

# Persuasion With Motivated Beliefs

*David Hagmann and George Loewenstein*  
*Department of Social and Decision Sciences*  
*Carnegie Mellon University*  
*November 1, 2017*

## **Abstract**

Considerable empirical research finds that people derive utility not only from consumption, but also from their beliefs about themselves and the world. Rather than dispassionately updating their views in response to new information, such belief-based utility implies that people at times avoid information and use other strategies to protect their existing beliefs. We present a two-stage model of persuasion in the presence of belief-protecting strategies and test it in an incentive-compatible task. In the experiment, persuaders seek to shift receivers' subjective numeric estimates related to emotionally charged topics, such as abortion and racial discrimination. We manipulate whether the persuader first acknowledges her own lack of certainty and whether she first has an opportunity to build rapport with the receiver. Though these elements of communication ought to be irrelevant or even backfire under the standard account, our theory predicts they will enhance persuasiveness.

# Introduction

There are many philosophical questions where we expect disagreement among informed people. *To what extent should financial markets be regulated? How paternalistic should government be when it comes to promoting healthy food choices or retirement savings?* But there are also questions where disagreement should not persist. *During President Obama's time in office, did the stock market or the unemployment rate in the United States increase or decrease?* Such questions can be resolved with readily available information, and indeed might have an intuitive answer, given that the US economy had just come out of a recession at the beginning of that period. And yet, a survey found that, among supporters of President Trump, 39% believed the stock market had declined and 67% believed unemployment had increased (Public Policy Polling 2016). To the extent that such beliefs influence both voting behavior and economic decisions, holding inaccurate views may impose substantial economic costs. For example, someone deciding whether to start a small business may choose not to do so, perceiving an environment of weak economic growth, even when indicators suggest the economy had recovered. In this paper, we explore why such entrenched misbeliefs persist and why people may not be persuaded by experts, drawing on recent developments in belief-based utility (Golman, Hagmann, and Loewenstein 2017).

One possible reason for why people hold inaccurate beliefs is that they do not have access to the relevant data or have not been trained in how to interpret it. Many topics are indeed complex and difficult to understand, so that disagreement may simply be explained by insufficient information. On the topic of climate change, for example, we might expect the most partisan polarization among those least able to process scientific findings. Rather than judging the accuracy of the information or the expertise of a source, they might be limited to relying on partisan cues. Yet, Kahan et al. (2016) find that beliefs on whether human activity is primarily responsible for climate change are most polarized not among the uninformed, but among those who score highest on a test of scientific understanding. Rather than using their knowledge to process credible information in an unbiased way and converge

to a shared belief, it appears people instead use their cognitive capacities to bolster their existing view. Such a finding suggests that merely providing people with information may not be effective at resolving disagreements and indeed may backfire when individuals engage in motivated reasoning (Kunda 1990; Babcock et al. 1995).

When presented with information that contradicts an existing belief, decision-makers can either take into account the information and update their belief, or they can dismiss the information and the source along with it. Fischhoff (2007) has cautioned that this may occur when scientists take policy positions on climate change. Rather than persuading the public to take action, they may instead undermine trust in climate science in particular and the scientific process in general. This may even happen to a source who is trusted in other domains. After Pope Francis began to advocate for action on climate change, conservatives in the United States did not update their view on the scientific issue. Instead, they updated their view of the Pope, whose favorability declined among conservative Catholics from 72% to 45% (Gallup 2015). We propose that this discounting of sources that challenge current views occurs particularly when individuals are motivated to defend their beliefs.

For a potential persuader, such findings propose a considerable challenge. If conveying expert information cannot change minds, are we doomed to disagreement on the issues that most divide society? Existing models of persuasion in economics may not help us answer this question, because they have not considered motivated reasons for holding a belief. Such models broadly assume aligned interests (Crawford and Sobel 1982) or perfect Bayesian updating (Kamenica and Gentzkow 2011). But in a world with partisan motivation (Bolsen, Druckman, and Cook 2014), widespread belief in fake news (Allcott and Gentzkow 2017), and asymmetric updating that favors favorable and dismisses unfavorable information (Eil and Rao 2011; Möbius et al. 2014; Sunstein et al. 2016; Tappin, Leer, and McKay 2017), these assumptions do not hold and may explain why we see relatively little effective belief change in the field (DellaVigna and Gentzkow 2010).

Conveying information and persuading others is at the heart of many occupations.

Lawyers, columnists, politicians, and marketers are just a few examples of professions that seek to influence how their audience sees the world. McCloskey and Klammer (1995) estimate that as much as a quarter of US economic activity is dedicated to purposes that could be broadly construed as such persuasion; an estimate that does not take account of the persuasion individuals encounter in everyday situations, as exemplified by the situation of choosing a restaurant with a group of opinionated colleagues. We seek to look at settings that individuals (and societies) are deeply invested in and where persuasion may consequently be particularly challenging.

In this paper, we draw on research in belief-based utility to propose a model of persuasion in which a receiver's motivation extends beyond being accurate. We incorporate the notion that receivers of a persuasive message value their existing beliefs and derive utility from maintaining them, much like they derive utility from their material consumption. Engaging with new information then carries both rewards (to future consumption utility or to obtain new evidence that bolsters existing beliefs), but also risks (undermining existing belief utility). We propose that persuaders who can make themselves appear less threatening to existing beliefs will be more effective at changing the minds of those who are motivated to defend their views. We then test our theory with an experiment, introducing a game in which a persuader aims to change the mind of a receiver on an emotionally and politically charged topic. Both persuader and receiver are incentivized to be accurate and persuaders are further incentivized to shift the receiver's beliefs. We experimentally manipulate whether the sender has an opportunity to lower the receiver's defenses prior to delivering the persuasive message, either by building rapport or by expressing doubt and validating beliefs different from the sender's own.

