Online Appendix: Using Models to Persuade

Joshua Schwartzstein
Harvard Business School

Adi Sunderam
Harvard Business School

B Basic Model Properties

The *impact of persuasion* for agent j is the expected change in j's payoff: $\mathbb{E}[V^j(h, m(h)) - V^j(h, d)]$, where $\mathbb{E}[\cdot]$ is taken over the true distribution of histories h, i.e., with respect to the prior and the true likelihood (μ_0, π). We can decompose the impact of persuasion on the receiver as:

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

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

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

Observation A.1 (Model Properties)**.**

1. The information component of persuasion is positive for the receiver: $\mathbb{E}[V^R(h, m^T) - V^R(h, d)] \geq 0$.

2. Assume the persuader can propose the true model, $m^T \in M$, and $\Pr(h|m^T, \mu_0) \geq \Pr(h|d, \mu_0)$. The framing component of persuasion is positive for the persuader and negative for the receiver: $\mathbb{E}[V^j(h, m(h)) - V^j(h, m^T)]$ is positive for $j = S$, negative for $j = R$, and strictly positive for $j = S$ if and only if it is strictly negative for $j = R$.

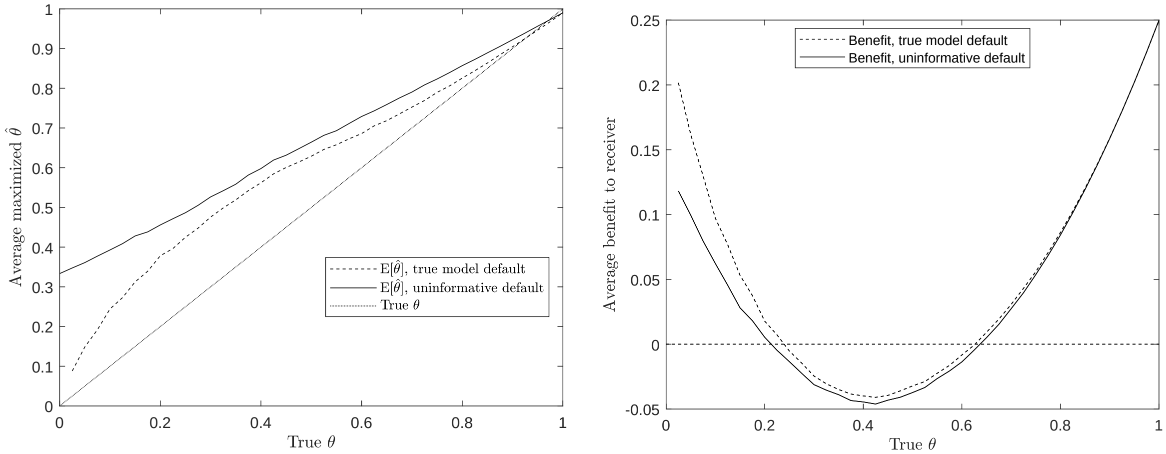
The first property is that the information component of persuasion is positive for the receiver: the receiver clearly cannot be made worse off on average by using the true model instead of the default model. The second property is that the framing component of persuasion is positive for the persuader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

C Highlighting Strips and Characteristics

C.1 Highlighting Strips of Data

This section analyzes the highlighting strips of data example. Recall that in the example the coin is flipped t times, where it yields heads with probability θ. While θ is drawn once and for all at the beginning of time from a density ψ, the persuader can propose models of the form "the last K periods are relevant for whether the coin comes up heads" for $K \geq 1$. We denote the receiver's posterior expectation of the probability of heads as $\hat{\theta}$. In our simulations, we pick a value of the true θ, and draw $t = 100$ random coin flips where the probability of heads is θ. We then find the optimal model for the persuader to propose, subject to the "truthteller constraint": the persuader's model must be more compelling than the truth. We use the "tc" superscript as short-hand for truthteller constrained. Finally, we compute the receiver's post-persuasion beliefs assuming $\psi \sim U[0,1]$. We run 5,000 simulations and report statistics aggregating across those simulations.

The left panel of Figure A.1 shows the receiver's average post-persuasion beliefs as a function of the true probability of heads θ. It draws a curve depicting the situation where the receiver has an uninformative default, so that he believes anything the persuader says, as well as a curve depicting the situation where the receiver's default is the true model. When the true model is the default, the receiver's post-persuasion beliefs are lower. The truthteller constraint prevents the persuader from proposing models that focus on very short favorable sequences. This reduces the scope for persuasion, particularly for low values of the true probability of heads θ. However, the figure shows that the scope for persuasion remains substantial, particularly for intermediate values of θ. Intuitively, there is always positive probability of a history with a long string of tails followed by a long string of heads, i.e., $(0 \ldots, 0, 1, \ldots 1)$. As an example, a politician can point to their recent "momentum" and thus limit voters' attention to a window of recent polls. In expectation, this increases voters'



(a) Beliefs

(b) Benefit from Persuasion

Figure A.1: Simulated Impact of Persuasion on Beliefs and Welfare when the Receiver's Default Model is the Truth

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

assessment of the politician's likelihood of winning. Similarly, a mutual fund company will choose to advertise with frames such as "be bullish" that emphasize past performance only when that past performance boosts investors' beliefs that future returns will be high (Mullainathan, Schwartzstein, and Shleifer 2008; Phillips, Pukthuanthong, and Rau, 2016; Koehler and Mercer, 2009).

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

Two other patterns from the left panel of the figure are worth noting. First, persuaders are constrained in the beliefs they can induce: on average, the receiver's estimate $\hat{\theta}$ is increasing in the true θ because it influences the expected number of heads in the history. A politician with a greater chance of winning will on average be more successful at increasing voters' assessments

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

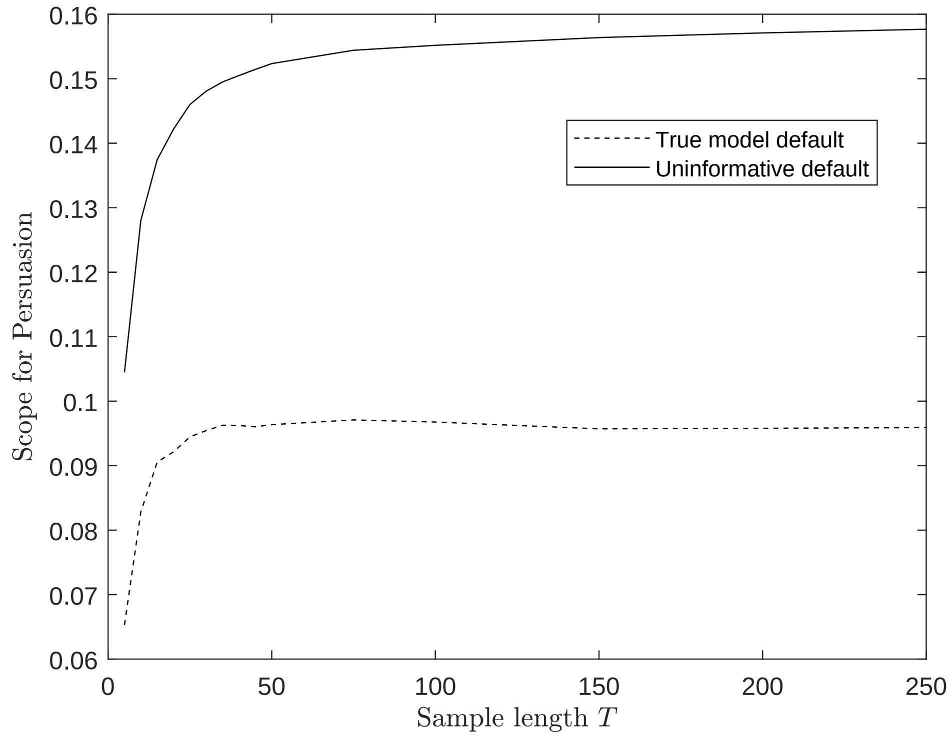


Figure A.2: Simulated Impact of the Truthteller Constraint on the Scope for Persuasion
This figure presents results on the impact of persuasion from simulations of the coin-flipping example for the case where $\psi \sim U[0,1]$ and $K \geq 1$. For each sample length, we compute the average over θ of the difference between the receiver's average post-persuasion beliefs and beliefs under the true model.

of her likelihood of winning. Similarly, a mutual fund with past successes will on average be more successful at increasing investors' assessment of future returns. Second, persuasion tends to attenuate the relationship between people's beliefs and the truth. For example, by inflating all political candidates' perceived chances of winning, persuasion reduces the average reaction of perceptions to reality.

