Naïvely Dialectic Belief Formation

James D. Campbell*
Providence College

October 31, 2018

Abstract

I model a person who interprets the world by adjudicating between the best case and worst case explanations of observable evidence. The model has a parameter, interpreted as skepticism or credulity, to capture how much the person discounts less likely explanations. I apply the model to belief formation, subjective assessment of probabilities, and susceptibility to political spin. In all three cases, someone who is excessively credulous towards far-fetched explanations displays well-known behavioral anomalies. The calibrated model can therefore simultaneously rationalize things like stubbornness in the face of new evidence, systematically biased beliefs, the overweighting of small probabilities, and the existence of maximally extreme yet effective political messaging.

*Email: jcampb10@providence.edu. I am grateful to Bernhard Ganglmair and Stephanie Tilden for extremely helpful discussions.

1 Introduction

The same world can look quite different depending on your perspective. That piece of bad news you just got was just bad luck. You're doing great, and you'll get them next time for sure. Or was it? That piece of bad news means that *you're* bad and you should quit now. Both stories are consistent with the evidence. Isn't the truth somewhere in between?

Dialectics like this are all around us. The glass could be half-full, or it could be half-empty. Anxiety drives us to distraction with what-if scenarios. Maybe you'll win the lottery, but maybe you're wasting two bucks. Political parties use spin to try to sell us the view of the world they would prefer us to hold. The 'good cop, bad cop' routine shows you a reassuring story and a scary one. On the one hand, things could be great, but on the other hand, they could be terrible, and in a complicated world, thinking about a more subtle third hand on top of that is a lot to ask.

In this paper I propose a way to model a person who forms opinions or beliefs based on this kind of binary reasoning. This person processes the world by adjudicating dialectics between competing, consistent, yet polar opposite explanations of what they see. These explanations are the best case and worst case scenarios that are consistent with pieces of observed evidence. Depending on the application we are interested in, the explanations could come from internal deliberation by the individual or be strategically selected by self-interested outsiders trying to exert influence over them. The process I suggest is a heuristic: the person takes a weighted average of the two explanations, where the weights carry a penalty for the explanation that is less precise and less likely to have generated the observed evidence.

The penalty is specified in the model as a parameter called *skepticism*. It represents how much the person being modeled discounts far-fetched explanations of the observed evidence. We could alternatively think of this parameter as being related to naïveté or credulity—the more skeptical a person is, the higher their value of this parameter, and the more credulous a person is, the lower their value of this parameter. The parameter acts as a penalty for implausibility, but it can be either 'too high' or 'too low' relative to true mathematical probability: excessive skepticism and insufficient skepticism are both covered by the model.

This process is systematically different from the mathematics of probability. I think, though, that the dialectic approach is a reasonable description of how people might process information.

The reason why I think it is worthy of attention is that—setting aside its arguable descriptive appeal—it generates predictions that are consistent with well-established behavioral anomalies across three superficially dissimilar settings. In each of the three cases the model predicts things that are consistent with our understanding of how people behave. The three settings are:

Example 1: forming beliefs about the world

A person observes noisy signals about the state of the world. An optimistic view would hold that unfavorable signals are just bad luck, while a pessimistic view would hold that unfavorable signals reflect the true, unhappy state of affairs. How does an adjudication between those competing explanations differ from a Bayesian view of the evidence?

Example 2: interpreting probabilities in risky lotteries

A decision maker is faced with a gamble that offers two outcomes with some probability. They are not good at interpreting probabilities, so form an idea of the relative likelihood of the outcomes by considering a lucky version of the future with an unlucky version of the future. How will their interpretation be difference from the true probability?

Example 3: political spin

In political discourse, competing parties spin evidence to voters in the most favorable possible way for their side, while the truth may be more innocuously in between. How far should each party spin the evidence, and how is the result different than if voters assessed the evidence directly?

It turns out that behavioral anomalies in the first two examples and structural features of the third example can all be explained by a common value of the skepticism parameter. The common value must be too lenient, in the sense that it discounts far-fetched explanations less than Bayesian updating would dictate. In short, a person who forms beliefs and opinions according to my model and who is insufficiently skeptical of far-fetched explanations will be predictably error-prone and exploitable in consistent but quite different ways.

The three examples differ in important respects. The first example proposes a psychological explanation for a person's systematic bias away from Bayesian posteriors. In this setting, the model implies excessive stubbornness of beliefs and the irrelevance of new evidence to a person's beliefs. I also discuss how the population distribution of beliefs can be systematically skewed away from the Bayesian standard under a symmetrically distributed skepticism parameter in the population.

The second example takes only a single piece of information—the probability of an outcome versus an alternative—as the input to the model. The model can in this case explain the hypothesized shape of the decision weights function in Prospect Theory. Experimental evidence on the shape of subjective assessments of probability matches well the shape of the dialectic interpretation of probabilities, where the skepticism parameter governs the skewness of the subjective assessment away from the truth.

The third example particularly invites us to consider the *strategic* selection of frames for the observed evidence. It turns out that as long as the decision maker does not use too stringent a plausibility penalty when evaluating the frames, the Nash equilibrium of the strategic version of the model is for strategic framers to choose the most extreme possible frames. This application is therefore consistent with extreme and polarized political messaging. I show that a predictable partisan ordering and one party's hostility to evidence and the media arise here purely as a function of skewed evidence interpreted dialectically by people of different skepticism.