# Theory

Persuasion appears at first-glance straight forward: an individual is exposed to new information from a (possibly more knowledgeable) sender and faces the task of updating her beliefs. When incentives of both sender and receiver are aligned, this becomes an exercise in applying Bayes' rule. The receiver has some prior belief, obtains a signal, and updates her prior in the direction of the signal to form a posterior belief. Examples of persuasive efforts with aligned incentives are not hard to name: public health organizations seek to persuade doctors to prescribe fewer antibiotics and narcotics, seeking to limit drug resistance and addiction. Biologists try to persuade policymakers and the public to be accepting of evolution, promoting the scientific consensus. Maybe most relevantly today, climate scientists seek to persuade the public of the dangers of global climate change, preventing costly repercussions for the public.

Such persuasion is unexciting in a Bayesian world, where information is of neutral valence and is processed without bias. In such a world, the public adopts the scientific consensus and thus everyone would accept evolution and the reality of climate change. And yet, there is continued debate about these topics, with just above half of US residents stating they believe that humans are responsible for climate change and fewer than a third believing in human evolution by natural selection (Pew Research Center for the People and the Press 2009). When persuasion aims to change minds, evidence from the field suggests that it is only of limited effectiveness or may even backfire and undermine trust in the speaker (DellaVigna 2009; Fischhoff 2007). And despite the personal, professional, and economic relevance of the topic, we have surprisingly few explanations for why receivers of persuasive messages are so reluctant to change their beliefs, even (or especially) when the stakes are high. What might account for this paucity of success stories?

Ever since Stigler (1961) first introduced the concept of information as a scarce commodity into economics, one of the standard assumptions has been that people desire information (only) to the extent to which it improves their decision-making. Individuals stand to gain

consumption utility from making better decisions, and if the information is free, incur no costs. That information is desirable is also at the heart of the economic view of persuasion: when information is presented to a Bayesian agent, she updates her view in an unbiased fashion (e.g. Kamenica and Gentzkow 2011). In this paper, we argue that recent insights on belief-based utility undermine this key assumption: beliefs are not merely means to maximize our consumption utility, but they directly enter our utility function. A persuasive message that aims to correct a decision maker’s false belief may increase her consumption utility, but may in turn also lower her belief utility.

The notion that beliefs enter directly into utility dates back to Abelson (1986), who argues that “beliefs are like possessions” in the sense that we are reluctant to surrender them, even for other beliefs that might be more valuable. Models accounting for direct benefits from holding beliefs have since also found acceptance in economics (e.g. Loewenstein 1987; Caplin and Leahy 2004; Kőszegi 2010). A key implication of belief-based utility is that there may be substantial cost to changing a belief. People often enjoy immediate benefits from holding (or professing to hold) a particular belief, especially to the extent that others share it (Abelson and Prentice 1989). Someone who is part of a religious community, for example, may find that many of their relationships are contingent on maintaining their belief. As Marshall (2014) observes, people may similarly be motivated to resist believing in climate change in order to avoid the anxiety it may generate. Even without such motivation, individuals who have made costly decisions based on beliefs may have incurred sunk costs that they are reluctant to write off (Thaler 1980).

Consequently, decision-makers now have an incentive to avoid information, precisely because it may be persuasive and make the existing belief no longer maintainable (Fels 2015). A rich literature on “information avoidance” finds that people indeed choose not to obtain information even when it is known to be available, useful, and available at no cost or is even costly to avoid (for a review, see Golman, Hagmann, and Loewenstein 2017). They may avoid information in order to avoid disappointment (Kőszegi 2006), maintain their sense

of optimism (Oster, Shoulson, and Dorsey 2013), or avoid anxiety (Persoskie, Ferrer, and Klein 2013). When faced with the prospect of a devastating medical diagnosis, for example, individuals may avoid a medical test that poses the threat of confirming that they indeed have the condition. Obtaining and incorporating undesirable information, like a medical diagnosis, could make it impossible for a decision-maker to maintain a (false) belief, when they have much at stake in maintaining their existing belief.

We propose that similar motivation is at play with hotly contested political issues. Those favorable (unfavorable) towards affirmative action, for example, may not be accepting of evidence that minority students are performing worse (better) than they thought. Although facts about student performance may not matter on their own, they are important when they tackle a higher-order belief: if minority students indeed were to perform poorly, that could challenge the appropriateness of affirmative action. And those who believe abortion should be legal (illegal) under any circumstance may seek to downplay (exaggerate) regret experienced by women who went through the procedure. After all, if most women were to regret the decision, policies imposing mandatory waiting times may be reasonable and not merely punitive. Accepting what at first appears to be neutral and objective information may thus threaten to undermine (potentially costly) investments in an existing and deeply held, higher-order belief.

Consequently, decision-makers seeking to protect their belief utility have an incentive to protect themselves against incongruous information. Rather than changing their views on affirmative action, for example, individuals will seek to avoid threatening information. They may do so by limiting their news consumption to sources that are ideologically aligned with their views (Mullainathan and Shleifer 2005; Bakshy, Messing, and Adamic 2015). When such physical avoidance fails, individuals have a number of other strategies available to them (reviewed in Golman, Hagmann, and Loewenstein 2017). Most relevantly, they may simply fail to update their beliefs and thus not take into account the undesirable information (Eil and Rao 2011; Möbius et al. 2014; Rabin and Schrag 1999; Sunstein et al. 2016; Tappin,

Leer, and McKay 2017) or choose not to commit it to memory (Akerlof and Dickens 1982; Benabou and Tirole 2011; Shu and Gino 2012).

We propose that biased processing of information is the result of “psychological defenses” and that the establishment of the defenses is an important first step that has so far not been accounted for in models of persuasion. When a decision-maker encounters a situation in which she may be persuaded (for example when she reads the headline of a news article), she first determines whether this information is likely to validate or threaten her beliefs. Based on this assessment, and other factors we discuss shortly, she makes an investment in psychological defenses that dampen the impact of the information. In the second stage, she receives the information and updates her beliefs in a biased fashion, where the extent of the bias is determined by her defensive investment.