We next study how the impact of persuasion varies with sample size. Figure A.2 plots the average difference between the receiver's post-persuasion beliefs and beliefs under the true model, $\mathbb{E}_\psi[\hat{\theta}] - \mathbb{E}_\psi[\theta]$, as a function of the length of the sample. Strikingly, we see that additional data actually *benefits* the persuader at the expense of the receiver. The intuition is that more data gives the persuader flexibility to propose compelling models that highlight favorable sequences—that is, to propose models that are beneficial to the persuader and overfit the historical data.[2] For

[2]On the other hand, the impact of persuasion does not go up with the data if the number of models the receiver is willing to consider decreases or stays the same as the amount of data increases. For example, the effectiveness of persuasion weakly decreases if the persuader can only choose between models that throw out the first 1, 2, or 3 flips. In this case, the impact of persuasion will go away for large t since the impact of the first three flips will become negligible. Such restrictions seem less plausible than the setting we consider.

instance, for political candidates, a longer history in the public eye is both a blessing and a curse. The candidate has a larger set of positives to highlight, but their opponent also has a larger set of potential negatives to highlight.

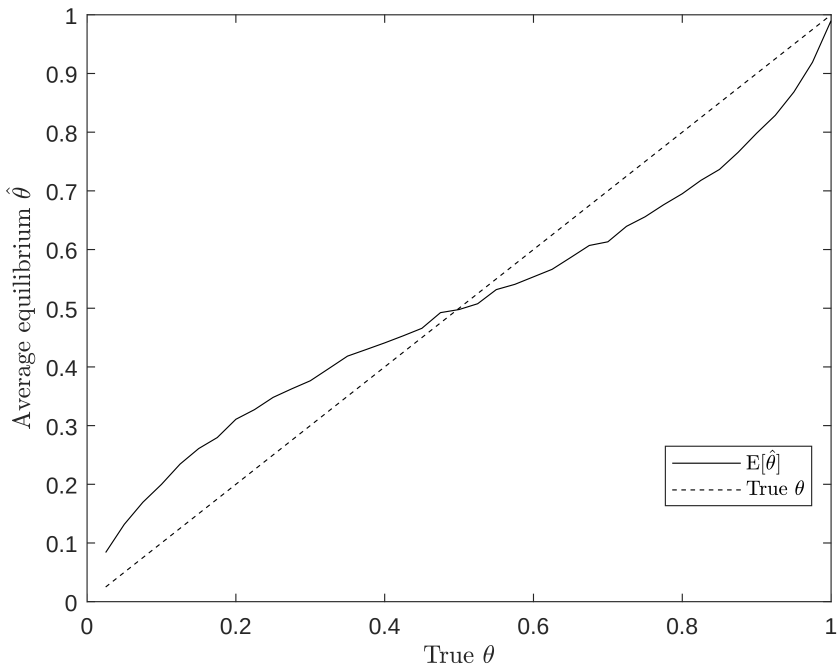


Figure A.3: Simulated Equilibrium with Competition
This figure presents results on the impact of competition in persuasion from simulations of the coin-flipping example for the case where $\psi \sim U[0,1]$ and $K \geq 1$. For each of 40 values of θ, we plot the average equilibrium post-persuasion beliefs of the receiver over $5,000$ sample paths, each of length 100.

Finally, we analyze the effect of competition between persuaders in this setting. We assume that there are two persuaders, one trying to persuader the receiver that θ is high and one trying to persuade the receiver that θ is low. Each persuader proposes a model after observing the history. The receiver then adopts the best-fitting model and updates her beliefs using Bayes rule. In the simulations, we again pick a value of the true θ, and draw $t = 100$ random coin flips where the probability of heads is θ. We then compute the receiver's equilibrium beliefs after the optimistic and pessimistic persuaders propose their optimal models. Figure A.3 reports statistics aggregating across 5,000 simulations. The figure shows that competition pushes equilibrium beliefs towards the receiver's prior of 0.5. For low values of θ, competition increases equilibrium beliefs relative to the single persuader case in Figure A.1. For high values of θ, competition reduces equilibrium beliefs. However, in contrast to Proposition 4 in the main text where we assumed that receivers were maximally open to persuasion, in this setting with a limited model space competition does not push beliefs fully to the receiver's prior.

C.2 Highlighting Characteristics

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

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

We write $c_m(x)$ to denote the element of the partition that contains x under model m. We assume the persuader can always propose the finest partition, where each x is in its own cell. To illustrate, if each x is described as a vector of attributes, $x = (x_1, x_2, \ldots, x_J)$, and m is a model where only the first three attributes are relevant to success, then $c_m(\tilde{x}) = \{x \in X : (x_1, x_2, x_3) = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)\}$. If the receiver adopts model m, the success probability he ascribes to each element of the partition is based on the number of successes in the data within that element of the partition. Formally, let $k_m(x, h)$ denote the number of times an element in $c_m(x)$ appears in the history h, and $s_m(x, h)$ denote the number of times an element in $c_m(x)$ appears in h as a success rather than a failure. If $\psi = \text{Uniform}[0, 1]$, by the Beta-Binomial updating formula the probability the receiver attaches to the actor being successful given model m is $\hat{\theta}(x) \equiv \mathbb{E}[\theta(x)|m, h] = (s_m(x, h)+1)/(k_m(x, h)+2)$.

D When Are Receivers Persuadable? Details

Recall from Section II that 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.

We illustrated the fourth point in Section II. To illustrate each of the first three points, we make use of the following definition: We say persuasion *is ineffective at history h given default d* when $\Pr(h|d, \mu_0) > \Pr(h|m, \mu_0)$ for all $m \in M$ satisfying $V^S(h, m) > V^S(h, d)$. That is, persuasion is ineffective when the persuader is unable to convince the receiver of any interpretation of the data more favorable to the persuader than the receiver's default interpretation.

The first factor affecting persuadability is how well the receiver's default model fits the history. When receivers' defaults fit the data well, they are hard to persuade. Formally, holding fixed history h, consider defaults d and d' such that d' fits the history better but induces the same posterior: $\Pr(h|d', \mu_0) > \Pr(h|d, \mu_0)$ and $\mu(h, d') = \mu(h, d)$. If persuasion is ineffective at history h given default d, then it is also ineffective at h given default d'. When receivers have defaults that overfit the data, they are hard to persuade. For instance, academics and benevolent financial advisers have a hard time convincing individual investors that stock returns are unpredictable because individual investors falsely perceive patterns in stock prices.

The analysis is similar across histories: there is less scope for persuasion under histories that fit the default better. Under conditions we will make precise below, if persuasion is ineffective at history h given default d, then it is also ineffective at any history $\tilde{h}$ given default $\tilde{d}$ that induces the same beliefs and fits better: $\mu(\tilde{h}, \tilde{d}) = \mu(h, d)$ and $\Pr(\tilde{h}|\tilde{d}, \mu_0) > \Pr(h|d, \mu_0)$. For instance, receivers are more persuadable that an abnormally cold month signals a hiatus in global warming than that an abnormally warm month signals a hiatus. A cold month poorly fits the default model that global warming is taking place, creating space for the persuader to propose an alternative.

Proposition A.1. *Suppose persuasion is ineffective at history h given default d and prior μ_0.*

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

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

This proposition formalizes the two ways that increasing the fit of the receivers' default reduces how persuadable they are. With a bit more structure, similar intuitions also apply if we modify the prior to change how well the same default interpretation fits the data. When the data and default interpretation imply something that the receiver viewed as ex ante unlikely, there is more space for persuasion. For instance, it is easier to persuade a voter that a bad gaffe by the candidate is meaningless if the voter ex ante believed the candidate to be competent than if the voter thought the candidate was incompetent.[3]

A second factor affecting persuadability is the expected (ex ante, prior to h being realized) difficulty receivers will have explaining the data under their default interpretation.[4] Receivers find technical analysis compelling in interpreting prices and trading volumes in financial markets; they would not when explaining patterns in their bank-account balances.

Receivers are also less persuadable when the data is less open to interpretation. If persuasion is ineffective at history h and default d given model space M, then it is ineffective for any $M' \subset M$. The receiver's openness to different models for interpreting data creates space for misleading persuasion. When signals have a natural interpretation, the persuader cannot do much to change minds; vague signals (Olszewski 2018), on the other hand, are ripe to be framed.

E Further Extensions, Robustness, and Examples

E.1 Generalizations and Refinements Capturing Receiver Knowledge

Consider the situations described in Section I.C where we think of the receiver's utility as being over actions and outcomes rather than over actions and latent states of the world. As we emphasize in that section, we assume the persuader cannot propose models that directly alter the prior distribution over states or the relationship between states and payoff-relevant outcomes.

One way of viewing this restriction is to broaden our conception of a model to be a joint distribution over outcomes y, histories h, and states ω: that is, a model m indexes a joint distribution $\Pr_m(y, h, \omega)$. With this broader conceptualization, our assumptions are:[5]

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

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

[5]For simplicity, this discussion assumes spaces are discrete.