So, in sum, dialectics of the type we consider here may take a few different forms. They could be internal moral struggles, like the angel and the devil on your shoulders.[1] They could be representative of different psychological states, such as the confident and anxious versions of oneself. They might be be induced by strategic actors, for example by a good cop/bad cop routine or the propaganda of political spin. But what they have in common is that a decision maker is setting up or is presented with a dilemma of interpretation. The tendency toward this kind of dialectic is a familiar way to deal with a complicated world, and so the idea here is to suggest a common behavioral framework for the notion of the dialectic and derive its implications in different decision making problems.

2 Dialectics, multiple selves, and behavioral anomalies

The notion of resolution between competing explanations is in the tradition of Hegelian dialectics (Hegel, 1991). This is the 'thesis, antithesis, synthesis' framework, so named by Johann Gottlieb Fichte Solomon (1985). I have in mind here a person who forms a synthesis between two frames,

[1]See Black (2014) for an excellent, varied history of the idea of this concept, from Plato's *Phaedrus* to *The Simpsons*.

thesis and antithesis, that arise from either internal reflection or external prompting. The thesis and antithesis are truly opposites, in the sense that they are in maximal disagreement about the truth, subject to the constraint of the observable evidence.

Popper (1940) is critical of this dialectic triad, highlighting inconsistencies and absurdities. Although I am not here engaging with the concept of the dialectic triad itself, I want to draw an analogy that sidesteps the critique. I am interpreting the decision maker as an explicitly imperfect dialectician, for whom the triad is a convenient heuristic rather than something designed to draw impeccable conclusions.

The primary inspiration for this paper is Froeb et al. (2016). The idea of that paper is to model adversarial decision making as the choice of interpretations of evidence. A court adjudicates between a plaintiff and defendant but cannot perfectly judge whose interpretation was likelier to have generated the evidence. In the equilibrium of the interpretation selection game, the decision of the court is biased in favor of the unlikelier, more extreme interpretation. The version of the model we develop in Section 6 is closest to the Froeb et al. (2016) case, since there we consider a strategic version of the model and discuss how the plausibility penalty affects the optimal choices of adversarial political messengers.

This paper differs from that work both in the nature of the model as well as the nature of the applications. I combine the idea of the adversarial process with the idea of multiple selves (Ambrus and Rozen, 2013) or multiple rationales (Kalai et al., 2002; De Clippel and Eliaz, 2012) in a decision making process. I use a less general version of the adversarial framework than Froeb et al. (2016) in which the frames are uniform distributions, and consider either non-strategic, behavioral processes or strategic outside influencers as being behind the frames. This lets me parameterize the model in a parsimonious but meaningful way.

There are several prior strands of literature that consider intrapersonal conflict as a motivator for beliefs and decisions. Most notably there is the literature on tension between present and future versions of oneself, for example from Peleg and Yaari (1973) on consumption paths with changing tastes, O'Donoghue and Rabin (1999) on sophisticated and naïve anticipation of future preferences, Kahneman (2011) on impulsive and deliberative selves, and so on. In another vein, Ding (2007) considers a class of intraperson games (motivated by the angel and devil on one's shoulders)

between with competing selves that care about efficiency and equity. Jamison and Wegener (2010) discusses neuroscientific evidence for the multiple selves approach. Another strand is the 'multiple utilities' approach (an excellent review of this concept in philosophy and economics can be found in White, 2006). An example is the idea of a selfish self and ethical self that are mediated by a third, overarching self.

My approach in this paper has affinity with multiple selves models, but it is a little distinct. There is indeed a tension between two frames, but these could be the result of personal reflection or of outside influence. The decision maker is not wrestling with another version of themselves who wants different things, but instead is a single flawed adjudicator who is muddling through the evidence using an imperfect heuristic.

The two anomalies of individual choice we focus on are the inability to accurately interpret and compare stated probabilities (Kahneman and Tversky, 1979) and the inability to accurately update beliefs (Ouwersloot et al., 1998), but I will discuss this in more detail below after deriving the implications of the model.

3 Internal deliberation model

A decision maker would like to make a 'best guess' about the state of the world. They do this by balancing competing explanations offered by different perspectives on the observed data.

The domain of possible states is on the unit interval $[0, 1]$. There are two frames on the data that the decision maker will adjudicate. The frames are two continuous uniform distributions over a subset of the unit interval.[2] They each must be consistent with any available information or evidence, but are otherwise unconstrained. Some possible interpretations of the competing frames are the angel and devil on your shoulders, the best and worse case scenarios, or good cop/bad cop. The common idea is that this decision maker is good at imagining extreme or deterministic scenarios but less good at shades of grey or probabilistic scenarios.

Call the mean of the two frames μ_L and μ_H, and the width of the two frames w_L and w_H. The decision maker comes up with their best guess about the true state by computing a weighted

[2]Depending on the application, the frame may be interpreted as, for example, a suggestion of a particular uniform distribution that could have generated some observed data (this is the example in Section 4) or an ad hoc 'visualization' of the relative size of probabilities (Section 5).

average of the two frames, where a frame gets higher weight if its width is less:

$$\hat{\mu} = \frac{w_H^s}{w_L^s + w_H^s}\mu_L + \frac{w_L^s}{w_L^s + w_H^s}\mu_H \tag{3.1}$$

The decision maker has a *skepticism parameter* that punishes the frame that is relatively wider. The parameter $s > 0$ is capturing the magnitude of this penalty. When s is small, the decision maker weighs more equally the suggestions of the two frames, without considering much which frame is less plausible. When s is large, the decision maker heavily discounts the frame that is relatively less plausible.