## **A Two-Stage Model of Persuasion**

In some situations, which are the ones we focus on in this paper, a decision maker faces the choice of whether to obtain, or be receptive to, new information that presents a trade-off between its hedonic and instrumental value. In some cases, obtaining information may threaten cherished values, but offer the promise of improving decision making. Receiving critical feedback on a manuscript draft, for example, may undermine how competent an author believes herself to be, but could enhance the chance that the manuscript gets accepted. In other, much rarer, cases, information may offer the promise of enhancing cherished values but the threat of undermining the quality of decision making. Information conveying expertise in a topic, for example, may make a student feel competent, but lead her to study less and hence perform worse on an exam.

Eil and Rao (2011) show that when participants in a laboratory experiment initially obtain unfavorable information (about their attractiveness or IQ), they are subsequently less likely to desire additional information, which may similarly be unfavorable. We conceptualize this and other avoidant strategies as “psychological defenses” that people erect in order to



protect their cherished beliefs.

Such a defensive mechanism has important consequences for persuasion: in a purely Bayesian world, whether information threatens to challenge or promises to validate someone's belief is an irrelevant factor. In a world in which information has hedonic utility, whether a persuader is expected to provide supporting or disconfirming evidence may determine how willing the listener is to incorporate the information.

Suppose we have a sender who has aligned incentives and seeks to persuade a receiver to change a belief she is invested in. In a Bayesian framework, the sender should communicate (only) information aimed at changing the belief. Moreover, she should express the maximum degree of certainty to ensure the receiver puts high weight on the new information. However, this same strategy is most aversive to someone who does not want to see their beliefs challenged. Encountering an individual who is strongly opposed to one's deeply held belief and is forcefully aiming to attack it poses the greatest threat to one's own belief utility. To the extent that the receiver wants to avoid such threatening information, persuasion may be more successful when the sender appears less confident and driven to persuade.

### Stage 1

Before any communication takes place, a receiver chooses to make an investment in her defenses  $\mathcal{D}$ . This defensive investment has to be made prior to obtaining information and is hence going to affect both favorable and unfavorable information. We propose that three factors drive the extent of the defensive investment: (1) the expected alignment of beliefs between the sender and the receiver ( $\mathcal{A} \in [-1, 1]$ ), (2) the importance attributed to the belief by the receiver ( $\mathcal{S} \in [0, 1]$ ), and (3) the sender's expected expertise and persuasive skill ( $\mathcal{E} \in [0, 1]$ ).

Most straight-forward,  $\mathcal{D}$  is decreasing in  $\mathcal{A}$ . If a sender is believed to hold aligned views that can reinforce an existing belief, there is no reason to establish defenses. Conversely, if a sender is known to hold a strongly opposed belief, defenses become most useful. Defensive

investments are similarly increasing in  $\mathcal{I}$ . To the extent that defensive investments are costly (e.g. because the receiver misses out on additional consumption utility from holding an accurate belief), there is no reason to protect a belief that one does not care about. An example of such a belief may be the value of a number one has randomly been assigned to in an experiment, as in Eil and Rao (2011). Conversely, a belief that is central to one's identity (e.g. abortion for a devout Catholic) is worth protecting even at a cost to consumption utility.

The effect of the sender's expected expertise or persuasive skill on  $\mathcal{D}$  depends on  $\mathcal{A}$ . A sender who is an authority on a subject and is thought to hold aligned beliefs encourages a receiver to drop her defenses in order to absorb information that can protect the belief against future challengers. An authority with misaligned beliefs, on the other hand, poses a serious threat to the receiver who may not be able to counter their arguments.

## Stage 2

After establishing her defenses, the receiver obtains the sender's message and has to update her beliefs. Rather than updating as a pure Bayesian, however, she may choose to discount the sender's information, with the extent of the discounting driven by her defenses. This leads to a type of conservative Bayesian updating (Ward 2007) that is not uniformly conservative, but is rather driven by motivated reasoning. Consequently, a biased posterior is formed by a mixture of the unbiased Bayesian posterior and the prior, where  $\mathcal{D}$  determines the relative weight on the two components:

$$P(A|B) = (1 - \mathcal{D}) \cdot \frac{P(B|A) \cdot P(A)}{P(B)} + \mathcal{D} \cdot P(B|A)$$

Note that when  $\mathcal{D} = 0$ , that is when there are no defenses established by the receiver, the equation reduces to Bayesian updating. If, on the other hand,  $\mathcal{D} = 1$ , that is when the receiver has maximally invested in her defenses, then the posterior is simply equal to the prior, meaning that no updating has taken place. One such example might be that, after reading a headline in a newspaper, the decision-maker chooses not to read the article.

## Implications

The two-stage model of persuasion implies that factors that ought not to matter when information is non-threatening and processed without bias now become relevant. Indeed, in an experiment by Coffman and Niehaus (2014), participants in the role of a seller who worked for commission were able to increase a buyer’s valuation of a good through free-form communication. Successful sellers used their time to establish a rapport with the buyer, engaging in cheap talk that did not convey information about the good itself but made the sender more likable to the receiver.<sup>1</sup> Building rapport and making oneself more likable may be successful strategies to undermine a receiver’s defenses even as they would appear pointless (or even counterproductive) in a Bayesian framework.

Another strategy to bypass a receiver’s defenses may be to express doubt of one’s own view and acknowledge that conflicting beliefs are valid as well, prior to communicating a persuasive message. Doing so validates the recipient’s view of the world and may make the sender appear less threatening, thus prompting lower defenses by the receiver. In a Bayesian world, however, expressing doubt would lead a receiver to put less weight on the information than on that of a persuader who is thought to be confident. Thus, in our model, doubt and validation have the opposite predicted effect and enhance rather than diminish persuasiveness. Establishing likability and expressing doubt may circumvent the receiver’s defenses and end up as effective tools of persuasion where direct information provision fails, much like a small number of Greek soldiers hidden in a wooden horse succeeded where larger armies failed.

## Experimental Design

Our experimental design consists of three stages, which we run sequentially using an online labor market. In the first stage, senders are recruited to provide guesses on a series of percentage estimation tasks and to report their attitudes on related, controversial topics

---

<sup>1</sup>We note that “persuasion” in their study is not a change in beliefs, but an increase in willingness to pay, which increases the seller’s earnings.