1. Any model the persuader proposes must be consistent with the prior over states: for all $m \in M$ and $\omega \in \Omega$, $\sum_{y',h'} \Pr_m(y', h', \omega) = \mu_0(\omega)$.

2. Any model the persuader proposes must be consistent with the relationship between outcomes and latent states of the world: for all $m \in M, y \in Y$, and $\omega \in \Omega$,

$$\frac{\sum_{h'} \Pr_m(y, h', \omega)}{\sum_{y',h'} \Pr_m(y', h', \omega)} = f(y|\omega).$$

The default model is $\Pr_d(y, h, \omega) = \Pr_d(y|h, \omega) \cdot \Pr_d(h|\omega) \cdot \mu_0(\omega) = f(y|\omega) \cdot \pi_d(h|\omega) \cdot \mu_0(\omega)$.Writing our assumptions this way suggests generalizations and refinements of our approach. We first consider generalizations that relax (1) and (2), and then consider further refinements.

E.1.1 Two Generalizations that Highlight The Role of Prior Knowledge

It is crucial for our results that the receiver has prior knowledge that the persuader cannot influence without referencing data. Relaxing the first assumption of the previous section, imagine that the persuader could propose a prior over states in addition to proposing a likelihood function: under this alternative formulation, a model m indexes a (prior, likelihood function) combination (μ_m, π_m). In this case the persuader can always implement any beliefs she likes if the receiver is maximally open to persuasion:

Proposition A.2. *Suppose the receiver is maximally open to persuasion and consider the alternative formulation where the persuader is able to propose a prior over Ω as well as a likelihood function, so a model $m \in M$ indexes (μ_m, π_m). Fixing d, μ_0, and h, the persuader is able to induce any target belief $\tilde{\mu} \in \Delta(\Omega)$.*

The idea behind this result is that, for any desired belief $\tilde{\mu}$, the persuader is always able to propose a model that says: "the true prior belief over states is $\tilde{\mu}$ and what you just saw was inevitable". This model fits the data perfectly and induces the persuader's prior belief. Under this alternative formulation, there is no tradeoff between fit and movement.

This is not what we consider model persuasion to be. With model persuasion, a persuader influences the receiver's beliefs by proposing a model to make sense of data. Under this alternative formulation, the persuader is able to influence the receiver's beliefs by merely asserting the receiver should hold those beliefs. In the context of a jury trial, for example, this alternative formulation would allow the mere claim that the defendant is guilty to be persuasive, even if the prosecutor does not reference the evidence. Of course, there may be situations described by this alternative form of persuasion. It may operate in situations where the receiver has not given a problem enough thought to truly have prior beliefs, for example in the context of certain laboratory experiments.

However, in the large number of situations like the ones we focus on in this paper where receivers do have prior beliefs, we think it is more realistic to assume that the persuader cannot move those beliefs without referencing data.

In situations where the receiver's utility is over actions and outcomes, the persuader is similarly unconstrained if we relax the second assumption of the previous section and imagine the receiver is willing to entertain any mapping between states and outcomes, $f(y|\omega)$. In such situations what the persuader ultimately cares about is the receiver's ultimate belief in the distribution over Y because this is what determines the receiver's action. Given a target distribution μ^y, the persuader could get the receiver to believe in this distribution by proposing a model that implies $f_m(y|\omega) = \mu^y(y)$ for all $\omega \in \Omega$ and $y \in Y$. For example, an entrepreneur could argue: "it does not matter how good I am, all startups will for sure perform well next year". Under this alternative formulation, again there is no tradeoff between fit and movement. Again, we believe that this relaxation of our assumptions would not capture the essence of model persuasion: it abstracts from the idea that people find models compelling when they help make sense of data. While persuaders sometimes just assert that something is true, they often point to data to make their case.

E.1.2 Further Refinements

The broader conceptualization of a model above also suggests principled ways of restricting the model space M to reflect the receiver's knowledge and/or to require models to be internally consistent. To take some examples:

- Refinements might reflect *knowledge about the distribution over observables* h. For example, the receiver might be certain about the likelihood that an entrepreneur's first startup will be successful. At the extreme where the receiver is certain that the distribution over h is given by $q(\cdot)$, a refinement would be to require that for all $m \in M$ (including the default) and $h \in H$, $\sum_{y',\omega'} \Pr_m(y', h, \omega') = q(h)$. The fit constraint in this extreme case amounts to requiring that a receiver adopts the persuader's model only if it is equally compelling to the default model.

- Refinements might reflect *internal consistency requirements.* The receiver might have some knowledge that aspects of h must be related to the state in the same way. For example, the receiver might know that if drinking could give a person a heart attack at age 45 it cannot be good for them at age 47. Letting h_1 and h_2 be two components of h that the receiver is certain relates to the outcome in the same way, for example, the persuader might be restricted to models implying $\Pr_m(y|h_1, \omega) = \Pr_m(y|h_2, \omega)$ for all $y \in Y, \omega \in \Omega$.

- Refinements might reflect *knowledge about the true model.* The receiver might have some knowledge of how aspects of h are in fact related to the state. We saw examples of such a restriction, for example, in Section II where we discussed data that is closed to interpretation.

A single persuader is not constrained by refinements of the first form if receivers are otherwise maximally open to persuasion:

Proposition A.3. *Fix d, μ_0, and h and assume the receiver is willing to entertain a model m if and only if it satisfies $\Pr(h|m, \mu_0) = \Pr(h|d, \mu_0)$ for all $h \in H$. The persuader is able to induce target belief $\tilde{\mu} \in \Delta(\Omega)$ if and only if inequality* (2) *holds.*

This result, which follows immediately from the proof of Proposition 1, shows that a single persuader is not constrained per se by receiver knowledge about the distribution over observables. The receiver might know, for example, the likelihood that an entrepreneur's first startup will be successful. In the extreme where receivers are certain about this distribution, the persuader is constrained to propose models that are equally compelling to the default interpretation. While this restriction abstracts from the crucial psychology that receivers find models compelling when they *better* make sense of the data than the default interpretation, the behavioral implications are the same with a single persuader when receivers are maximally open to persuasion. Intuitively, needing to explain the data with an equally compelling model is less constraining to the persuader than needing to explain the data with a more compelling model. With competition, matters are more complicated because the analysis hinges a lot on how receivers are assumed to break ties between equally compelling models.

Persuaders, however, are constrained per se by receiver knowledge about the relationship between observables and payoff-relevant outcomes. For example, the persuader is constrained by the receiver knowing (or thinking he knows) that the success of an entrepreneur's first startup is positively correlated with the success of her second startup.

We leave a fuller analysis of such refinements for future work.

E.2 Receiver Skepticism

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

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

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

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

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

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

Proposition A.4. *Fix d, μ_0, and h and suppose the receiver is σ-skeptical. There is an M under which the persuader is able to induce target belief $\tilde{\mu} \in \Delta(\Omega)$ if and only if*

$$\tilde{\mu}(\omega) \leq \frac{\mu_0(\omega)}{\Pr(h|d,\mu_0)} \forall \omega \in \Omega \textit{ and } V^S(m(\tilde{\mu}),h) < V^S(d(h),h) \tag{A.1}$$

or

$$\tilde{\mu}(\omega) \leq \frac{\mu_0(\omega)}{\Pr(h|d,\mu_0)} \cdot (1-\sigma)\ \forall \omega \in \Omega \textit{ and } V^S(m(\tilde{\mu}),h) \geq V^S(d(h),h), \tag{A.2}$$

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

This result collapses to Proposition 1 when the receiver is not skeptical ($\sigma = 0$). Greater receiver skepticism ($\sigma > 0$) places restrictions on beliefs the persuader is able to induce. When receiver skepticism is sufficiently large ($\sigma > 1 - \Pr(h|d,\mu_0)$), the receiver will never adopt the persuader's proposed model when it is known to be in the persuader's interest. When receiver skepticism is slightly smaller ($\sigma \approx 1 - \Pr(h|d,\mu_0)$), the receiver will only adopt a model that is known to be in the persuader's interest when it induces beliefs that are close to the receiver's prior. Indeed, when $\sigma = 1 - \Pr(h|d,\mu_0)$, then the only beliefs that satisfy (A.2) are $\tilde{\mu} = \mu_0$.

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

E.3 Gathering, Revealing and Framing Data

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

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

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

Given these assumptions, Corollary 1 implies that the persuader's payoff for fixed likelihood function π, prior μ_0, and ambiguous data h is

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

This implies that the persuader's expected payoff is

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

given π.