We may interpret s as a behavioral parameter that could differ in the population. It is related in some sense to skepticism but need not be taken literally this way. It may be something that can be primed or manipulated by the framing of questions or the inducement of a particular feeling in a decision-maker. One may also imagine that a decision maker's could 'learn' to have a different s based on their past experiences. I should note here that I am not totally comfortable with calling this parameter skepticism. What I have in mind could also be thought of as an inverse measure of credulity, naïveté, or gullibility. To avoid being confusing I have picked a name that matches the direction in which the parameter is increasing: bigger s, more skepticism. Nevertheless I want to keep an open mind about what exactly is being measured.

4 Forming a belief based on evidence

The decision maker observes events that are drawn from an underlying uniform distribution over a subset of the $[0, 1]$ interval.[3] The mean of the underlying distribution is the true state of the world, but the decision maker does not know the true state or what the bounds of the distribution are. The underlying uniform distribution is capturing the idea that the true state of the world is observed with noisy signals. The decision maker entertains two competing conjectures about what the underlying distribution might be, a pessimistic conjecture and an optimistic conjecture. The pessimistic frame 'prefers' the decision maker to consider the true state of the world to be as close as possible to 0, while the optimistic frame 'prefers' the decision maker to consider the true state

[3]We do not consider the case in which the decision maker themselves is responsible for seeking out evidence.

of the world to be as close as possible to 1.

The decision maker forms a posterior point belief about the true state of the world. We may consider this belief as something the decision maker will act on or that will influence their state of mind. The model, applied to this situation, proceeds with a the point belief being formed by taking a weighted average of the means of the competing pessimistic and optimistic frames. The weights are the relative widths of the other frame's uniform distribution. This skepticism means that the more implausible the frame—in the sense that the observed data is more skewed within the distribution—the less weight that frame will get in the decision maker's evaluation.

Take an example. In Figure 1 there are two observed data points, 0.6 and 0.8. The pessimistic frame claims that these have been generated from a uniform distribution over $[0, 0.8]$ and the optimistic frame claims that they came from a uniform distribution over $[0.6, 1]$. The two frames are the most extreme frames in the direction of the preferred end of the spectrum that still include the observed data.

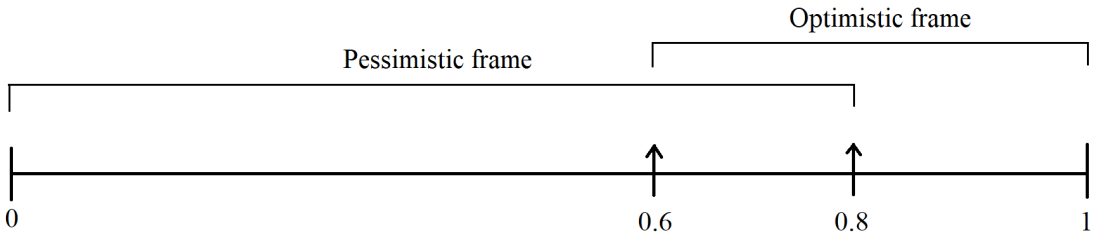

Figure 1: Pessimistic and optimistic frames with two observations

The pessimistic and optimistic frames here are assumed to include the preferred endpoints of each point of view.[4] On one hand, what's the worst-case scenario? On the other, what's the best-case scenario? How the decision maker weighs the two frames depends on the parameter s. Figure 2 shows how the posterior belief depends on the value of s in the example with evidence 0.6 and 0.8.

[4]In Section 6 we discuss strategic selection of the frames by motivated senders.

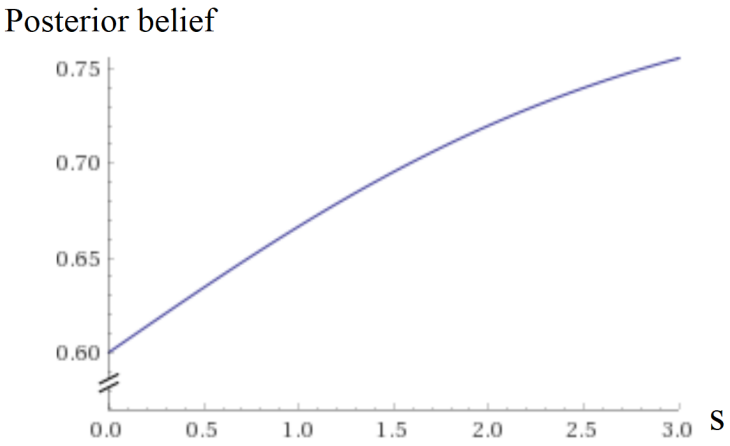

Figure 2: Posterior point belief as a function of skepticism s

As a second example, consider a similar computation if there had been two pieces of evidence at 0.2 and 0.3. Figure 3 shows the posterior belief in that case.

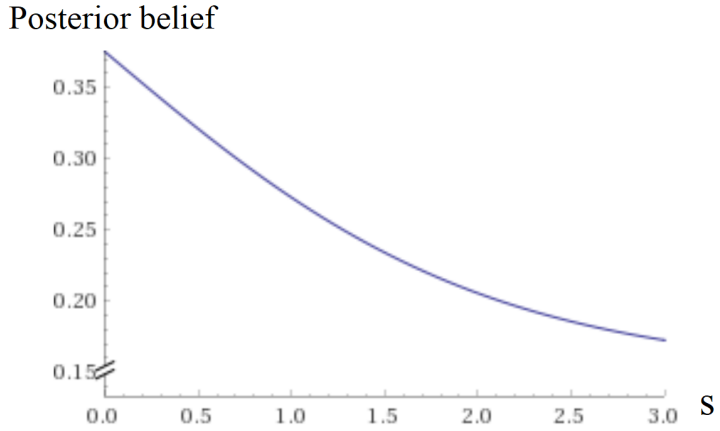

Figure 3: Posterior point belief as a function of skepticism s

From this case we see that for extreme values of skepticism, the decision maker's posterior belief after adjudicating the frames can be outside the scope of the available evidence. This is true in both directions around the Bayesian best guess, consistent with experimental evidence that some people underreact and some overreact to new information relative to Bayes (Epstein et al., 2010).