(from strongly disagree to strongly agree). After providing their responses, they are asked to write a message persuading another participant who had made a lower (higher) guess on a question to increase (decrease) their estimate. They are incentivized for accuracy on their estimates and for persuasiveness. Senders also draft a short message intended to build rapport and are nudged to express some doubt about their persuasive message. All messages are then rated by a group of judges and the most persuasive messages are used in the third stage.

Receivers also begin by reporting their initial guesses and higher-order beliefs and are then matched with a sender. After reading the sender’s persuasive message, receivers have the opportunity to make another estimate on the question on which they were persuaded. We define the shift in the receiver’s estimate preceding and following the sender’s message as the message’s persuasiveness. Receivers are moreover randomly assigned to one of three conditions. Prior to seeing the persuasive message, they either are shown an open-form communication from the sender aimed at building rapport, a fixed message expressing some doubt about the following advice (selected by the sender), or no message in the control group. We now describe the tasks and the procedure in more detail.

## **Estimation & Higher-Order Beliefs**

Both senders and receivers report their beliefs on five estimation tasks and are incentivized for their accuracy. The estimation tasks elicit beliefs about verifiable quantities, for example the proportion of convicted criminals among deported undocumented immigrants in a particular year. Along with these verifiable quantities, they also report to what extent they agree or disagree with related policies, for example whether they favor a path to legal residency for undocumented immigrants. Although we cannot incentivize a change in the higher-order belief (support or opposition for a policy), we can (and do) incentivize persuasion on the verifiable estimate. The estimation questions and the corresponding higher-order beliefs are shown in table 1.

Table 1: All senders and receivers are incentivized for accuracy on the verifiable beliefs and self-report the corresponding higher-order belief.

Verifiable Belief	Corresponding Higher-Order Belief
Of women who had an abortion, what percentage do you believe report regretting the decision afterward?	Abortion should be legal under any and all circumstances.
What percentage of food stamps (SNAP) recipients do you believe are employed full time or part time (but earned an amount that is difficult to live on) as a fraction of all SNAP recipients (i.e. those who are both employed and unemployed)?	The government spends too much money on social welfare programs and should reduce benefits.
Of African Americans who graduated with a Bachelor’s degree at an accredited 4 year college/university in the U.S., what percentage do you believe obtained a GPA of 3.5 or higher?	Universities should use affirmative action in admissions decisions to increase enrollment of racial minorities.
What percentage of undocumented immigrants deported in 2015 do you believe were convicted criminals?	Undocumented immigrants in the U.S. who meet certain requirements should have a way to remain the country legally.

Verifiable Belief	Corresponding Higher-Order Belief
Stop and Frisk is a practice in some cities in which police officers stop and question a pedestrian, then frisk them for weapons and other contraband. What percentage of Caucasians who are searched under the policy in New York City do you believe get arrested?	The police should rely on “Stop and Frisk” policies, in which the police stop and search people without specific cause.

## **Senders**

We recruit participants from Amazon Mechanical Turk ( $n = 402$ ). Senders first report their beliefs on the five quantitative estimates and the matching higher-order beliefs from table 1.<sup>2</sup> We then match them to one of the estimation tasks and ask them to persuade a receiver who has made a lower (higher) estimate to increase (decrease) their guess. Participants are matched to an estimate for which they reported strong feelings on the associated higher-order belief, if possible, and the direction in which they argue depends on whether they made a low (or high) guess initially. The aim of this matching is to increase the likelihood that senders are invested in their persuasive message and that we can match them with a receiver in the third stage.

We then ask senders to repeat all quantitative estimates again. Although they did not receive any new information, the act of persuading someone to increase (decrease) their estimate may also affect the sender, as suggested by research on self-persuasion (Janis and King 1954; Briñol, McCaslin, and Petty 2012).

Finally, we elicit messages intended to express doubt and build rapport as follows. We

---

<sup>2</sup>We counterbalance whether the quantitative estimates are elicited first or second.

first ask senders to select one of three pre-written messages that most closely resembles the degree of doubt that they have about their advice. The messages, shown in table 2, are written such that one expresses extreme confidence and no doubt whatsoever, one expresses no confidence whatsoever, and the third offers a balanced and reasoned expression of doubt and validates other beliefs. The two extreme messages are intended to be decoys, allowing us to keep the expression of doubt fixed across all senders. Indeed, 92% of participants chose the middle option. To give senders an opportunity to build rapport, we ask them to write a short account of the last time when someone did something nice for them. We expect that this would increase the sender’s likability.

Table 2: Senders select one of these messages to express how confident they are about their recommendation.

---

Messages Expressing Doubt

---

I cannot understand how someone could make a higher or lower estimate than I did. Providing an estimate is easy and I think someone would be foolish not to listen to my argument.

I can see how someone could make a higher or lower estimate than I did. Providing an estimate is difficult, but I believe that I thought carefully about it. Although I am not completely convinced that my estimate is right, I think my argument will help someone make up their mind and make a good decision.

I cannot understand how someone could answer this question at all. Providing an estimate was extremely difficult and I answered pretty randomly. I do not think someone should listen to my argument, because they probably will make a worse decision if they do so.

---

Senders are incentivized on the accuracy of their estimates as well as on their persua-

siveness. We select one estimate from either the first or second elicitation at random and pay participants up to \$2, according to a quadratic scoring rule. Choosing one at random prevents participants from hedging, such that any change we observe in the second elicitation should be due to additional contemplation rather than due to strategic considerations. We incentivize persuasiveness by selecting only from among the most persuasive messages (as rated by other MTurk participants), offering a \$5 fixed bonus if the sender’s message is selected and an additional \$0.50 bonus for each percentage point that the receiver updates her estimate in the direction of the sender.

