NBER WORKING PAPER SERIES

WHY DOES BALANCED NEWS PRODUCE UNBALANCED VIEWS?

Edward L. Glaeser
Cass R. Sunstein

Working Paper 18975
http://www.nber.org/papers/w18975

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2013

Why Does Balanced News Produce Unbalanced Views?
Edward L. Glaeser and Cass R. Sunstein
NBER Working Paper No. 18975
April 2013
JEL No. K0

ABSTRACT

Many studies find that presentation of balanced information, offering competing positions, can promote polarization and thus increase preexisting social divisions. We offer two explanations for this apparently puzzling phenomenon. The first involves what we call asymmetric Bayesianism: the same information can have diametrically opposite effects if those who receive it have opposing antecedent convictions. Recipients whose beliefs are buttressed by the message, or a relevant part, rationally believe that it is true, while recipients whose beliefs are at odds with that message, or a relevant part, rationally believe that the message is false (and may reflect desperation). The second explanation is that the same information can activate radically different memories and associated convictions, thus producing polarized responses to that information, or what we call a memory boomerang. An understanding of these explanations reveals when balanced news will produce unbalanced views. The explanations also account for the potential influence of "surprising validators." Because such validators are credible to the relevant audience, they can reduce the likelihood of asymmetric Bayesianism, thus promoting agreement.

Edward L. Glaeser
Department of Economics
315A Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@harvard.edu

Cass R. Sunstein
Law School
University of Chicago
1111 E. 60th St.
Chicago, IL 60637
cass_sunstein@law.uchicago.edu

Very preliminary draft 4/3/13
Subject to revision

Why Does Balanced News Produce Unbalanced Views?

Edward Glaeser* and Cass R. Sunstein**

Abstract

Many studies find that presentation of balanced information, offering competing positions, can promote polarization and thus increase preexisting social divisions. We offer two explanations for this apparently puzzling phenomenon. The first involves what we call asymmetric Bayesianism: the same information can have diametrically opposite effects if those who receive it have opposing antecedent convictions. Recipients whose beliefs are buttressed by the message, or a relevant part, rationally believe that it is true, while recipients whose beliefs are at odds with that message, or a relevant part, rationally believe that the message is false (and may reflect desperation). The second explanation is that the same information can activate radically different memories and associated convictions, thus producing polarized responses to that information, or what we call a memory boomerang. An understanding of these explanations reveals when balanced news will produce unbalanced views. The explanations also account for the potential influence of "surprising validators." Because such validators are credible to the relevant audience, they can reduce the likelihood of asymmetric Bayesianism, thus promoting agreement.

I. Introduction

One of the underlying principles of a system of freedom of expression, and indeed of democracy itself, is that if "there be time to expose through discourse the falsehood and fallacies, to avert the evil by the process of education, the remedy to be applied is more speech, not enforced silence."[1] The principle seems both important and unexceptionable, but it rests on empirical assumptions that might not always hold true. Under what circumstances is "more speech" actually a remedy? Might more speech be ineffective and even counterproductive? If so, can we identify the reasons and the circumstances, and perhaps specify remedies that do not simply involve "more speech"?

It is well-known that when like-minded groups deliberate, they tend to polarize, in the sense that they generally end up in a more extreme position in line with their predeliberation tendencies.[2] For example, people who are inclined to believe that climate change is not occurring, and that it is some kind of hoax, are likely to become more

* Fred and Eleanor Glimp Professor of Economics, Harvard University.
** Robert Walmsley University Professor, Harvard University and Harvard Law School.
[1] See Whitney v. California, 247 U.S. 357, 377 (1927) (Brandeis, J., concurring).
[2] See Cass R. Sunstein, Going to Extremes (2008).

unified, more confident, and more extreme in that belief after discussion with one another.[3] The phenomenon of *group polarization* can be explained in part by rational updating as information is exchanged among group members.[4] We have suggested, however, that this explanation is inadequate and that group polarization occurs in part as a result of Credulous Bayesianism: Group members insufficiently adjust for idiosyncratic features of particular environments and put excessive weight on the statements of others in the face of common sources of information and unrepresentative group membership.[5]

In light of group polarization and its underlying mechanisms, it might be thought that consistent with widespread faith in the potential power of "more speech," the provision of balanced, objective information would be a valuable corrective, helping to produce a consensus, and perhaps even a rational one, where it did not exist before. If people begin with highly disparate beliefs about climate change, the provision of balanced information might be expected to lead them to converge, with the degree of the effect depending on the trustworthiness of the source. But numerous studies cast serious doubt on this reasonable speculation. On the contrary, balanced information has been found to increase polarization.[6] No less than discussion by like-minded people, such information can led people to have greater confidence and conviction about their antecedent beliefs – and thus to make antecedently divided opinion even more divided than it was before. In short, balanced news can unbalance views.