Here are a few implications of this model of belief formation:

1. *Irrelevance of new evidence*

The arrival of new evidence is irrelevant to the decision maker's posterior belief if it is within the scope of the original evidence. This is because new evidence of that type does not change the width and mean of either frame. This is consistent with evidence that people's beliefs do not move toward an increasing volume of evidence that is distinct from their position. The evaluation of new evidence will seem to be contingent on how a person was evaluating the original evidence, an effect highlighted in Lord et al. (1979).

2. *Everyone makes mistakes*

   The 'right' skepticism for one problem, in the sense that it generates a posterior belief that agrees with the maximally likely state, is wrong for another. If s is a fixed behavioral parameter carried by an individual, this means that no-one will get the right posterior belief every time.

3. *Aggregate belief is systematically far-fetched*

   For a reasonable distribution of s in the population (in the sense of the various s being either sometimes accurate or close to it), posterior beliefs are asymmetrically skewed toward the more outlandish explanation. Mistakes are not symmetrically distributed, as we can see from Figures 2 and 3: the 'misses' are bigger in the direction of less skepticism.

4. *Stubborn beliefs*

   For low values of the skepticism s, the decision maker's posterior assessment after deliberation over the two frames is systematically closer to the ex ante mean (0.5) than the maximum likelihood estimate of the mean considering the observed data. This is consistent with 'stubborn' belief formation that is stickier and slower to respond to new evidence than Bayes' rule would imply.

The last of these implications and the arguments below in Section 5 and Section 6 lead me to conjecture that an appropriate value for the skepticism parameter in calibrating the model is less than 1. Stubborn belief formation, overweighing of small probabilities, and maximally extreme political messaging are all consistent with the implications of this model for $s < 1$.

The notion of stubborn beliefs can also be an alternative explanation for the well-studied 'disposition effect' in finance, that investors have a "tendency to sell assets that have gained value

('winners') and keep assets that have lost value ('losers')" (Weber and Camerer, 1998). Also in a finance application, the representative heuristic and conservatism identified in Barberis et al. (1998) as possible psychological explanations for underreaction to news and overreaction to a *series* of news are also simultaneously reconcilable here.

4.1 Pathologies of skepticism

In the model of dilemmas we are working with, skepticism is a single behavioral parameter, an inverse measure of credulity, that captures how the decision maker weighs competing explanations. The psychological interpretation of the frames invites the possibility that the weights on the optimistic and pessimistic frames might be systematically biased, over and above this behavioral parameter.

While this is outside the scope of the simple model in Section 3, it would be straightforward to think of additional ad hoc weights on the frames. In the belief formation application, this could capture a persistent mental state of the decision maker, such as depression or anxiety; in the political messaging example, it could capture an idiosyncratic preference for a political party or worldview.

In order to disentangle the extra ad hoc weight from a particular value of skepticism it would be necessary to look for asymmetry in effects. To see why, consider the example of depression. Say that a person wakes up one day with a suddenly very low skepticism parameter. Such a person would suddenly be much more pessimistic if things were going well for them (positive evidence) but much more optimistic if things were going poorly for them (negative evidence). If instead a person wakes up one day with a suddenly large ad hoc weight on the low frame, then they will be much more pessimistic if things were going well, but similarly pessimistic if things were going poorly. All of which is to say that there is no reason that we could not include systematic bias in this model, but it would be something different, over and above the effect of the skepticism parameter s.

4.2 Code-switching

The sociolinguistic concept of code-switching—based on language, culture, race, place of origin, and so on—has something in common with the idea that we are working with here (Demby, 2013). A person can have core elements and yet have two or more quite different ways of presenting themselves. One's words and actions then depend on what version of ourselves is doing the speaking and acting. The tension and unease of being pulled in one direction or another is consistent with the systematic difference in my model between evidence and dialectic conclusion.

In this interpretation, the 'evidence' is the analogy of core identity, and the 'frames' the analogy of the different languages or registers employed by the individual. The potential for systematic difference between the implications of the core identity and the result of the dialectic between the frames is reflected in the predictions of the model.

5 Interpreting probabilities

Consider a second application of the model. A person is evaluating a stated probability. The evidence, in the language of the previous application, is in this case a single point, simply the stated number, between 0 and 1. The two frames are the person's ability to consider the two outcomes, weighted by their relative likelihood, but perhaps imperfectly.

The effect of the skepticism parameter in this application will be to boost the weight on the relatively unlikely outcome in the mind of the individual. A low probability event is less likely to occur; a person's credulity determines how readily they can envision this event occurring relative to more likely events. If the value of the skepticism parameter is very low, small probabilities will be disproportionately overweighted by the decision maker.

For concreteness, say that the probability is of some favorable outcome happening (this lets us continue to use the language of the optimistic and pessimistic frames).