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

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

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

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

$$f\left(|H| \cdot \mu_0(1)\right), \tag{A.3}$$

which is implemented by an ambiguous data process π that features $\pi(h|\omega) = 1/|H|$ for all h, ω.[6] *So the persuader maximizes his payoff by collecting ambiguous data that is completely uninformative.* Noting that (A.3) is increasing in $|H|$, we see that *the persuader maximizes his payoff by collecting and reporting as much of this uninformative, ambiguous data as possible.* Finally, noting that (A.3) is decreasing in mean-preserving spreads of $\mu_0(1)$ since f is concave, we see that *the persuader does not want to collect and reveal any unambiguous information.*

The intuition behind why the persuader wants to collect and reveal completely uninformative ambiguous information is that she is (weakly) risk averse and eliminates the risk of having difficulty framing the information by making it uninformative.[7] This is also the intuition for why the persuader, who is assumed to have no prior informational advantage over the receiver about ω, does not want to collect and reveal any unambiguous information. The intuition for why the persuader wants to collect and reveal as much uninformative ambiguous information as possible is that this maximizes the wiggle room the persuader has to frame the information.

F Dynamics

This appendix considers three ways to extend our analysis to dynamic environments and establishes preliminary results. Unless otherwise noted, this appendix analyzes a simple two-period setting: (1) there is a first signal h_1, the persuader proposes a model m_1; (2) there is a second signal h_2, the persuader proposes another model m_2, and finally the receiver makes a decision. In particular, the receiver only makes a single decision even in this dynamic setting.

F.1 Three Ways to Extend Our Analysis to Dynamic Environments

There are at least three ways to extend our framework to such a dynamic environment. We start by giving intuitive descriptions of these three extensions and then provide more formal definitions.

Under *prior dynamics*, the model the persuader proposes in the first period influences the receiver's prior—and thus the models she will find compelling in the second period. If the persuader's model in the first period convinces the receiver that an entrepreneur is likely of high

[6] This π is not necessarily the unique solution to the maximization problem. But note that this π also minimizes the expected value of $\Pr(h|m^T, \mu_0)$—creating the most expected space for persuasion—no matter μ_0.

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

quality, then in the second period the receiver will favor models that explain why the most recent data is expected if the entrepreneur is of high quality.

Under *consistency dynamics*, the model the persuader proposes in the second period has to be consistent with the model the persuader proposed in the first period. For example, the persuader cannot propose a first period model suggesting that first-period signal h_1 indicates a low quality entrepreneur and a second period model suggesting that the same signal indicates a high quality entrepreneur.

Under *default dynamics*, the model the persuader proposes in the first period becomes the receiver's default model going into the second period. If the persuader gets the receiver to believe in a model that explains the entrepreneur's initial failure quite well, then the receiver will only find a future proposed model compelling if it explains the entrepreneur's initial failure even better.

More formally, in each dynamic specification the receiver adopts the persuader's first model, m_1, if and only if it fits first-period data h_1 as well as the default: $\Pr(h_1|m_1, \mu_0) \geq \Pr(h_1|d, \mu_0)$. The three forms of dynamics we consider then differ in how m_1 constrains the persuader in the second period:

1. *Prior dynamics*. The persuader's first model, m_1, influences the receiver's prior, so the receiver's prior in the second period becomes $\mu(h_1, m_1)$. The receiver then adopts the persuader's second model if and only if $\Pr(h_2|m_2, \mu(h_1, m_1)) \geq \Pr(h_2|d, \mu(h_1, m_1))$. Note that in this setup a model only specifies how to interpret a single signal and the default is independent of previous models proposed by the persuader.

2. *Consistency dynamics*. The persuader always has to propose models that are consistent with previous statements: model m_2 must satisfy $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega) = \pi_{m_1}(h_1|\omega)$ for all $\omega \in \Omega$. The receiver adopts the persuader's model in the second period only if $\Pr(h_2, h_1|m_2, \mu_0) \geq \Pr(h_2, h_1|d, \mu_0)$. Note that, in this setup as in prior dynamics, a model only specifies how to interpret signals that have arrived so far. Unlike prior dynamics, there are default interpretations of both single and multiple signals, and the second-period model interprets all previous signals.[8]

3. *Default dynamics.* The persuader's first model, m_1, becomes the receiver's default interpretation. The receiver adopts the persuader's second model only if $\Pr(h_2, h_1|m_2, \mu_0) \geq \Pr(h_2, h_1|m_1, \mu_0)$. Note that in this setup a model specifies how to interpret both the first and second signals, whether or not it is proposed in the first or second period.

We think there are realistic features of each specification: the lens through which receivers evaluate models over time is probably influenced by previous models they adopted as in prior dynamics; re-

[8]Note that prior dynamics also embeds a kind of consistency constraint by updating priors to be consistent with first-period models and having second-period models only explain second-period data.

ceivers may penalize persuaders for proposing a model today that is inconsistent with a model they proposed before as in consistency dynamics; and receivers' default interpretations are probably influenced by previous models they found compelling as in default dynamics.

As a first step to analyzing dynamics, we examine each of these three cases. Future work could more carefully examine combinations of these approaches, or the factors that influence which approach most realistically describes different situations.

F.2 A Simple Example to Build Intuition

We first show how these three specifications work in a single example: one where an entrepreneur seeks investment for a third startup before and after information on the success of his second startup arrives. As in the example throughout the paper, we imagine there is a success probability θ that governs the success of the entrepreneur's third startup and receivers have a uniform prior $U[0,1]$ over this probability. The default model is the true model. The persuader wants the receiver's belief in the success probability to be as large as possible. In the static framework of the main paper, it was not important that the persuader know the true model in this example. In these dynamic setups, we assume the persuader understands that the true expected value of θ is increasing in the empirical success frequency.

Listing the most recent outcome first (i.e., ordering (h_2, h_1)), we compute the outcome for each dynamic specification and the static setup in each of four scenarios (i) (failure, failure), (ii) (success, failure), (iii) (failure, success), and (iv) (success, success). We present the full calculations for scenario (i) and then consolidate the rest of the results into a table for brevity.

Scenario (i) (failure, failure)

Static prediction: We have that $\Pr(\text{failure, failure}|d, \mu_0) = \int (1-\theta)^2 d\theta = 1/3$. As a result, the best the persuader is able to do is to propose the model: "entrepreneurs in the top $1/3$ of future success probability fail the first two times; all other entrepreneurs have an initial success".[9] That is, the persuader proposes a model with

$$\pi(\text{failure, failure}|\theta) = \begin{cases} 1 & \text{if } \theta \geq 2/3 \\ 0 & \text{otherwise.} \end{cases}$$

This model yields posterior expectation $\hat{\theta}^{\text{static}}(\text{failure, failure}) = 5/6$.

[9]Footnote 25 shows that such a cutoff model is optimal.

Prior dynamics: In the second period, the persuader must propose a model m_2 that satisfies

$$\Pr(\text{failure}|m_2, \mu(\text{failure}, m_1)) \geq \Pr(\text{failure}|d, \mu(\text{failure}, m_1)) = \int (1-\theta) d\mu(\text{failure}, m_1) \equiv k(\text{failure, failure}).$$

The persuader will propose the following model: "the top k(failure, failure) of entrepreneurs fail their second startup; all other entrepreneurs succeed with their second startup." This model yields posterior expectation

$$\hat{\theta}^{\text{prior dynamics}}(\text{failure, failure}) = 1 - k(\text{failure, failure})/2.$$

Knowing this, in the first period the persuader proposes a model that maximizes

$$\mathbb{E}[\theta|\text{failure}] \cdot (1 - k(\text{success, failure})/2) + (1 - \mathbb{E}[\theta|\text{failure}]) \cdot (1 - k(\text{failure, failure})/2)\,,$$

where $\mathbb{E}[\theta|\text{failure}] = 1/3$ and $k(\text{success, failure}) = \Pr(\text{success}|d, \mu(\text{failure}, m_1)) = \int \theta d\mu(\text{failure}, m_1)$. That is, the persuader proposes an m_1 that solves

$$\max_{m_1} 1/3 \cdot (1 - k(\text{success, failure}|m_1)/2) + 2/3 \cdot (1 - k(\text{failure, failure}|m_1)/2)\,.$$

This is equivalent to solving:

$$\min_{m_1} k(\text{success, failure}|m_1) + 2k(\text{failure, failure}|m_1) = \min_{m_1} \int \theta d\mu(\text{failure}, m_1) + 2\int (1-\theta) d\mu(\text{failure}, m_1),$$

which in turn is equivalent to maximizing $\int \theta d\mu(\text{failure}, m_1)$. This means that in the first period, the persuader proposes the model

$$\pi_{m_1}(\text{failure}|\theta) = \begin{cases} 1 & \text{if } \theta \geq 1/2 \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the persuader first proposes model m_1 that the top half of entrepreneurs fail in their first startup and then proposes the model m_2 that the top quarter of entrepreneurs fail in their first two startups. Entrepreneurs and venture capitalists often use these kinds of models, arguing that lessons from early failures are the key to success.[10] So $k(\text{failure, failure}|m_1) = 2\int_{1/2}^{1}(1-\theta)d\theta =$