## Judges

We recruit judges from Amazon Mechanical Turk ( $n = 319$ ) for a fixed payment of \$2 for a task taking 10-20 minutes. They rate a total of 10 messages, all of which are on the same estimation question and go in the same direction (e.g. higher or lower). Participants rate each message on seven dimensions: persuasiveness, the author’s knowledge, reliance on facts, reliance on emotions, and the author’s competency (all on a scale from 0-100), and whether the argument relied on a story or personal experience and whether there is a clear recommendation (both binary). Moreover, within each question and direction pairing, all judges see the same two questions to begin with: one that is coherently written and makes an argument and one that suggests the sender did not take the task seriously. We use these two ratings as an attention check, excluding judges who thought the latter message was more persuasive.

Each message is rated by at least 3 judges and for each question and direction pair, we select the most persuasive message as determined by the median persuasiveness rating.<sup>3</sup>

---

<sup>3</sup>We limit our selection to messages written by senders who selected the non-decoy doubt response, so as to keep this message consistent across receivers.



## Receivers

Lastly, we recruit participants for the role of receiver ( $n = 606$ ) and randomly assign them to one of three experimental conditions. All receivers first complete the five estimation tasks and self-report their associated higher-order beliefs. They are incentivized for accuracy using a quadratic scoring rule with earnings of up to \$1. We then match receivers to a random question on which they expressed a strongly held associated belief, if any. If they did not respond “strongly agree” or “strongly disagree” to any policy question, we assign them to a random estimation task. We choose this matching procedure because we predict our manipulations to increase persuasiveness when receivers are invested in their beliefs. Although we include participants without such a strong view in our analyses, we are interested only in those with a strong view.

Participants in the “baseline” condition then immediately proceed to the sender’s persuasive message and make a second estimate on that question only, for which they are incentivized up to \$1 (again using a quadratic scoring rule). Participants in the “doubt” and “rapport” conditions see an additional screen prior to receiving the sender’s persuasive message. In the “doubt” condition, receivers see the doubt message selected by the sender and are told that this message was selected by the sender to most closely reflect how certain they are about their upcoming advice. Receivers in the “rapport” condition are asked to read the sender’s message about the last time someone did something nice for them. After this additional message, participants proceed to the message and the estimation task.

## Results

Throughout our analysis, we define “persuasion” as the change in beliefs between the first and the second estimate. If the message argues for increasing the estimate, persuasion is equal to  $B_2 - B_1$ , if it argues for decreasing the estimate, it is equal to  $-(B_2 - B_1)$ . Thus, persuasion is always positive if updating occurs in the direction of the message and negative

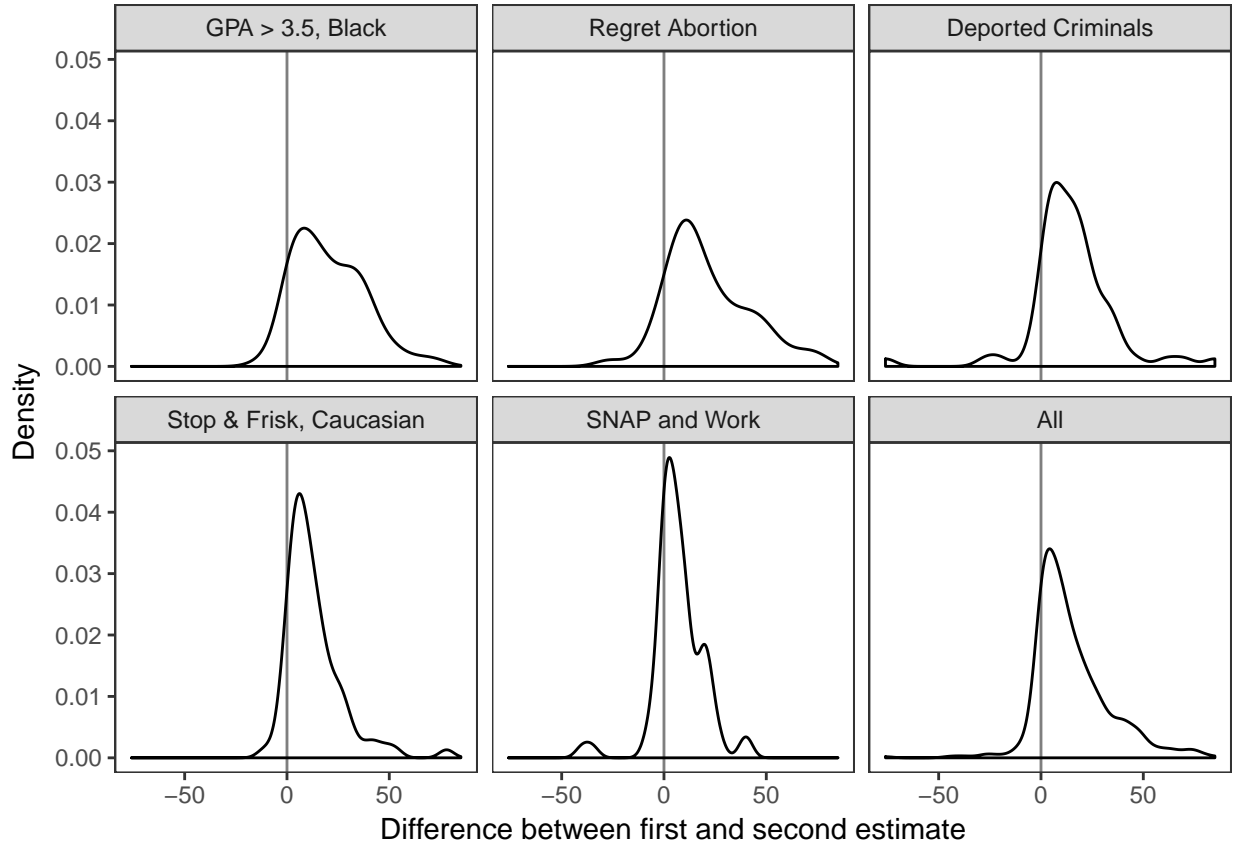


Figure 1: Change in second reported estimate in the direction of the persuasive message.

if updating occurs in the opposite direction.

We first show the degree to which receivers are persuaded across all conditions. Figure 1 plots the density separately for each belief as well as for all topics combined. On average, receivers adjust their estimate 15 points into the direction of the sender. Some receivers are, however, dissuaded by the sender and update in the opposite direction of the argument ( $n = 44$ , or 7% of receivers).