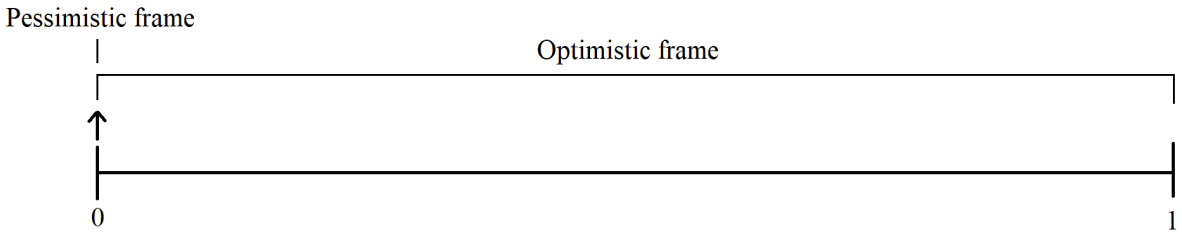

Figure 4: For stated probability 0 or 1, one frame is a point

Figure 4 shows the interpretation of a stated probability of zero. The pessimistic frame here is a zero-width point. In the computation of the perceived probability, the optimistic frame therefore gets zero weight, and the interpreted probability is precisely zero.

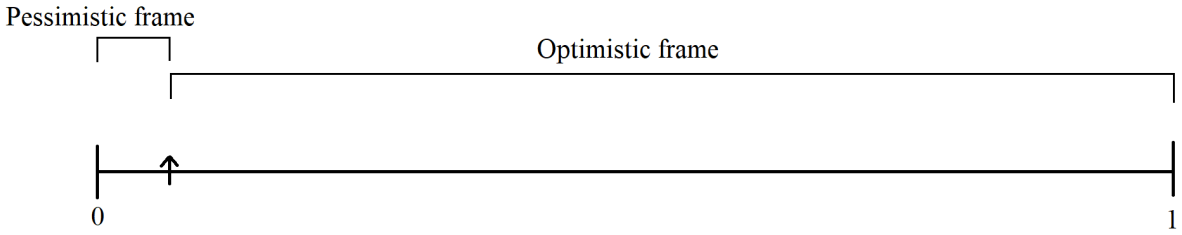

Figure 5: For stated probability close to an endpoint, both frames have width

If instead the stated probability is small and close to zero, the pessimistic frame has width—it may be slim, but there is a chance that things will not fall in the decision maker's favor. The computation of the perceived probability therefore puts positive weight on both frames. The interpreted probability is positive.

The magnitude of the decision maker's interpreted probability depends on the skepticism parameter s. If $s < 1$, the interpreted probability is skewed toward $\frac{1}{2}$, in this case meaning higher than the stated probability. For $s < 1$, when the stated probability is closer to $\frac{1}{2}$, the interpreted probability after the internal deliberation process will be closer to the truth. Toward the endpoints, the discrepancy is larger, and at the endpoints themselves the function relating the stated to interpreted probability accelerates rapidly toward the 45 degree line. The decision maker who is insufficiently skeptical perceives either/or probabilities to be too similar to 50-50 shots.

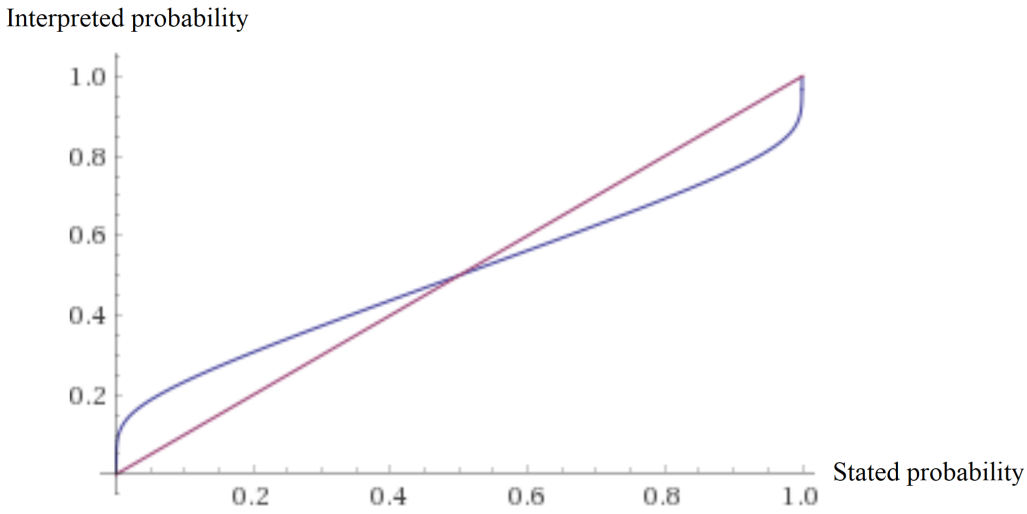

Figure 6: Interpreted probabilities for $s = 0.25$

Figure 6 illustrates the shape of the interpreted probability function for a value of skepticism $s = 0.25$.

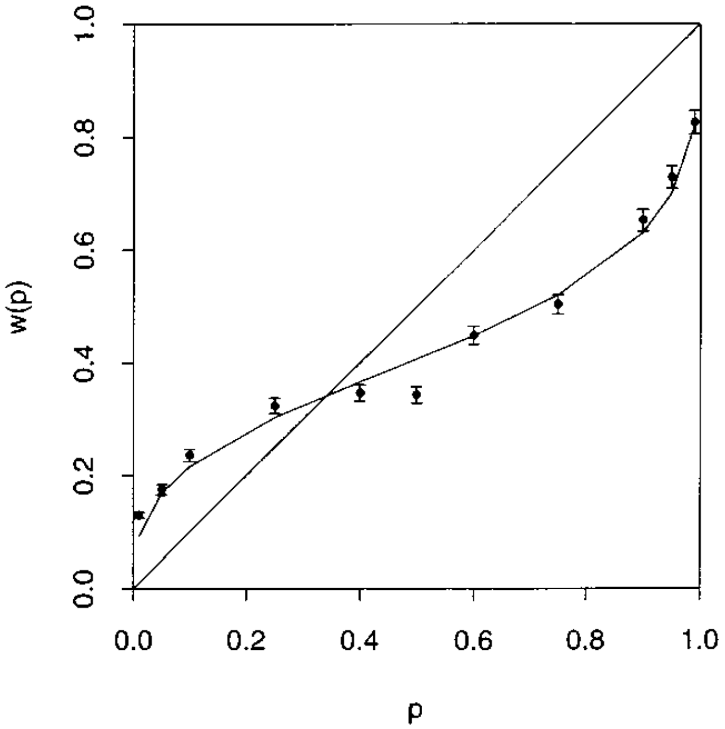

Figure 7: Calibrated decision weight function from Gonzalez and Wu (1999)

