pi_m(h|g)\mu_0^{\text{optimist}}(g) \geq \Pr{}^{\text{optimist}}(h|d) \iff$$

$$\text{(A.11)} \quad \pi_m(h|b) \geq \frac{\Pr^{\text{optimist}}(h|d) - \mu_0^{\text{optimist}}(g)\pi_m(h|g)}{1-\mu_0^{\text{optimist}}(g)}.$$

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

$$\text{(A.12)} \quad \frac{\pi_m(h|g)\mu_0^{\text{pessimist}}(g)}{\pi_m(h|g)\mu_0^{\text{pessimist}}(g) + \pi_m(h|b)(1-\mu_0^{\text{pessimist}}(g))} \geq 1/2$$

$$\text{(A.13)} \quad \frac{\mu_0^{\text{pessimist}}(g)\pi_m(h|g)}{1-\mu_0^{\text{pessimist}}(g)} \geq \pi_m(h|b).$$

Since the right hand side of Eq (A.11) is decreasing in $\pi_m(h|g)$ and the left hand side of Eq (A.13) is increasing in $\pi_m(h|g)$, there is a message that simultaneously satisfies the two inequalities if and only if Eq (8) holds.

Now we establish that Eq (8) 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

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

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 (A.13) 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 (8) 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

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

But the existence of a message that satisfies this condition when Eq (8) 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 (8) boils down to

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

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

If receivers share the same default interpretation, then

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

which means that Eq (8) 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.

□

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. 2016. "Fragility of Asymptotic Agreement Under Bayesian Learning." *Theoretical Economics*, 11(1): 187–225.

Alonso, Ricardo, and Odilon Câmara. 2016. "Persuading Voters." *The American Economic Review*, 106(11): 3590–3605.

Andreassen, Paul B. 1990. "Judgmental Extrapolation and Market Overreaction: On the Use and Disuse of News." *Journal of Behavioral Decision Making*, 3(3): 153–174.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics*, 127(2): 1017–1055.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2014. "The Role of Age in Jury Selection and Trial Outcomes." *The Journal of Law and Economics*, 57(4): 1001–1030.

Aragones, Enriqueta, Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler. 2005. "Fact-Free Learning." *American Economic Review*, 95(5): 1355–1368.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Bajgrowicz, Pierre, and Olivier Scaillet. 2012. "Technical Trading Revisited: False Discoveries, Persistence Tests, and Transaction Costs." *Journal of Financial Economics*, 106(3): 473–491.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics*, 49(3): 307–343.

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

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

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

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. "Narratives, Imperatives and Moral Reasoning."

Bergstresser, Daniel, John MR Chalmers, and Peter Tufano. 2008. "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry." *The Review of Financial Studies*, 22(10): 4129–4156.

Bhattacharya, Utpal, Andreas Hackethal, Simon Kaesler, Benjamin Loos, and Steffen Meyer. 2012. "Is Unbiased Financial Advice to Retail Investors Sufficient? Answers From a Large Field Study." *The Review of Financial Studies*, 25(4): 975–1032.

Billingsley, Randall S, and Don M Chance. 1996. "Benefits and Limitations of Diversification Among Commodity Trading Advisors." *Journal of Portfolio Management*, 23(1): 65.

Broockman, David, and Joshua Kalla. 2016. "Durably Reducing Transphobia: A Field Experiment on Door-to-Door Canvassing." *Science*, 352(6282): 220–224.

Bruner, Jerome. 1991. "The Narrative Construction of Reality." *Critical Inquiry*, 18(1): 1–21.

Busby, Ethan, DJ Flynn, James N Druckman, and Scott Hall. 2018. "Studying Framing Effects: Existing Research and Lingering Questions."

Cain, Daylian M, George Loewenstein, and Don A Moore. 2005. "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest." *The Journal of Legal Studies*, 34(1): 1–25.

Campbell, John Y. 2006. "Household Finance." *The Journal of Finance*, 61(4): 1553–1604.

Chalmers, John, and Jonathan Reuter. 2012. "Is Conflicted Investment Advice Better Than No Advice?" National Bureau of Economic Research.

Chater, Nick, and George Loewenstein. 2016. "The Under-Appreciated Drive for Sense-Making." *Journal of Economic Behavior & Organization*, 126: 137–154.