**## Warning: package 'bindrcpp' was built under R version 3.4.1**

In figure 2, we collapse across all topics and show persuasion across experimental conditions. We separate receivers according to whether they had a strong higher-order belief, i.e. either agreed strongly or disagreed strongly with the matching policy. Note that we assigned receivers whenever possible to a belief on which they had a strong opinion, so this is not the result of random assignment. Recall that we chose this matching because our

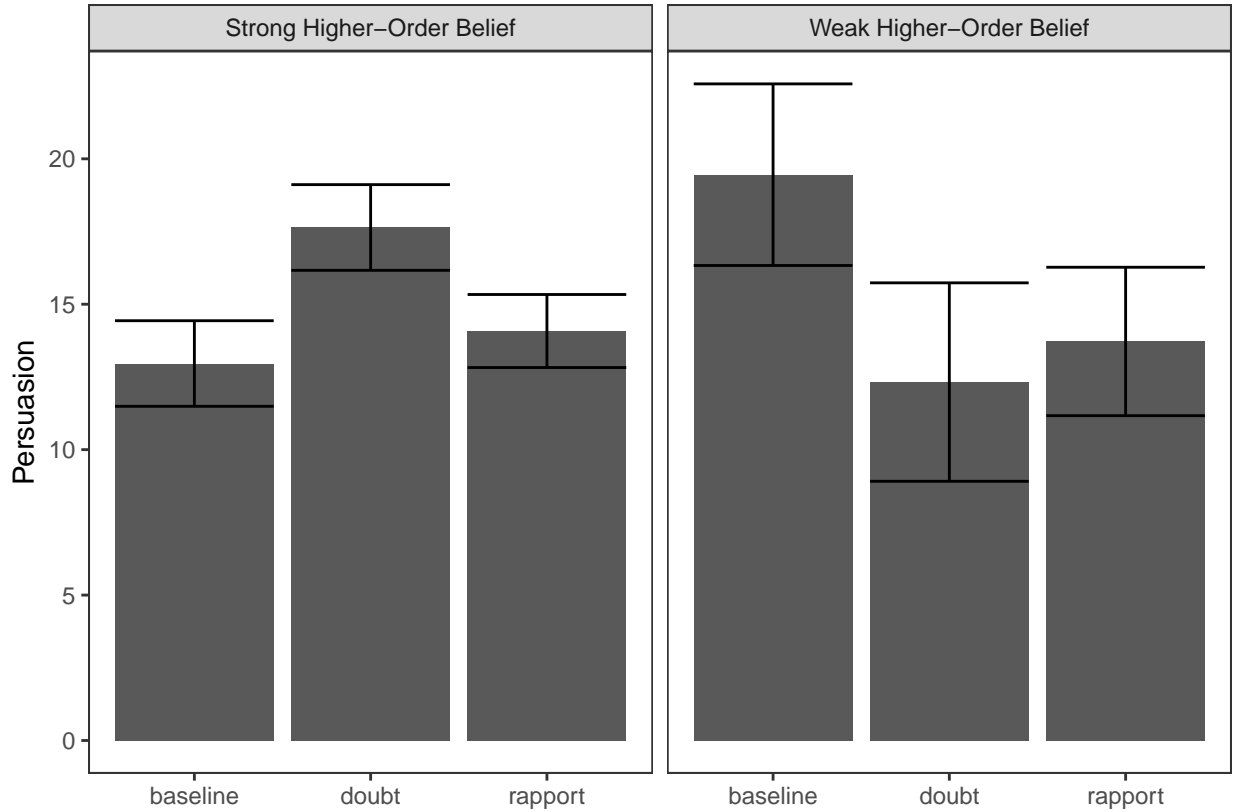


Figure 2: Error bars show standard errors.

predictions relate only to those with strongly held beliefs. We show all respondents for sake of completeness, but focus our analysis on those with a strong belief ( $n = 503$ ).

A t-test comparing the means of belief change for those in the doubt and baseline conditions (with strong beliefs) reveals a significant difference ( $t = -2.246$ ,  $p = 0.025$ ), suggesting that receivers who were exposed to the doubt message prior to being persuaded update more into the direction of the sender. The difference between the rapport and the baseline conditions, however, is not significantly different ( $t = -0.578$ ,  $p = 0.564$ ).

Table 3 shows the results of OLS regression on persuasion. In the baseline model, in which we aggregate across all participants irrespective of their belief strength, we see that receivers changed their estimate on average by 15 points. When we look at the effect of our manipulation separately for those with and without strong beliefs, we see that an expression of doubt increase persuasion for those with strong beliefs, but (directionally) decreases them

for those without. Moreover, those with strong beliefs update significantly less. The third model adds message fixed effects to the regression. The coefficient on the doubt and strong belief interaction remains significant.

Table 3:

	Percentage Points Persuaded		
	Baseline	Interaction	Message FE
Doubt	2.411 (1.821)	-7.131* (4.162)	-5.895 (3.999)
Rapport	-0.269 (1.832)	-5.735 (4.565)	-3.686 (4.383)
Strong Belief	-0.770 (1.989)	-6.493** (3.107)	-4.739 (2.992)
Doubt x Strong Belief		11.808** (4.625)	10.951** (4.454)
Rapport x Strong Belief		6.852 (4.983)	5.677 (4.779)
Intercept	14.978*** (2.020)	19.455*** (2.748)	24.742*** (6.070)
Message FE	No	No	Yes
Observations	606	606	606
R <sup>2</sup>	0.005	0.015	0.117

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Self-Persuasion & Self-Dissuasion

Lastly, our data also allow us to explore whether senders persuaded themselves after writing a message to the receiver. For each reported belief, we calculate “self-persuasion” as the change between the second estimate and the first estimate. This change is positive if it is in the direction in which an argument was made (higher or lower) and negative if it is in the

opposite direction. Self-persuasion may occur when someone makes a strong argument for why the receiver should adjust her estimate in the desired direction. After generating the argument, the sender’s own estimate may now seem too low. On the other hand, dissuasion may occur when the sender realizes that she cannot make a strong argument for the receiver to increase her estimate and ends up revising downward her own response. We calculate the change in beliefs for all estimation tasks. Estimations on which the sender did not attempt to persuade a receiver may vary as well, but any such change should be due to noise. This leads us to estimate the following model:

$$SP = \beta_0 + \beta_1 P + \beta_2 M + \beta_3 (P \cdot M),$$

where  $SP$  is self persuasion, i.e. the change in beliefs between the second and the first estimate,  $P$  is the mean persuasiveness of the message as rated by the judges, and  $M$  is whether the sender wrote a message for that belief (which is equal to 1 for one estimate and 0 for the remaining four estimates).