This description fits the shape of the 'decision weight' function that relates stated to interpreted probability in Kahneman and Tversky (1979): overweighing of small probabilities, and discontinuous jumps at the endpoints (Figure 7 is an experimentally estimated example). Recall from Section 4 on belief formation that an implication of $s < 1$ was that a decision maker's posterior belief after observing some signals can be outside of the range of those signals. A decision maker with $s < 1$ would therefore display a high degree of both misinterpretation of probabilities *and* a stubbornness in the face of evidence. They place 'too much' weight on implausible interpretations of data relative to standard mathematical tools.

By contrast, a skepticism of precisely 1 in this application generates interpreted probabilities that are identical to the true, stated probabilities. In the examples of Section 4, this penalty was still so low as to overweigh the implausible frames. Loosely speaking, this means that pathologies of belief formation are 'easier' mistakes to make than errors in the interpretation of probabilities, since more extreme values of s are required in the latter case. A decision maker with a fixed s could display behavioral anomalies in belief formation, but perceive probabilities quite accurately.

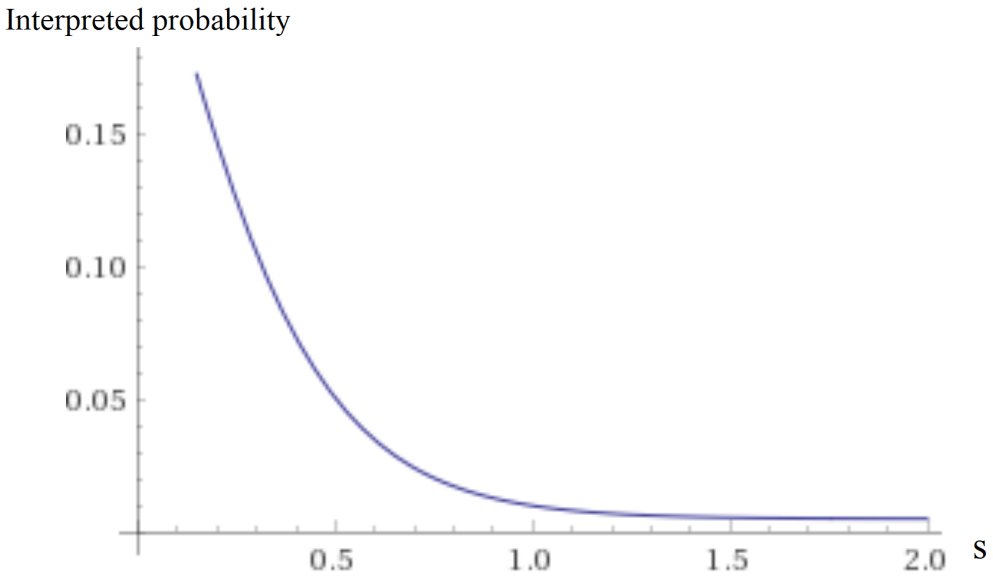

Figure 8: Interpreted probability for a stated probability of 0.01, by plausibility penalty p

A distribution of shapes for the probability weighting function is consistent with available experimental evidence on the best fit for such a function for different decision makers. Gonzalez and Wu (1999) presents evidence for weighting functions with different rotation, exactly the type

of variation that is consistent with different values of the skepticism in the model in this paper.

6 Political messaging

In his dissent in *Abrams v. United States*, Oliver Wendell Holmes wrote that "the best test of truth is the power of the thought to get itself accepted in the competition of the market"[5]. The 'marketplace of ideas' is now a familiar analogy. It is used to support allowing any and all thoughts to be expressed, to be relaxed about falsehood, exaggeration, or conspiracy, since the truth will win out.

There are two steps to this kind of argument. First there is the notion that the marketplace of ideas will resolve around some broadly accepted or dominant idea. Second, there is an assumption that the winning idea will be the truth. Should we see the truth as tautologically whatever comes from the competition among ideas, or is it something distinct that might be missed?

If we apply the model of credulous dialectic to the marketplace of ideas, we get something quite different from Holmes's ideal. The model is consistent with a world with competing political spin that is both as extreme as possible and effective. If the way that the representative decision maker process ideas is through frames strategically chosen by messengers, then we should not expect the 'truth' (in the sense of the maximally likely explanation) to emerge. And if different decision makers embody different skepticism parameters, then we should not expect consensus after the fact. Neither of the two key aspects of the 'marketplace of ideas' emerges here: we would not predict a single, dominant idea to win out, nor would we predict that the most dominant idea or the average idea to match the truth. Instead, a defining feature of the interpretations of ideas will be persistent disagreement that is driven by how credulous different decision makers are to outlandish interpretations.

Political messaging is just one example of a class of *strategic* applications of the model. In strategic applications the competing frames that are adjudicated by the decision maker are selected by strategic, motivated players. In the case of politics, competing operatives seek to spin the facts in their favor. A citizen with limited time or attention to spend thinking about politics interprets the competing narratives without thinking too hard about the fundamentals.