[10] For instance, Microsoft co-founders Bill Gates and Paul Allen reflect on their career: "Even though [our previous firm] wasn't a roaring success, it was seminal in preparing us to make Microsoft's first product a couple of years later." https://archive.fortune.com/magazines/fortune/fortune_archive/1995/10/02/206528/index.htm

$1/4$ and $k(\text{success, failure}|m_1) = 2\int_{1/2}^{1} \theta d\theta = 3/4$. This all means then that

$$\hat{\theta}^{\text{prior dynamics}}(\text{failure, failure}) = 1 - k(\text{failure, failure})/2 = 1 - 1/8 = 7/8.$$

Consistency dynamics: Model m_2 must satisfy $\sum_{h_2'} \pi_{m_2}(h_2', \text{failure}|\omega) = \pi_{m_1}(\text{failure}|\omega)$ for all $\omega \in \Omega$. The receiver adopts the persuader's model in the second period only if $\Pr(\text{failure, failure}|m_2, \mu_0) \geq \Pr(\text{failure, failure}|d, \mu_0)$. First note that consistency dynamics adds constraints to the persuader in the second period, relative to the static setup. Thus, she can at best do as well as she does in the static setup. The question is thus: can she do exactly as well as in the static problem? We show that she indeed does as well in this case by constructing models that deliver the same receiver's posterior as in the static case:

$$\pi_{m_1}(\text{failure}|\theta) = \begin{cases} 1 & \text{if } \theta \geq 1/2 \\ 0 & \text{otherwise} \end{cases}$$

$$\pi_{m_2}(\text{failure, failure}|\theta) = \begin{cases} 1 & \text{if } \theta \geq 2/3 \\ 0 & \text{otherwise} \end{cases}$$

$$\pi_{m_2}(\text{success, failure}|\theta) = \begin{cases} 1 & \text{if } \theta \in [1/2, 2/3) \\ 0 & \text{otherwise.} \end{cases}$$

We have $\sum_{h_2'} \pi_{m_2}(h_2', \text{failure}|\theta) = \pi_{m_1}(\text{failure}|\theta)$, and since $\pi_{m_2}(\text{failure, failure}|\theta)$ is the same as the static case, the persuader does exactly as well: $\hat{\theta}^{\text{consistency dynamics}}(\text{failure, failure}) = 5/6$.

Default dynamics: In the second period, the persuader needs to propose a model m_2 that satisfies:[11]

$$\Pr(\text{failure, failure}|m_2, \mu_0) \geq \Pr(\text{failure, failure}|m_1, \mu_0) \equiv k(\text{failure, failure}).$$

The persuader then will propose the following model: "the top $k(\text{failure, failure})$ of entrepreneurs fail their first two startups; all other entrepreneurs succeed with one of their first two startups." This model yields posterior expectation

$$\hat{\theta}^{\text{default dynamics}}(\text{failure, failure}) = 1 - k(\text{failure, failure})/2.$$

Knowing this, in the first period the persuader proposes a model that maximizes

$$\mathbb{E}[\theta|\text{failure}] \cdot (1 - k(\text{success, failure})/2) + (1 - \mathbb{E}[\theta|\text{failure}]) \cdot (1 - k(\text{failure, failure})/2),$$

[11] As a preliminary observation, the persuader could always propose d in the first period, so in expectation the persuader does better by proposing m_1 than she would do in the static model.

Table A.1: Summarizing What the Three Dynamic Specifications Imply in the Example

	(f,f)	(s,f)	(f,s)	(s,s)
Bayesian Benchmark	1/4	1/2	1/2	3/4
Static	5/6	11/12	11/12	5/6
Prior Dynamics (m_1 influences μ_1)	7/8	5/8	5/8	7/8
Consistency Dynamics (m_2 consistent with m_1)	5/6	11/12	11/12	5/6
Default Dynamics (m_1 becomes default)	1	3/4	3/4	1

This table shows the receiver's posterior expectation of success in the example for each dynamic specification following every history. The table uses "f" as shorthand for "failure" and "s" as shorthand for "success".

where $\mathbb{E}[\theta|\text{failure}] = 1/3$ and $k(\text{success, failure}) = \Pr(\text{success, failure}|m_1, \mu_0)$. That is, the persuader proposes an m_1 that solves

$$\max_{m_1} 1/3\,(1 - k(\text{success, failure}|m_1)/2) + 2/3 \cdot (1 - k(\text{failure, failure}|m_1)/2)\,,$$

which is equivalent to solving:

$$\min_{m_1} k(\text{success, failure}|m_1) + 2 \cdot k(\text{failure, failure}|m_1),$$

subject to $k(\text{success, failure}|m_1)+k(\text{failure, failure}|m_1) \geq 1/2$. The solution is $k(\text{failure, failure}|m_1) = 0$ and $k(\text{success, failure}|m_1) = 1/2$. This all means then that

$$\hat{\theta}^{\text{default dynamics}}(\text{failure, failure}) = 1 - k(\text{failure, failure})/2 = 1.$$

Summary Table

From the Summary Table (Table A.1), we see the following:

1. No matter the dynamic specification, the persuader does better than under the Bayesian benchmark in every contingency.

2. If a persuader does worse in some contingency under a dynamic specification than under the static specification, then there is another contingency where the persuader does better under the dynamic specification than under the static specification.

   (a) The way the persuader is able to do better in some contingencies under prior and default dynamics is to propose first-period models that will not, in expectation, fit well going forward; i.e., "opening minds" or "increasing puzzlement". Following a failure, the persuader suggests a model that fits a subsequent failure poorly; following a success, the persuader proposes a model that fits a subsequent success poorly. In this way, the persuader relaxes the fit constraint in expectation. This means that, unlike in the static model, with prior or default dynamics the persuader benefits more from consistent than inconsistent histories.

3. There is a sense in which needing to be consistent with previous statements does not constrain the persuader. The persuader does as well under consistency dynamics as she would in the static specification. Likewise, needing to propose models that fit even better than previously suggested models does not constrain the persuader: The persuader does better on average under default dynamics than under the static specification. On the other hand, it may be constraining for the persuader to have the model he proposes today influence the lens through which receivers evaluate models tomorrow: In this example, the persuader does worse on average under prior dynamics than under the static formulation of the model.

F.3 Some Lessons That Hold More Broadly

Continuing to consider the same dynamic setting, we generalize a sense in which dynamics need not be constraining to the persuader. When receivers are maximally open to persuasion, the persuader is neither constrained by needing to propose models that are consistent with previous models, nor by earlier proposed models becoming receivers' default models.

Proposition A.5. *Fix d as well as μ_0, and suppose the receiver is maximally open to persuasion. Let $\tilde{\mu}(h_2, h_1)$ be a mapping between possible values (h_2, h_1) and beliefs over Ω that are implementable in the static formulation of the model—i.e., such beliefs satisfy (2).*

1. *Under consistency dynamics, for every h_1 there exists a first-period model $m_1(h_1)$ such that for every h_2 the persuader is able to implement $\tilde{\mu}(h_2, h_1)$ with a model $m_2(h_2, h_1)$ that satisfies (i) $\sum_{h'_2} \pi_{m_2}(h'_2, h_1|\omega) = \pi_{m_1}(h_1|\omega)$ for all $\omega \in \Omega$ and (ii) $\Pr(h_2, h_1|m_2, \mu_0) \geq \Pr(h_2, h_1|d, \mu_0)$.*

2. *Under default dynamics, for every h_1 there exists a first-period model $m_1(h_1)$ such that for every h_2 the persuader is able to implement $\tilde{\mu}(h_2, h_1)$ with a model $m_2(h_2, h_1)$ that satisfies $\Pr(h_2, h_1|m_2, \mu_0) \geq \Pr(h_2, h_1|m_1, \mu_0)$.*

The first part means that, when the receiver is maximally open to persuasion, the persuader always does equally well under consistency dynamics as she does under the static formulation of the model. The reason for this is that, following h_1, the persuader is always able to say "h_1 was inevitable". So long as the persuader would be able to induce $\tilde{\mu}(h_2, h_2)$ if she didn't have the opportunity to say anything in the first period, she is also able to induce that belief in a way that's consistent with the earlier statement. To leave herself maximum flexibility, the persuader wants to tell the audience there is nothing to learn from the initial signal. Note that this is different from telling the audience to, e.g., stick with the default interpretation because this could in fact constrain the persuader under consistency dynamics: the persuader wants to move the audience's beliefs away from the default in the first period towards a model that says more data is needed to learn anything.