This finding is puzzling as well as disturbing, and it raises empirical doubts about the effects of more speech in counteracting falsehoods and fallacies. It also raises immediate questions: Why would a presentation of competing views increase polarization? When does it do so? Our principal goal in this Article is to answer these questions. We offer two explanations. The first involves the relationship between the informational signal and people's antecedent beliefs, which can lead to what we call *asymmetric Bayesianism*. We show that if antecedent beliefs are sharply divided, the same signal (whether it is balanced information about a familiar topic, balanced information about an unfamiliar topic, or some kind of purported factual correction) may produce highly disparate responses, leading to even sharper divisions.[7]

The mechanisms in the model involve signals that are produced in two different situations. Consider corrections that backfire. We assume that truthful messages are usually easy to send, which leads to a group of truth-tellers, genuinely seeking to correct

---

[3] See David Schkade et al., What Happened on Deliberation Day?, 95 Cal. L. Rev. 1515 (2007).
[4] See Roger Brown, Social Psychology (2d ed. 1984).
[5] See Edward Glaeser and Cass R. Sunstein, Extremism and Social Learning, 1 J. Legal Analysis (2009).
[6] See, e.g, Charles Lord et al., Biased Assimilation and Attitude Polarization, 37 J Pers and Social Psych 2098 (1979).
[7] For related discussion, emphasizing preferences for like-minded advisers and the resulting polarization, see Wing Suen, The Self-Perpetuation of Biased Beliefs, 114 Ec. Journal 377 (2004).

misunderstandings; but false messages often have the biggest expected impact, which leads to a group of desperate deceivers, trying to persuade people to believe falsehoods. Responses are disparate because people begin with different prior beliefs and hence different degrees of skepticism about the motives of the messenger. Individuals who believe that the messenger is a truth-teller largely have their beliefs buttressed, while individuals who are skeptical think that the message is deceitful, which reinforces and even increases their skepticism.

The second explanation involves memory and salience. An informational signal will focus people's attention on a problem and a range of associated memories and convictions, when they might otherwise be neglected or in the background. Once the problem becomes salient, it may activate a set of memories and fit with an assortment of associated beliefs, whose content depends again on antecedent convictions; the result is to produce polarized reactions. We refer to settings when new information brings back more powerful memories that go in the opposing direction as involving a *memory boomerang*. Of course asymmetric Bayesianism and memory boomerang may be simultaneously at work, with different proportions in different contexts.

The two explanations shed light on a pervasive difficulty. Influential intermediaries, such as general interest newspapers and magazines, often attempt to present both sides of an issue, with the honorable goal of informing, and not merely reinforcing, people's opinions. If the consequence of such efforts is merely to increase people's commitment to what they thought before, and thus to increase polarization, there is a natural question: What is the point?

If the second "memory boomerang" model is correct, it may be difficult to eliminate the polarization, as any new information will tend to bring up memories that reinforce current views. A great deal of information may be necessary to unsettle those views. Under the second model, the identity of those who provide the information should not matter. But if the first model, focusing on actual and perceived signaler incentives, provides the central explanation, then there is a natural solution to the problem. Messages need to come from sources that are seen as credible to the relevant audience and not as likely to be lying (and especially not doing so out of desperation).

Our principal suggestion involves *surprising validators*. When balanced information produces polarization, it is in part because people credit information that is consistent with their preexisting convictions while dismissing information that is inconsistent with those convictions. But when information that is unwelcome (in the sense that it casts doubt on one's prior beliefs) comes from someone who is highly credible and difficult to dismiss, a change in view is more likely. In this respect, surprising validators can overcome asymmetric Bayesianism.

II. Empirical Puzzles: A Quartet of Findings

A seemingly diverse collection of studies attests to the possibility that balanced informational signals, and truthful corrections, sometimes intensify polarization. As we

shall show, one explanation involves the interaction between those signals and people's antecedent beliefs; another explanation involves the activation of different memories and associations. These explanations help to unify the disparate empirical findings that we trace here. It is important to note that those findings involve political beliefs, not market behavior, and it is reasonable to wonder whether the same kinds of polarization would be observed in the market domain (for example, with respect to willingness to pay for goods and services). We believe that the answer is a qualified yes, and we will speculate about that issue in due course.