[5] Abrams v. United States (1919).

To analyze a strategic version of the game, consider a rewritten version of Equation 3.1, the decision maker's formation of $\hat{\mu}$. Call the lower and upper bounds on the frame of the low framer $\underline{\mathrm{L}}$ and $\bar{L}$, and similarly for the high framer. We can rewrite both the mean and the width of each frame using these bounds:

$$\hat{\mu} = \frac{w_H^s}{w_L^s + w_H^s}\mu_L + \frac{w_L^s}{w_L^s + w_H^s}\mu_H \tag{6.1}$$

$$= \frac{1}{2}(\underline{\mathrm{L}} + \bar{L})\left[\frac{(\bar{H} - \underline{\mathrm{H}})^s}{(\bar{L} - \underline{\mathrm{L}})^s + (\bar{H} - \underline{\mathrm{H}})^s}\right] + \frac{1}{2}(\underline{\mathrm{H}} + \bar{H})\left[\frac{(\bar{L} - \underline{\mathrm{L}})^s}{(\bar{L} - \underline{\mathrm{L}})^s + (\bar{H} - \underline{\mathrm{H}})^s}\right] \tag{6.2}$$

There is no benefit to either party from moving $\underline{\mathrm{H}}$ away from the lower bound of the evidence, or $\bar{L}$ away from the upper bound of the evidence. Each frame must accommodate the evidence, but it is self-defeating for either messenger to include in their frame possibilities outside the evidence that are more favorable to their opponent.

The interesting cases we are left with are $\underline{\mathrm{L}}$ and $\bar{H}$. Can either framer benefit by moving the bound of their frame away from their preferred extreme pole? To answer this, we need to think about the sign on $\frac{d\bar{\mu}}{d\underline{\mathrm{L}}}$ and $\frac{d\bar{\mu}}{d\bar{H}}$. If the low framer can reduce the decision maker's assessment $\bar{\mu}$ by tightening the bounds on their frame, or the high framer can increase the the decision maker's assessment $\bar{\mu}$ by tightening the bounds on their frame, then we will not observe maximally extreme frames as an equilibrium of the framing game.

The tradeoff for each of the framers is that tightening the bounds of their frame moves the mean of their frame further away from their preferred point, but also reduces the decision maker's weight on the other framer's frame. The relative importance of these considerations depends on skepticism—how much the frame widths matter in the decision maker's weighting.

An example is sufficient to show the nature of the answer. Consider again the example from Section 4: two evidence points at 0.6 and 0.8.

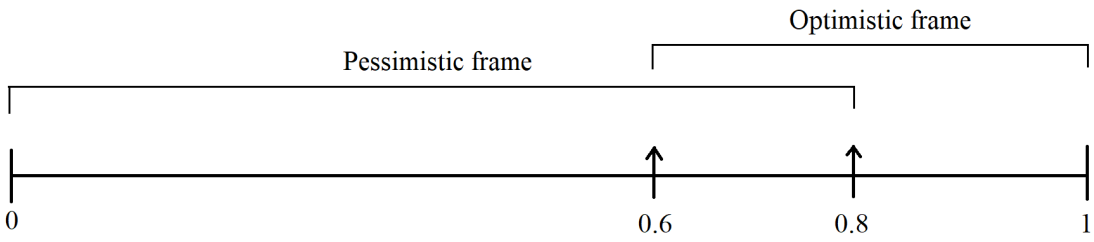

Figure 9: Pessimistic and optimistic frames with two observations

The question at hand here is whether a strategic low framer would prefer a frame other than $[0, 0.8]$ and whether a high framer would prefer a frame other than $[0.6, 1]$.

Say that the high framer selects the most extreme frame, $[0.6, 1]$. It turns out that there is a threshold for s such that the best response of the low framer might or might not be to select their most extreme frame. That threshold lies between $s = 1$ and $s = 2$. For s below that threshold, the best response of the low framer is to be maximally extreme, since the receiver of the message is very credulous For s above that threshold, the best response of the low framer is to be less extreme, since the receiver discounts more heavily the width—the implausibility—of the frame. The maximally extreme frame is a best response as long as the receiver is credulous enough.

This illustrative example generalizes. The Nash equilibrium of the spin game is for both players to choose the most extreme frame that accommodates the evidence—that is, the frame that was assumed to be in the mind of the decision maker in the previous Sections 4 and 5—as long as the skepticism of the receiver is sufficiently low. For sufficiently small s, neither player can induce the decision maker to take a posterior belief that is closer to their preferred point by choosing a uniform distribution that does not include the endpoint. This degree of skepticism is not onerous enough for either framer to do better by choosing a tighter, more plausible, but less favorable explanation of the data.

Of particular interest is that the kind of skepticism that is consistent with behavioral anomalies in Sections 4 and 5 is also sufficient here to generate maximally extreme framing in the strategic version. The threshold on s is not so restrictive as to be irrelevant; the s required for maximally extreme frames to be a bad choice is large relative to the kind of s I have argued is plausible based on the micro behavioral examples. Therefore, for a plausible parameterization of the model,

extremely skewed political spin is optimal for the senders, and the more skewed, the more effective at influencing the receiver.

6.1 Susceptibility to spin

If we turn the question on its head, another way to view this application is to ask which receivers will be susceptible or immune to spin. If the senders are selecting the most extreme frames, the aggregate interpretation of them will depend on the distribution of the skepticism in the population. Ex post, it will seem that the spin 'worked' on some people but not others—while those with low s are successfully pulled in the direction of the far-fetched spin, those with high s are pushed toward the more sober spin. Another implication is that, like in the application in Section 4, new, reinforcing arguments are absorbed by the individual without changing their ex post position. This is in the style of the 'disconfirmation bias' of Taber and Lodge (2006).