The second part means that the persuader on average does weakly better under default dynamics as she does under the static formulation of the model. The reason is that the persuader is always able to just supply the default model in the first period, which under default dynamics will make him as well off as under the static formulation of the model.

Conclusions under prior dynamics are more nuanced. In the example of the previous section, the persuader does worse on average under prior dynamics than under the static formulation of the model. However, there are also examples where the persuader does better on average under prior dynamics than he does under the static formulation of the model. As an illustration, suppose that the persuader is trying to convince an audience that a coin will come up heads on the third flip and the coin ex ante is an "always heads" coin with probability $.4$ and an "always tails" coin with probability $.6$. The default interpretation is that the same coin is flipped each time, so a single flip reveals whether the coin will come up heads. Under the static formulation of the model (as well as consistency dynamics), by Proposition 1 the persuader is able after two tails to convince the audience that the likelihood of heads on the third flip is $.4/.6 = 2/3$. Under prior dynamics, the persuader is able to do even better (again, invoking Proposition 1): after the first tails, the persuader is able to get the audience to believe the likelihood of "always heads" is $2/3$. This then means that, in the second period, $\Pr(\text{tails}, \text{tails}|d, \mu(h_1, m_1)) = 1/3$ and the persuader is able to get the audience to believe that the likelihood of heads on the third flip is $\min\{1, (2/3)/(1/3)\} = 1$.

An interesting avenue for future work is to consider what happens when the receiver takes actions each period. While we will say little here, we establish a basic result for consistency dynamics: The conclusion that the persuader does as well as under the static formulation of the model still holds in special cases of the receiver taking actions each period, such as where states are ordered and the persuader benefits from beliefs that are more positive in the sense of first-order stochastic dominance. As with the example of the previous section, the persuader could induce as positive beliefs as possible following h_1 in a way that's consistent with any attempt to induce as positive beliefs as possible following (h_2, h_1):

Proposition A.6. *Suppose states are ordered $\Omega = \{\omega_1, \ldots, \omega_N\}$ with $\omega_1 < \ldots < \omega_N$. Fix d as well as μ_0, and suppose the receiver is maximally open to persuasion. Let $\tilde{\mu}(h_2, h_1)$ be a mapping between possible values (h_2, h_1) and beliefs over Ω that are implementable in the static formula-*

tion of the model—i.e., such beliefs satisfy (2)—and first-order stochastically dominate any other beliefs over Ω that are implementable in the static formulation of the model given (h_2, h_1). For every h_1 there exists first-period model $m_1(h_1)$ that (i) implement beliefs that first-order stochastically dominate any other beliefs over Ω that are implementable given h_1 and (ii) allow that, for every h_2, the persuader is able to implement $\tilde{\mu}(h_2, h_1)$ with a model $m_2(h_2, h_1)$ that satisfies (ii.a) $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega) = \pi_{m_1}(h_1|\omega)$ for all $\omega \in \Omega$ and (ii.b) $\Pr(h_2, h_1|m_2, \mu_0) \geq \Pr(h_2, h_1|d, \mu_0)$.

The intuition for why this is true generally follows from the example above: interpreting (h_2, h_1) to induce beliefs that are as positive as possible will always be consistent with interpreting h_1 to induce beliefs that are as positive as possible.

F.4 Sequential Competition

This section uses an example to show how sequential competition may look different than static competition under certain dynamic specifications. In particular, we return to the jury example from Section VI.C and show that the order in which the defense and prosecution present arguments may influence jurors' ultimate decisions under prior dynamics. There is a first mover advantage because the party that moves first is able to preemptively frame evidence to its advantage.

Suppose evidence h is fixed and maximally open to interpretation, but arguments are sequential. Further, to limit the number of cases suppose that under the receiver's default interpretation h provides evidence neither for guilt nor innocence. To be able to directly port the definitions above with dynamics, we make the contrived assumption that "half of" h, h_1, arrives prior to the first argument and the other half, h_2, arrives prior to the second argument. In keeping with the idea that h_1 and h_2 each comprise "half of" h and are uninformative of guilt under the receiver's default interpretation, assume $\Pr(h_1|d, \mu_0) = \Pr(h_2|d, \mu_0)$ and $\Pr(h_1|d, \mu)$ is independent of μ. Under prior dynamics, the model the first-moving persuader proposes, m_1, influences the receiver's prior, so the receiver's prior in the second period becomes $\mu(h_1, m_1)$. The receiver adopts the second-moving persuader's model only if $\Pr(h_2|m_2, \mu(h_1, m_1)) \geq \Pr(h_2|d, \mu(h_1, m_1))$.

Imagine the jury's prior is to acquit the defendant: payoffs are such that the cutoff for acquittal is 1/2 and priors are such that $\mu_0(g) < 1/2$. The prosecution goes first in closing arguments and the defense goes second. Is there a model the prosecution can propose that induces $\mu(g|h_1, m_1) > 1/2$ and prevents the defense from framing the data so that the jury acquits?

Under many parameters, the answer is yes. The prosecution will select m_1 to induce the maximum possible belief in guilt, which by Proposition 1 is $\min\{\mu_0(g)/\Pr(h_1|d, \mu_0), 1\}$. If this equals 1, there is nothing the defense can then do to bring $\mu(g)$ back down below $1/2$. If it is below 1, then the defense moves the jury's belief that the defendant is not guilty, $\mu(ng)$, up to the maximum possible amount $\mu(ng|h_1, m_1)/\Pr(h_2|d, \mu(h_1, m_1)) = 1/\Pr(h_2|d, \mu_1) - \mu_0(g)/(\Pr(h_1|d, \mu_0) \times$

$\Pr(h_2|d, \mu_1))$ (again, by Proposition 1). This implies that

$$\mu(g|h, m_2) = \frac{\Pr(h_2|d, \mu_1) - 1}{\Pr(h_2|d, \mu_1)} + \frac{\mu_0(g)}{\Pr(h_1|d, \mu_0) \times \Pr(h_2|d, \mu_1)}.$$

This final belief is above $1/2$ if and only if

$$\mu_0(g) > \frac{\Pr(h_1|d, \mu_0)\,(2 - \Pr(h_2|d, \mu_1))}{2} = \frac{\Pr(h_1|d, \mu_0)\,(2 - \Pr(h_1|d, \mu_1))}{2} \equiv \tilde{\mu},$$

where the equality is a result of the assumption that h is uninformative about guilt under the default interpretation. Two implications follow. First, $\tilde{\mu} < 1/2$ whenever $\Pr(h_1|d, \mu_0) = \Pr(h_2|d, \mu_1) < 1$: the first mover has the advantage of influencing the lens through which the receiver evaluates the second-mover's model. Second, $\tilde{\mu}$ is increasing in $\Pr(h_1|d, \mu_0) = \Pr(h_2|d, \mu_1)$: the more surprising the signals are under the jury's default, the larger the first mover advantage for the prosecution.

G Examples: Details

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

G.1 Persuading an Investor: Technical Analysis

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

At date t, the state is $\theta_t \in \{0.25, 0.5, 0.75\}$, where θ_t is the probability AMZN stock rises at date $t+1$. The persuader frames the history h of returns from 1/8/2019 to 1/28/2019 to influence the receiver's posterior on $\theta_{1/29}$. Suppose the receiver's prior is evenly distributed across the three possible states: $\mu_0(\theta_{1/29} = 0.25) = \mu_0(\theta_{1/29} = 0.5) = \mu_0(\theta_{1/29} = 0.75) = 1/3$.