A. The Effect of Balanced Presentations

For over three decades, it has been well-known that information might not produce consensus, even if is balanced and appears directly to address the concerns that led to divided views in the first place. The underlying phenomenon is usually described as *biased assimilation*.[8] The basic idea is that people assimilate information in a way that is skewed in the direction of support for their antecedent beliefs.[9]

The initial studies involved capital punishment.[10] People were asked to read several studies arguing both in favor of and against the view that capital punishment has deterrent effects. A key finding was that both supporters and opponents of the death penalty were far more convinced by the studies supporting their own beliefs than by those challenging them. After reading the opposing studies, both sides reported that their beliefs had shifted toward a stronger commitment to what they originally thought. One consequence is that the two sides were more polarized than they were before they began to read.

Similar findings have been made in many contexts.[11] In one experiment, both confirming and disconfirming information was provided on the questions whether sexual orientation has a genetic component and whether same-sex couples are likely to be good parents. After receiving that information, people's preexisting beliefs were strengthened, and there was greater, not less, polarization on those questions.[12] In studies of this kind, people are provided with "pro" and "con" arguments, and at least under certain

---

[8] See Lord et al., supra note; Geoffrey D. Monro et al., Biased Assimilation of Sociopolitical Arguments: Evaluating the 1996 U.S. Presidential Debate, 24 Basic and Applied Social Psychology 15 (2002); John McHoskey, Case Closed? On the John F. Kennedy Assassination: Biased Assimilation of Evidence and Attitude Polarization, 17 Basic and Applied Social Psychology 395 (2002).

[9] See Geoffrey D. Munro and Peter Ditto, Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information, 23 Pers Soc Psychol Bull 636 (1997).

[10] Lord et al., supra note.

[11] See note supra.

[12] See note supra.

conditions, provision of such arguments leads to an increase in polarization, even on questions of fact.[13]

B. When Corrections Backfire

Suppose that a society is divided on some proposition. The first group believes A and the second group believes not-A. Suppose that the first group is correct. Suppose finally that truthful information is provided, not from members of the first group but from some independent source, in support of A. It would be reasonable to suppose that the second group would come to believe A. But in important settings, the opposite happens. The second group continues to believe not-A, and even more firmly than before. The result of the correction is to increase polarization.

In a relevant experiment,[14] people were exposed to a mock news article in which President George W. Bush defended the Iraq war, in part by suggesting (as President Bush in fact did) that there "was a risk, a real risk, that Saddam Hussein would pass weapons or materials or information to terrorist networks." After reading this article, they read about the Duelfer Report, which documented the lack of weapons of mass destruction in Iraq. Subjects were then asked to state their agreement, on a five-point scale (from "strongly agree" to "strongly disagree") with the statement that Iraq "had an active weapons of mass destruction program, the ability to produce these weapons, and large stockpiles of WMD."

The effect of the correction greatly varied by political ideology. For very liberal subjects, there was a modest shift in favor of disagreement with this statement; the shift was not significant, because very liberal subjects already tended to disagree with it. But for those who characterized themselves as conservative, there was a statistically significant shift in the direction of *agreeing* with the statement. "In other words, the correction backfired – conservatives who received a correction telling them that Iraq did not have WMD were more likely to believe that Iraq had WMD than those in the control condition."[15] It follows that the correction had a polarizing effect; it divided people more sharply, on the issue at hand, than they had been divided before.

An independent study confirmed the more general effect. People were asked to evaluate the proposition that cutting taxes is so effective in stimulating economic growth that it actually increases government revenue. They were then asked to read a correction.

---

[13] For related findings, with a close connection with our memory model, see David Hardistey et al., A Dirty Word or a Dirty World? Attribute Framing, Political Affiliation, and Query Theory, 21 Psych. Sci. 86 (2010). In particular, the authors find that framing a surcharge for carbon emissions as a "tax" or a "surcharge" had essentially no effect on the preferences of Democrats, but a significant effect on the preferences of independents, and an exceedingly large effect on the preferences of Republicans.

[14] See Brendan Nyhan and Jason Reifler, When Corrections Fail: The Persistence of Political Misperceptions, 32 Polit Behav. 303 (2010).

[15] Id.

The correction actually increased people's commitments to the proposition in question. "Conservatives presented with evidence that tax cuts do not increase government revenues ended up believing this claim more fervently than those who did not receive a correction."[16]

Or consider a test of whether apparently credible media corrections alter the belief, supported and pressed by former Alaska Governor Sarah Palin, that the Affordable Care Act would create "death panels."[17] Among those who viewed Palin favorably but had limited political knowledge, the correction succeeded; it also succeeded among those who views Palin unfavorably. But the correction actually backfired among Palin supporters with a high degree of political knowledge. After receiving the correction, they became *more* likely to believe that the Affordable Care Act contained death panels.