This is all before we get to the possibility that the receivers of the frames may themselves have a bias toward one end of the spectrum or another. This may manifest as motivated gathering, for example the second effect of 'confirmation bias' in Taber and Lodge (2006). But note that we do not need this idiosyncratic bias to exist for ex post beliefs to appear systematically skewed. If the evidence congregates more toward one side of the spectrum, then the more far-fetched explanations will be consistently observed from the partisan of the opposite side. Even without idiosyncratic preferences by the receiver, ex post beliefs will be skewed to the far-fetched. This means that an individual's degree of skepticism can influence their (seemingly partisan) dialectic response to frames and in turn their political position, as well as the opposite direction in which the individual's political position influences their response to information.

A striking example of the effect on partisanship is in Stephen Colbert's quip that "[i]t is a well known fact that reality has a liberal bias" (Colbert, 2006). Of the two major political parties in the United States, one is routinely confronted by expert advice and opinion that is quite far away from its preferred platform, for example on climate change or the effect of tax cuts. The frame used to square the evidence with the platform then becomes a tortured one. However, in a world with a low skepticism, such a strategy will be both optimal and effective. This would also be enough to generate polarization in the response to spin, and systematic differences in skepticism

among adherents of different party messages.

This is consistent with evidence that, for example, Republicans appear to discount scientific evidence on climate change more than Democrats (Campbell and Kay, 2014). It is consistent also with evidence that the media's supposed liberal bias is the result of conservative elites' claims of liberal bias (Watts et al., 1999; Lee, 2005): a disinterested, Bayesian media does not reach the same conclusions as the dialectic model, especially for very low skepticism.

7 Common elements from the examples

In all three examples, empirically plausible quirks are consistent with the same feature of the proposed model of internal deliberation. Belief formation that is more stubborn than Bayes' rule, the overweighing of small probabilities and the certainty effect, and the fruitful manipulation of voters via extreme partisan framing are all consistent with the same, low skepticism in deliberation over competing frames.

While there are surely many ways to think about the idea of multiple selves or non-Bayesian reasoning, I think the idea of the naïve dialectic is an interesting model to consider. The idea of dialectic deliberation has a long history in philosophy, psychology, and literature, and I have argued that the model based on this idea is consistent with our understanding of some ways in which the world contrasts with a world of Bayesians. According to this line of thinking, a little more skepticism, a little less credulity, in the population would be a valuable thing.

References

Abrams v. United States (1919). 250 U.S. 616.

Ambrus, A. and Rozen, K. (2013). Rationalising choice with multi-self models. *The Economic Journal*, 125(585):1136–1156.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307 – 343.

Black, R. E. (2014). From charioteer myth to shoulder angel: A rhetorical look at our divided soul. *Colloquy*, 10:36–49.

Campbell, T. H. and Kay, A. C. (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology*, 107(5):809.

Colbert, S. (2006). Correspondents' Dinner Speech. Presented at the 2006 White House Correspondents' Association Dinner, Washington, D.C.

De Clippel, G. and Eliaz, K. (2012). Reason-based choice: A bargaining rationale for the attraction and compromise effects. *Theoretical Economics*, 7(1):125–162.

Demby, G. (2013). How Code-Switching Explains The World. NPR. `https://www.npr.org/sections/codeswitch/2013/04/08/176064688/how-code-switching-explains-the-world`. Accessed October 27th, 2018.

Ding, M. (2007). A theory of intraperson games. *Journal of Marketing*, 71:2.

Epstein, L. G., Noor, J., and Sandroni, A. (2010). Non-bayesian learning. *The B.E. Journal of Theoretical Economics*, 10(1).

Froeb, L. M., Ganglmair, B., and Tschantz, S. (2016). Adversarial decision making: Choosing between models constructed by interested parties. *The Journal of Law and Economics*, 59(3):527–548.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38(1):129–166.

Hegel, G. W. F. (1991). The Encyclopedia Logic. trans. T. Geraets, W. Suchting, and H. Harris.

Jamison, J. and Wegener, J. (2010). Multiple selves in intertemporal choice. *Journal of Economic Psychology*, 31(5):832 – 839. Special Issue on Decision Neuroscience.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291.

Kalai, G., Rubinstein, A., and Spiegler, R. (2002). Rationalizing choice functions by multiple rationales. *Econometrica*, 70(6):2481–2488.

Lee, T.-T. (2005). The liberal media myth revisited: An examination of factors influencing perceptions of media bias. *Journal of Broadcasting & Electronic Media*, 49(1):43–64.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

Ouwersloot, H., Nijkamp, P., and Rietveld, P. (1998). Errors in probability updating behaviour: Measurement and impact analysis. *Journal of Economic Psychology*, 19(5):535 – 563.

Peleg, B. and Yaari, M. E. (1973). On the existence of a consistent course of action when tastes are changing. *The Review of Economic Studies*, 40(3):391–401.

Popper, K. R. (1940). What is Dialectic? *Mind*, 49(196):403–426.

Solomon, R. C. (1985). *In the spirit of Hegel*. Oxford University Press.

Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769.

Watts, M. D., Domke, D., Shah, D. V., and Fan, D. P. (1999). Elite cues and media bias in presidential campaigns: Explaining public perceptions of a liberal press. *Communication Research*, 26(2):144–175.

Weber, M. and Camerer, C. F. (1998). The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organization*, 33(2):167 – 184.

White, M. D. (2006). Multiple utilities and weakness of will: A kantian perspective. *Review of Social Economy*, 64(1):1–20.