The default model is a history-dependent version of the random walk model—at each date, AMZN is equally likely to rise or fall:[12]

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

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

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

The model is formally:[13]

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

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

$$\Pr(h|RS, \mu_0) = \pi_{RS}(h|\theta = 0.75)\mu_0(\theta = 0.75) = (0.50)(0.75)^3(1/3) = 0.07.$$

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

$$\Pr(h|d, \mu_0) = \pi_d(h|\theta = 0.5)\mu_0(\theta = 0.5) = (0.50)^4(1/3) = 0.02.$$

Performing the analogous calculations on the actual AMZN price path from 1/8/2019 to 1/28/2019,

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

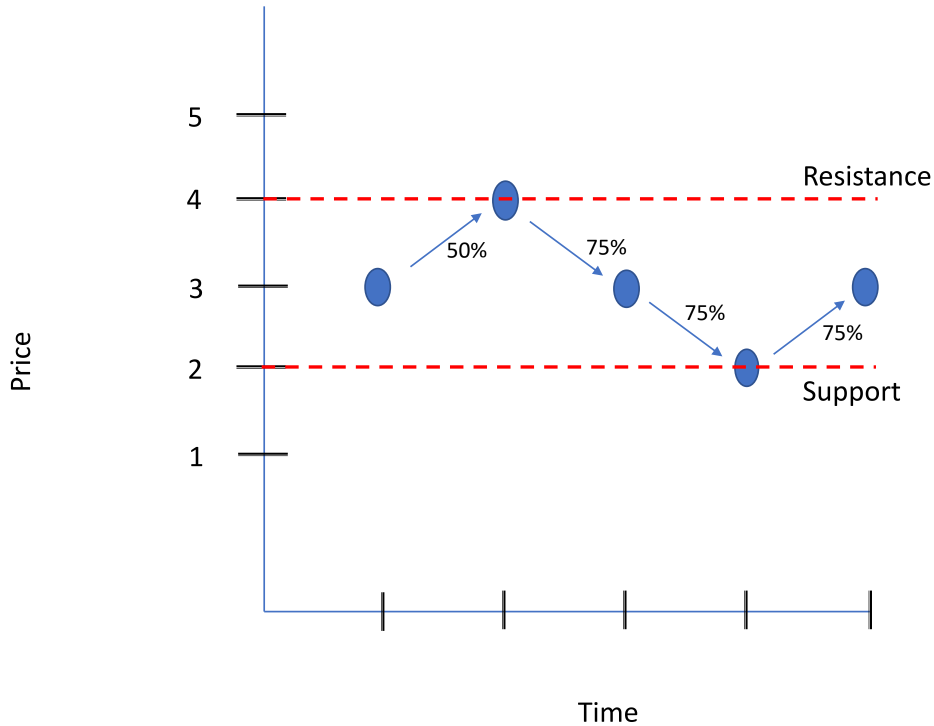


Figure A.4: Applying the Support and Resistance Model to a Stylized Price Path

the support-resistance model implies that AMZN is likely to rise, as AMZN had most recently hit its support.[14] The probability of the history is more than four times higher under the support-resistance model than the random walk model. Note that this means that even if we modified the setup so that the prior strongly favored the random walk model (e.g., $\mu(\theta = 0.5) = 2 \times \mu(\theta = 0.75)$), the receiver would still find the support-resistance model more compelling.

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

This section provides a somewhat more elaborate formulation of individual investment advice, relative to the one presented in Section II. Suppose there are N investments with investment j having characteristics $(x_1^j, \dots, x_K^j)$. Each investment will either be successful or not. If the investor's investment is successful he gets a payoff of $s > 0$ and he gets a payoff of 0 otherwise. The investor may pay a cost $\chi \in (0, 1)$ to make an "active" choice in a particular investment or to pay a cost $\chi_L \in [0, \chi)$ to invest "passively" in one of the N investments selected at random—for illustrative purposes, we assume the investor is risk neutral so that there is no diversification motive and

[14]Calculated based on data from the Daily 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.

randomly selecting one investment has the same expected utility as holding all N. We normalize $\chi_L = 0$, so the investor will want to make an active choice of a particular investment if and only if he thinks he is able to predict which investment will be successful to an extent that justifies the cost χ. The person's prior μ_0 is that the success probability of each investment is independently drawn from a uniform distribution over a finite set centered around $1/2$.

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

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

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

Continue to assume that receivers are maximally open to persuasion, but suppose that Passive acts as a strategic truthteller. In this case, a simple application of Proposition 3 shows that Passive will always convince investors to make a passive investment because Passive is at an advantage: she wants investors' beliefs to stay at the prior. But this is only an advantage if Passive is willing to propose the wrong model. In addition, Passive's advantage would disappear if the receiver's prior favored Active instead.

G.3 Persuading a Client: Good to Great

This section shows how "Good to Great" advice from Collins (2001) is compelling according to our framework. The setup is very similar to the entrepreneur problem. The underlying state is the expected (log) return for the portfolio of 11 good-to-great companies highlighted by Jim Collins. We assume the receiver's prior is that expected log returns are distributed $\rho \sim N(\bar{\rho}, \sigma_\rho^2)$. We observe realized returns, which are expected returns plus noise: $r_t = \rho + \varepsilon_{it}$ where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Let $\tau_\rho = 1/\sigma_\rho^2$ and $\tau_\varepsilon = 1/\sigma_\varepsilon^2$.

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

$$\begin{aligned}\Pr(\mathbf{r}_t|\rho,d,\mu_0) &= \prod_t \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left\{-\frac{(r_{it}-\rho)^2}{2\sigma_\epsilon^2}\right\}.\\ &= \frac{1}{(2\pi)^{T/2}}\frac{1}{\sigma_\varepsilon^T}\exp\left\{-\frac{T}{2\sigma_\varepsilon^2}s^2\right\}\exp\left\{-\frac{T}{2\sigma_\varepsilon^2}(\bar{r}-\rho)^2\right\}.\end{aligned}$$

The prior for ρ is given by

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

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

$$\Pr(\mathbf{r}_t|d,\mu_0)=\frac{1}{(2\pi)^{(T+1)/2}}\frac{1}{\sigma_\varepsilon^T}\frac{1}{\sigma_\rho}\exp\left\{-\frac{T}{2\sigma_\varepsilon^2}s^2\right\}\int\left[\exp\left\{-\frac{T}{2\sigma_\varepsilon^2}(\bar{r}-\rho)^2-\frac{(\rho-\bar{\rho})^2}{2\sigma_\rho^2}\right\}\right]d\rho.$$

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

$$\Pr(\mathbf{r}_t|d,\mu_0)=\frac{\sigma_\varepsilon}{(\sqrt{2\pi}\sigma_\varepsilon)^T\sqrt{T\sigma_\rho^2+\sigma_\varepsilon^2}}\exp\left(\frac{-\sum_t r_t^2}{2\sigma_\varepsilon^2}-\frac{\bar{\rho}^2}{2\sigma_\rho^2}\right)\exp\left(\frac{\frac{\sigma_\rho^2T^2\bar{r}^2}{\sigma_\varepsilon^2}+\frac{\sigma_\varepsilon^2\bar{\rho}^2}{\sigma_\mu^2}+2T\bar{r}\bar{\rho}}{2(T\sigma_\rho^2+\sigma_\varepsilon^2)}\right). \tag{A.4}$$

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

$$\begin{aligned}&\frac{\sigma_\varepsilon}{(\sqrt{2\pi}\sigma_\varepsilon)^L\sqrt{L\sigma_\rho^2+\sigma_\varepsilon^2}}\exp\left(\frac{-\sum_{t=1}^L r_t^2}{2\sigma_\varepsilon^2}-\frac{\bar{\rho}^2}{2\sigma_\rho^2}\right)\exp\left(\frac{\frac{\sigma_\rho^2L^2\bar{r}_{[1,L]}^2}{\sigma_\varepsilon^2}+\frac{\sigma_\varepsilon^2\bar{\rho}^2}{\sigma_\rho^2}+2L\bar{r}_{[1,L]}\bar{\rho}}{2(L\sigma_\rho^2+\sigma_\varepsilon^2)}\right)\times\\ &\qquad\frac{\sigma_\varepsilon}{(\sqrt{2\pi}\sigma_\varepsilon)^{T-L}\sqrt{(T-L)\sigma_\rho^2+\sigma_\varepsilon^2}}\exp\left(\frac{-\sum_{t=L+1}^T r_t^2}{2\sigma_\varepsilon^2}-\frac{\bar{\rho}^2}{2\sigma_\rho^2}\right)\times\\ &\qquad\qquad\exp\left(\frac{\frac{\sigma_\rho^2(T-L)^2\bar{r}_{[L+1,T]}^2}{\sigma_\varepsilon^2}+\frac{\sigma_\varepsilon^2\bar{\rho}^2}{\sigma_\rho^2}+2(T-L)\bar{r}_{[L+1,T]}\bar{\rho}}{2((T-L)\sigma_\rho^2+\sigma_\varepsilon^2)}\right).\end{aligned}$$

Table A.2: Stock Return Performance of Good-to-Great Firms

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

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

Applying these formulas to the actual annual return path of 11 firms selected by Collins, we find that the "this time is different model" is over 8 times more likely to explain the data than the default model.[15] If we extend the data to the present day (an additional 20 years of data), then the "this time is different model" is 23% less likely to explain the data than the default model.

A naive regression approach gives similar results, presented in Table A.2. The constant says the firms average (log) returns of 6.2% per year before their leap and 6.2%+16.6% = 22.8% after the leap. The difference between pre- and post-leap is enormously statistically significant. If we extend the sample to the present day (an additional 20 years of data), after their leap firms average returns of 6.2%+7.0% = 13.2%, and the difference between pre- and post-leap is only barely statistically significant at the 10%. The last regression shows that if we split the post-leap period into the part Collins covered (post1) and the following years (post2), firms did worse in the years Collins did not cover than they did in the pre-leap period, but the difference is not statistically significant. It is as if the great firms just went back to being average. In untabulated results, we find similar patterns when we examine earnings, as opposed to stock returns.