Cheung, Yin-Wong, and Menzie David Chinn. 2001. "Currency Traders and Exchange Rate Dynamics: A Survey of the US Market." *Journal of International Money and Finance*, 20(4): 439–471.

Chong, Dennis, and James N Druckman. 2007. "Framing Theory." *Annual Review of Political Science*, 10: 103–126.

Collins, James Charles. 2009. *How the Mighty Fall: And Why Some Companies Never Give In.* Random House.

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

Crawford, Vincent P, and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 1431–1451.

CRSP. 1949–2018. "Monthly Stock File." The University of Chicago Booth School of Business. Accessed from Wharton Research Data Services (WRDS) on June 19, 2019. http://wrds–www.wharton.upenn.edu.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics*, 122(3): 1187–1234.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics*, 2(1): 643–669.

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

DiFonzo, Nicholas, and Prashant Bordia. 1997. "Rumor and Prediction: Making Sense (but Losing Dollars) in the Stock Market." *Organizational Behavior and Human Decision Processes*, 71(3): 329–353.

Egan, Mark L, Gregor Matvos, and Amit Seru. 2018. "Arbitration With Uninformed Consumers." National Bureau of Economic Research.

Eliaz, Kfir, and Ran Spiegler. 2019. "A Model of Competing Narratives."

Eliaz, Kfir, Ran Spiegler, and H Thysen. 2019. "Strategic Interpretations." mimeo.

Ely, Jeffrey C. 2017. "Beeps." *American Economic Review*, 107(1): 31–53.

Epstein, Larry G, and Martin Schneider. 2007. "Learning Under Ambiguity." *The Review of Economic Studies*, 74(4): 1275–1303.

Esponda, Ignacio, and Demian Pouzo. 2016. "Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models." *Econometrica*, 84(3): 1093–1130.

Eyster, Erik. 2019. "Errors in Strategic Reasoning." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2, 187–259. Elsevier.

Eyster, Erik, and Michele Piccione. 2013. "An Approach to Asset Pricing Under Incomplete and Diverse Perceptions." *Econometrica*, 81(4): 1483–1506.

Fama, Eugene F. 1965. "The Behavior of Stock-Market Prices." *The Journal of Business*, 38(1): 34–105.

Fischhoff, Baruch. 1975. "Hindsight is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty." *Journal of Experimental Psychology: Human Perception and Performance*, 1(3): 288.

Fisher, Walter R. 1985. "The Narrative Paradigm: In the Beginning." *Journal of Communication*, 35(4): 74–89.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein. 2018. "Channeled Attention and Stable Errors."

Galbraith, John Kenneth. 1967. *The New Industrial State.*

Galperti, Simone. 2019. "Persuasion: The Art of Changing Worldviews." *American Economic Review*, 109(3): 996–1031.

Gentzkow, Matthew, and Emir Kamenica. 2012. "Competition in Persuasion."

Gentzkow, Matthew, and Emir Kamenica. 2017. "Competition in Persuasion." *The Review of Economic Studies*, 84(1): 300–322.

Gentzkow, Matthew, and Jesse M Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy*, 114(2): 280–316.

Gentzkow, Matthew, and Jesse M Shapiro. 2008. "Competition and Truth in the Market for News." *Journal of Economic Perspectives*, 22(2): 133–154.

Gershman, Samuel J. 2018. "How to Never be Wrong." *Psychonomic Bulletin & Review*, 1–16.

Glazer, Jacob, and Ariel Rubinstein. 2004. "On Optimal Rules of Persuasion." *Econometrica*, 72(6): 1715–1736.

Glazer, Jacob, and Ariel Rubinstein. 2006. "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach." *Theoretical Economics*, 1: 395–410.

Goffman, Erving. 1974. *Frame Analysis: An Essay on the Organization of Experience.* Harvard University Press.

Hackethal, Andreas, Michael Haliassos, and Tullio Jappelli. 2012. "Financial Advisors: A Case of Babysitters?" *Journal of Banking & Finance*, 36(2): 509–524.

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

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

Heidhues, Paul, and Botond Kőszegi. 2018. "Behavioral Industrial Organization." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 1, 517–612. Elsevier.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack. 2018. "Unrealistic Expectations and Misguided Learning." *Econometrica*, 86(4): 1159–1214.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *The American Economic Review*, 101(6): 2590–2615.