The results are shown in table 4. We see that across all senders, there is no evidence of self-persuasion (baseline). However, once we control for the sender’s persuasiveness, we see the expected interaction: a sender at the high-end of persuasiveness with a mean rating of 90 (on a scale from 0 to 100) adjusted her estimate by 2.4 points in the direction of the advice she gave. On the low-end of persuasiveness, on the other hand, a sender with a mean rating of 10 revised her estimate away from her argument by 2.3 points.

## Discussion

When decision-makers derive utility not only from their consumption, but also from their beliefs, they may engage in efforts to protect themselves against conflicting information. Inspired by recent advances in belief-based utility, we propose a two-stage theory of persuasion and present results from an experiment. In the first stage, a decision-maker assesses the threat

Table 4:

	Self-Persuasion	
	Baseline	Interaction
	(1)	(2)
Persuasive Message	-0.539 (0.558)	-2.882*** (1.104)
Persuasiveness		-0.055 (0.183)
Persuasiveness x Persuasive Message		0.059** (0.024)
Intercept	-7.106 (4.457)	-5.763 (5.767)
Belief FE	Yes	Yes
Individual FE	Yes	Yes
Observations	2,005	2,005
R <sup>2</sup>	0.260	0.262

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

to her beliefs posed by an incoming persuader. The more threatening the persuader appears to be, the greater the investment in her mental defenses. After establishing her defenses, the receiver is exposed to the sender's message and updates her beliefs conservatively, with the extent of conservatism determined by her defenses.

We propose that supposedly irrelevant factors, like establishing rapport, and factors that under a standard account may make a sender less persuasive, like expressing doubt, can actually increase persuasiveness. In our experiment, we find evidence that an expression of doubt and acknowledgment of opposing views increases persuasiveness for those with a motivated reason to maintain their beliefs. We do not, however, find evidence that our rapport manipulation worked. We also explore the effect of writing a persuasive message on the sender and find that those who craft a persuasive message (as rated by independent judges) end up persuading themselves, while those who craft unpersuasive messages end up dissuading themselves, i.e. they update away from their recommendation.

Persuasion occurs in a wide range of domains from advertisement to policy. Indeed, much of political discourse aims to pit two opposing sides against each other in the hope that the better argument wins out. When people have motivated reasons to believe one argument over another, however, and update with bias, then such arguments may increase polarization, rather than bring convergence. This appears particularly prominent in important, high-stakes policy decisions, like how society should respond to the threat of global climate change. A better understanding of persuasion in such emotionally charged environments may lead to more productive political discourse and help reverse a growing trend of belief polarization.

## References

Abelson, Robert P. 1986. “Beliefs Are Like Possessions.” *Journal for the Theory of Social Behaviour* 16 (3): 223–50. doi:10.1111/j.1468-5914.1986.tb00078.x.

Abelson, Robert P., and Deborah A. Prentice. 1989. “Beliefs as Possessions: A Functional Perspective.” In *Attitude Structure and Function*, edited by Anthony R. Pratkanis, Steven J. Breckler, and Anthony G. Greenwald, 361–81. Hillsdale, N.J.: Erlbaum Associates.

Akerlof, George A., and William T. Dickens. 1982. “The Economic Consequences of Cognitive Dissonance.” *The American Economic Review* 72 (3): 307–19. <http://www.jstor.org/stable/1831534>.

Allcott, Hunt, and Matthew Gentzkow. 2017. “Social Media and Fake News in the 2016 Election.” Working Paper 23089. National Bureau of Economic Research. <http://www.nber.org/papers/w23089>.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. “Biased Judgments of Fairness in Bargaining.” *The American Economic Review* 85 (5): 1337–43. <http://www.jstor.org/stable/2950993>.

Bakshy, Eytan, Solomon Messing, and Lada Adamic. 2015. “Exposure to Ideologically Diverse News and Opinion on Facebook.” *Science*, May, aaa1160. doi:10.1126/science.aaa1160.

Benabou, R., and J. Tirole. 2011. “Identity, Morals, and Taboos: Beliefs as Assets.” *The Quarterly Journal of Economics* 126 (2): 805–55. doi:10.1093/qje/qjr002.

Bolsen, Toby, James N. Druckman, and Fay Lomax Cook. 2014. “The Influence of Partisan Motivated Reasoning on Public Opinion.” *Political Behavior* 36 (2): 235–62. doi:10.1007/s11109-013-9238-0.

Briñol, Pablo, Michael J. McCaslin, and Richard E. Petty. 2012. “Self-Generated Persuasion: Effects of the Target and Direction of Arguments.” *Journal of Personality and Social Psychology* 102 (5): 925–40. doi:10.1037/a0027231.

Caplin, Andrew, and John Leahy. 2004. “The Supply of Information by a Concerned



Expert.” *The Economic Journal* 114 (497): 487–505. doi:10.1111/j.0013-0133.2004.0228a.x.

Coffman, Lucas, and Paul Niehaus. 2014. “Pathways of Persuasion.” [https://www.econ.ucsb.edu/about\\_us/events/seminar\\_papers/l.%20coffman.pdf](https://www.econ.ucsb.edu/about_us/events/seminar_papers/l.%20coffman.pdf).