[15] These calculations assume $\bar{\rho} = 12\%$ per year (the average return across all stocks in CRSP over this period), $\sigma_\rho = 6\%$ (the standard deviation of expected returns under the assumption the CAPM holds), and $\sigma_\varepsilon = 17\%$ (the in-sample standard deviation of the portfolio of 11 stocks). Our results are robust to perturbing these numbers somewhat.

H Appendix Proofs

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

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

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

□

Proof of Proposition A.1. Since persuasion is ineffective given (h, d, μ_0), we have $\Pr(h|d, \mu_0) > \Pr(h|m, \mu_0)$ for all m such that $V^S(m, h) > V^S(d, h)$.

1. This means when $\Pr(h|\tilde{d}, \mu_0) > \Pr(h|d, \mu_0)$, we also have $\Pr(h|\tilde{d}, \mu_0) > \Pr(h|m, \mu_0)$ for all m such that $V^S(m, h) > V^S(d, h)$.

2. This means when receivers are maximally open to persuasion that no belief μ' simultaneously satisfies (2) and $V^S(\mu', h) > V^S(\mu(h, d), h)$. Given $\mu(\tilde{h}, \tilde{d}) = \mu(h, d)$ and $\Pr(\tilde{h}|\tilde{d}, \mu_0) > \Pr(h|d, \mu_0)$, this further implies that no belief μ' simultaneously satisfies (2) (now under $\tilde{h}$, $\tilde{d}$, μ_0) and $V^S(\mu', \tilde{h}) > V^S(\mu(\tilde{h}, \tilde{d}), h) = V^S(\mu(h, d), h)$. This is because the modification to $\tilde{h}$, $\tilde{d}$ tightens (2), so any belief that does not satisfy this constraint under h, d also does not satisfy it under $\tilde{h}$, $\tilde{d}$.

□

Proof of Proposition A.2. Take target belief $\tilde{\mu} \in \Delta(\Omega)$. For this target belief, let $\mu_m = \tilde{\mu}$ and $\pi_m(h|\omega) = 1$ for all $\omega \in \Omega$. This model will be adopted by the receiver since $\Pr(h|m, \mu_m) = 1$ and it leads to the target belief, since the receiver's posterior will just equal the persuader-proposed prior.

□

Proof of Proposition A.3. The proof of Proposition 1 establishes this result.

□

Proof of Proposition A.4. Follows the logic of the proof of Proposition 1.

□

Proof of Proposition A.5. For the first part, fix h_1. Let $m_1(h_1)$ satisfy $\pi_{m_1(h_1)}(h_1|\omega) = 1\ \forall \omega \in \Omega$. This will be adopted by the receiver in the first period because it perfectly fits the data. Now consider some h_2 and belief $\tilde{\mu}(h_2, h_1)$ that would be implementable in the static version of the model, which means that there is a model m that implements that belief and satisfies $\Pr(h_2, h_1|m, \mu_0) \geq \Pr(h_2, h_1|d, \mu_0)$, i.e., (ii). Consider model $m_2(h_2, h_1)$ that satisfies $\pi_{m_2}(h_2, h_1|\omega) = \pi_m(h_2, h_1|\omega)$ and $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega) = 1$ for all $\omega \in \Omega$. This model implements $\tilde{\mu}(h_2, h_1)$ while satisfying (i) and (ii).

For the second part, just let $m_1(h_1) = d$.

□

Proof of Proposition A.6. For any (h_2, h_1), let

$$\tilde{\mu}(\omega_i|h_2, h_1) = \min\left\{\frac{\mu_0(\omega_i)}{\Pr(h_2, h_1|d, \mu_0)}, 1 - \sum_{j=i+1}^{N} \tilde{\mu}(\omega_j|h_2, h_1)\right\}.$$

There is a cutoff $\tilde{i}(h_2, h_1)$ satisfying $\tilde{\mu}(\omega_i|h_2, h_1) > 0$ if and only if $i > \tilde{i}(h_2, h_1)$, with $\tilde{\mu}(\omega_i|h_2, h_1) = \mu_0(\omega_i)/\Pr(h_2, h_1|d)$ for all $i \geq \tilde{i}(h_2, h_1)+1$ and $\tilde{\mu}(\omega_{\tilde{i}(h_2,h_1)}|h_2, h_1) = 1-\sum_{j=\tilde{i}(h_2,h_1)+1}^{N} \tilde{\mu}(\omega_j|h_2, h_1)$. This mapping $\tilde{\mu}(h_2, h_1)$ is implementable in the static formulation of the model—i.e., such beliefs satisfy (2)—and first-order stochastically dominate any other beliefs over Ω that are implementable in the static formulation of the model given (h_2, h_1). A model that implements these beliefs satisfies

$$\pi_{m_2}(h_2, h_1|\omega_i) = \begin{cases} 1 & \text{if } i \geq \tilde{i}(h_2, h_1) + 1 \\ \frac{\left(1-\sum_{j=\tilde{i}(h_2,h_1)+1}^{N} \tilde{\mu}(\omega_j|h_2,h_1)\right)\Pr(h_2,h_1|d)}{\mu_0(\omega_{\tilde{i}(h_2,h_1)})} & \text{if } i = \tilde{i}(h_2, h_1) \\ 0 & \text{if } i \leq \tilde{i}(h_2, h_1) - 1. \end{cases}$$

Similarly, for any h_1, let

$$\tilde{\mu}(\omega_i|h_1) = \min\left\{\frac{\mu_0(\omega_i)}{\Pr(h_1|d)}, 1 - \sum_{j=i+1}^{N} \tilde{\mu}(\omega_j|h_1)\right\}.$$

There is a cutoff $\tilde{i}(h_1)$ satisfying $\tilde{\mu}(\omega_i|h_1) > 0$ if and only if $i > \tilde{i}(h_1)$, with $\tilde{\mu}(\omega_i|h_1) = \mu_0(\omega_i)/\Pr(h_1|d, \mu_0)$ for all $i \geq \tilde{i}(h_1) + 1$ and $\tilde{\mu}(\omega_{\tilde{i}(h_1)}|h_1) = 1 - \sum_{j=\tilde{i}(h_1)+1}^{N} \tilde{\mu}(\omega_j|h_1)$. This mapping $\tilde{\mu}(h_1)$ is implementable given h_1 and first-order stochastically dominate any other beliefs over Ω that are implementable given h_1. A model that implements these beliefs satisfies

$$\pi_{m_1}(h_1|\omega_i) = \begin{cases} 1 & \text{if } i \geq \tilde{i}(h_1) + 1 \\ \frac{\left(1-\sum_{j=\tilde{i}(h_1)+1}^{N} \tilde{\mu}(\omega_j|h_1)\right)\Pr(h_1|d,\mu_0)}{\mu_0(\omega_{\tilde{i}(h_1)})} & \text{if } i = \tilde{i}(h_1) \\ 0 & \text{if } i \leq \tilde{i}(h_1) - 1. \end{cases}$$

It remains to show that $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega) = \pi_{m_1}(h_1|\omega)$ for all $\omega \in \Omega$. Note that $\tilde{i}(h_2, h_1) \geq \tilde{i}(h_1)$ since $\Pr(h_2, h_1|d, \mu_0) \leq \Pr(h_1|d, \mu_0)$. This means that, for $i \geq \tilde{i}(h_2, h_1)+1$, $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega_i) = \pi_{m_1}(h_1|\omega_i) = 1$. Similarly, for $i \leq \tilde{i}(h_1) - 1$, $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega_i) = \pi_{m_1}(h_1|\omega_i) = 0$. The remaining states are for $i \in \left\{\tilde{i}(h_1), \ldots, \tilde{i}(h_2, h_1)\right\}$. For all such i, $\pi_{m_2}(h_2, h_1|\omega_i) \leq \pi_{m_2}(h_1|\omega_i)$ so it's possible to fill in $\pi_{m_2}(\cdot, h_1|\omega_i)$ to satisfy $\sum_{h_2'} \pi_{m_2}(h_2', h_1|\omega_i) = \pi_{m_1}(h_1|\omega_i)$.

□

References

Akerlof, George A, and Robert J Shiller. 2015. *Phishing for Phools: The Economics of Manipulation and Deception.* Princeton University Press.

CRSP. 2019. "Daily 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.

Koehler, Jonathan J, and Molly Mercer. 2009. "Selection Neglect in Mutual Fund Advertisements." *Management Science*, 55(7): 1107–1121.

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

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

Phillips, Blake, Kantura Pukthuanthong, and P Raghavendra Rau. 2016. "Past Performance May Be an Illusion: Performance, Flows, and Fees in Mutual Funds." *Critical Finance Review*, 5(2): 351–398.