Kass, Robert E, and Adrian E Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association*, 90(430): 773–795.

Kindleberger, Charles P, and Robert Aliber. 2010. *Manias, Panics, and Crashes.* . 6 ed., Palgrave Macmillan.

Lakoff, George. 2004. *Don't Think of an Elephant!: Know Your Values and Frame the Debate.* Chelsea Green Publishing.

Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By.* University of Chicago press.

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

Levy, Gilat, and Ronny Razin. 2020, forthcoming. "The Drowning Out of Moderate Voices: A Maximum Likelihood Approach to Combining Forecasts." *Theoretical Economics*.

Lo, Andrew W, Harry Mamaysky, and Jiang Wang. 2000. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation." *The Journal of Finance*, 55(4): 1705–1765.

Malmendier, Ulrike, and Devin Shanthikumar. 2007. "Are Small Investors Naive About Incentives?" *Journal of Financial Economics*, 85(2): 457–489.

Milgrom, Paul, and John Roberts. 1986. "Relying on the Information of Interested Parties." *The RAND Journal of Economics*, 18–32.

Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications." *The Bell Journal of Economics*, 380–391.

Mullainathan, Sendhil. 2002. "Thinking Through Categories."

Mullainathan, Sendhil, and Andrei Shleifer. 2005. "The Market for News." *American Economic Review*, 95(4): 1031–1053.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. "Coarse Thinking and Persuasion." *The Quarterly Journal of Economics*, 123(2): 577–619.

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

Olea, José Luis Montiel, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat. 2019. "Competing Models." *arXiv preprint arXiv:1907.03809*.

Olszewski, Wojciech. 2018. "Self-Esteem Preferences and Imaginary Effects in Information Processing."

Ortoleva, Pietro. 2012. "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News." *American Economic Review*, 102(6): 2410–36.

Pennington, Nancy, and Reid Hastie. 1986. "Evidence Evaluation in Complex Decision Making." *Journal of Personality and Social Psychology*, 51(2): 242.

Pennington, Nancy, and Reid Hastie. 1988. "Explanation-Based Decision Making: Effects of Memory Structure on Judgment." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3): 521.

Pennington, Nancy, and Reid Hastie. 1992. "Explaining the Evidence: Tests of the Story Model for Juror Decision Making." *Journal of Personality and Social Psychology*, 62(2): 189.

Pons, Vincent. 2018. "Will a Five-Minute Discussion Change Your Mind? A Countrywide Experiment on Voter Choice in France." *American Economic Review*, 108(6): 1322–63.

Roberts, Seth, and Harold Pashler. 2000. "How Persuasive is a Good Fit? A Comment on Theory Testing." *Psychological Review*, 107(2): 358.

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

Schulz, Laura E, and Jessica Sommerville. 2006. "God Does Not Play Dice: Causal Determinism and Preschoolers' Causal Inferences." *Child Development*, 77(2): 427–442.

Schwartzstein, Joshua. 2014. "Selective Attention and Learning." *Journal of the European Economic Association*, 12(6): 1423–1452.

Schwartzstein, Joshua, and Adi Sunderam. 2019. "Using Models to Persuade." *NBER Working Paper*, , (w26109).

Severin, J Werner, and James W Tankard. 2001. "Communication Theory: Origin, Methods, and Uses in The Mass Media."

Shiller, Robert J. 2015. *Irrational Exuberance.* Princeton University press.

Shiller, Robert J. 2017. "Narrative Economics." *American Economic Review*, 107(4): 967–1004.

Smith, David M, Christophe Faugère, and Ying Wang. 2014. "Head and Shoulders Above the Rest? The Performance of Institutional Portfolio Managers who use Technical Analysis." In *Research in Finance*. 167–189. Emerald Group Publishing Limited.

Spiegler, Ran. 2016. "Bayesian Networks and Boundedly Rational Expectations." *The Quarterly Journal of Economics*, 131(3): 1243–1290.

Taylor, Mark P, and Helen Allen. 1992. "The Use of Technical Analysis in the Foreign Exchange Market." *Journal of International Money and Finance*, 11(3): 304–314.

Weick, Karl E. 1995. *Sensemaking in Organizations.* Vol. 3, Sage.