Crawford, Vincent P., and Joel Sobel. 1982. “Strategic Information Transmission.” *Econometrica* 50 (6): 1431–51. <http://www.jstor.org/stable/1913390>.

DellaVigna, Stefano. 2009. “Psychology and Economics: Evidence from the Field.” *Journal of Economic Literature* 47 (2): 315–72. doi:10.1257/jel.47.2.315.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. “Persuasion: Empirical Evidence.” *Annual Review of Economics* 2 (1): 643–69. doi:10.1146/annurev.economics.102308.124309.

Eil, David, and Justin M Rao. 2011. “The Good News-Bad News Effect: Asymmetric Processing of Objective Information About Yourself.” *American Economic Journal: Microeconomics* 3 (2): 114–38. doi:10.1257/mic.3.2.114.

Fels, Markus. 2015. “On the Value of Information: Why People Reject Medical Tests.” *Journal of Behavioral and Experimental Economics* 56 (June): 1–12. doi:10.1016/j.socec.2015.02.006.

Fischhoff, Baruch. 2007. “Nonpersuasive Communication About Matters of Greatest Urgency: Climate Change.” *Environmental Science and Technology* 41 (21): 7204–8. doi:10.1021/es0726411.

Gallup. 2015. “Pope Francis’ Favorable Rating Drops in U.S. Gallup.com.” July 22. <http://news.gallup.com/poll/184283/pope-francis-favorable-rating-drops.aspx>.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. “Information Avoidance.” *Journal of Economic Literature* 55 (1): 96–135. doi:10.1257/jel.20151245.

Janis, Irving L., and Bert T. King. 1954. “The Influence of Role Playing on Opinion Change.” *The Journal of Abnormal and Social Psychology* 49 (2): 211–18. doi:10.1037/h0056957.

Kahan, Dan M., Asheley R. Landrum, Katie Carpenter, Laura Helft, and Kathleen Hall Jamieson. 2016. “Science Curiosity and Political Information Processing.” SSRN

Scholarly Paper ID 2816803. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2816803>.

Kamenica, Emir, and Matthew Gentzkow. 2011. “Bayesian Persuasion.” *American Economic Review* 101 (6): 2590–2615. doi:10.1257/aer.101.6.2590.

Köszegi, Botond. 2006. “Ego Utility, Overconfidence, and Task Choice.” *Journal of the European Economic Association* 4 (4): 673–707. doi:10.1162/JEEA.2006.4.4.673.

Köszegi, Botond. 2010. “Utility from Anticipation and Personal Equilibrium.” *Economic Theory* 44 (3): 415–44. doi:10.1007/s00199-009-0465-x.

Kunda, Ziva. 1990. “The Case for Motivated Reasoning.” *Psychological Bulletin* 108 (3): 480–98. doi:10.1037/0033-2909.108.3.480.

Loewenstein, George. 1987. “Anticipation and the Valuation of Delayed Consumption.” *The Economic Journal* 97 (387): 666.

Marshall, George. 2014. *Don't Even Think About It: Why Our Brains Are Wired to Ignore Climate Change*. 1 edition. Bloomsbury USA.

McCloskey, Donald, and Arjo Klamer. 1995. “One Quarter of GDP Is Persuasion.” *The American Economic Review* 85 (2): 191–95. <http://www.jstor.org/stable/2117917>.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2014. “Managing Self-Confidence.” *Mimeo*, August. <http://econweb.ucsd.edu/~pniehaus/papers/managingselfconfidence.pdf>.

Mullainathan, Sendhil, and Andrei Shleifer. 2005. “The Market for News.” *The American Economic Review* 95 (4): 1031–53. <http://search.proquest.com.proxy.library.cmu.edu/docview/233026221/abstract/80651ADE4E2B4E30PQ/1>.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. “Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease.” *The American Economic Review* 103 (2): 804–30.

Persoskie, Alexander, Rebecca A. Ferrer, and William M. P. Klein. 2013. “Association of Cancer Worry and Perceived Risk with Doctor Avoidance: An Analysis of Information

Avoidance in a Nationally Representative US Sample.” *Journal of Behavioral Medicine* 37 (5): 977–87. doi:10.1007/s10865-013-9537-2.

Pew Research Center for the People and the Press. 2009. “Section 5: Evolution, Climate Change and Other Issues.” The Pew Research Center. <http://www.people-press.org/files/legacy-pdf/528.pdf>.

Public Policy Polling. 2016. “Trump Remains Unpopular; Voters Prefer Obama on SCOTUS Pick.”

Rabin, Matthew, and Joel L. Schrag. 1999. “First Impressions Matter: A Model of Confirmatory Bias.” *The Quarterly Journal of Economics* 114 (1): 37–82. <http://www.jstor.org/stable/2586947>.

Shu, Lisa L., and Francesca Gino. 2012. “Sweeping Dishonesty Under the Rug: How Unethical Actions Lead to Forgetting of Moral Rules.” *Journal of Personality and Social Psychology* 102 (6): 1164–77. doi:10.1037/a0028381.

Stigler, George J. 1961. “The Economics of Information.” *Journal of Political Economy* 69 (3): 213–25. <http://www.jstor.org/stable/1829263>.

Sunstein, Cass R., Sebastian Bobadilla-Suarez, Stephanie C. Lazzaro, and Tali Sharot. 2016. “How People Update Beliefs About Climate Change: Good News and Bad News.” SSRN Scholarly Paper ID 2821919. Rochester, NY: Social Science Research Network. <http://papers.ssrn.com/abstract=2821919>.

Tappin, Ben M., Leslie van der Leer, and Ryan T. McKay. 2017. “The Heart Trumps the Head: Desirability Bias in Political Belief Revision.” *Journal of Experimental Psychology: General* 146 (8): 1143–9. doi:10.1037/xge0000298.

Thaler, Richard. 1980. “Toward a Positive Theory of Consumer Choice.” *Journal of Economic Behavior & Organization* 1 (1): 39–60. doi:10.1016/0167-2681(80)90051-7.

Ward, Edwards. 2007. “Conservatism in Human Information Processing.” In *Judgment Under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky, 359–69. Cambridge University Press.