Liberals are hardly immune to this effect.[18] In 2005, many liberals wrongly believed that President George W. Bush had imposed a ban on stem cell research. Presented with a correction from the New York Times or FoxNews.com, liberals generally continued to believe what they did before. By contrast, conservatives accepted the correction. Hence the correction produced an increase in polarization. Importantly but not surprisingly, it mattered, in terms of the basic effect, whether the correction came from the New York Times or Fox News: Conservatives distrusted the former more, and liberals distrusted the latter more. Source credibility is important – a point to which we will return.

C. Unfamiliar Issues

What if the underlying issue is not familiar? In that event, will balanced information produce polarization or instead consensus? A measure of agreement might well be expected, if only because people do not begin with strong antecedent convictions. A study of nanotechnology attempted to answer these questions.[19]

A large set of Americans was divided into two groups. In the "no information" condition, people were simply told that nanotechnology is a process for producing and manipulating small particles. In that condition, people did not divide about nanotechnology. Apparently the issue seemed highly technical, and the mere name and description did not split people along any relevant lines.[20]

---

[16] Id.
[17] See Brendan Nyhan et al., The Hazards of Correcting Myths About Health Care Reform, 51 Medical Care 127 (2013),
[18] Id.
[19] See Dan Kahan et al., Affect, Values, and Nanotechnology Risk Perceptions: An Experimental Investigation, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=968652 (2007)
[20] Id.

In "information exposed" condition, people were given factual material on the potential risks and benefits of nanotechnology. Exposure to such information sharply divided people in accordance with their preexisting political orientations. Those who tended to like free markets, and to distrust government interference, ended up far more favorably disposed toward nanotechnology. Those who tended to favor social equality, and to trust government to promote social goals, ended up far less favorably disposed. In the no-information condition, there was essentially no division between the two groups in their belief that the benefits of nanotechnology outweighed the risks. By a small majority (61 percent), both groups tended to accept that belief. But after exposure to balanced information, the split grew quite dramatically, from 0 to 68 percent, with 86 percent of free marketeers believing that the benefits outweighed the costs, and only 23 percent of egalitarians so believing.[21]

D. Learning from Good News, Neglecting Bad News

What happens if people receive either good news or bad news about themselves or about issues that concern them? It is tempting to predict that they would update their beliefs in accordance with what they learn, whatever its valence; but it should now be clear that this is not always what happens. On the contrary, people are subject to the "good news/bad news effect," by which they tend to credit good news and to update accordingly, but discount bad news and thus selectively ignore it.[22] This effect has been found with respect to people's ratings of their own intelligence, attractiveness, and susceptibility to bad events, such as getting cancer. In fact the effect appears to have neurological foundations.[23]

The good news-bad news effect is closely related to some of the forms of biased assimilation that we have outlined here. If, for example, people believe that affirmative action is a desirable policy, supportive information is good news and unsupportive information is bad news, and there is a tendency to update depending on whether the news is good. For those with disparate views about what counts as good news and what counts as bad, balanced information will produce polarization.

E. Politics and Markets: A Speculative Note

The political domain is distinctive in the sense that people have imperfect incentives to learn and to incorporate new information, and political beliefs are often cheap to hold.[24] As we have noted, it is reasonable to question whether similar effects

---

[21] Id.

[22] David Eil and Justin M. Rao, The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself, 3 *American Economic Journal: Microeconomics* 114 (2011).

[23] Tali Sharot et al., Selectively Altering Belief Formation in the Human Brain, 109 PROC. NAT'L ACAD. SCI. 17058 (2012), http://www.pnas.org/cgi/doi/10.1073/pnas.1205828109.

[24] Edward Glaeser, Paternalism and Psychology, 73 U. Chi. L. Rev. 133 (2006).

would be observed in markets. Suppose, for example, that certain products are said to contain nanotechnology and that the risks associated with nanotechnology become contested across predictable lines. Or suppose that people disagree, again across predictable lines, about the health and ecological consequences of genetically modified food. Would balanced information produce polarization with respect to willingness to pay? Would corrections backfire? Or would we fail to observe such effects in markets?

We are unaware of any empirical research on these questions; it would be quite valuable to have such research. But we would not expect to find any polarization where people face economic incentives and receive clear and prompt feedback. For example, men's razors either work or do not work (in the sense of cutting hair and not cutting faces). With respect to razors, factual claims, attempted corrections, or balanced information, provided in advance, should be immediately overcome by actual experience. The same is true with respect to a wide range of consumer products.

Suppose, however, that people do not receive clear and prompt feedback and that the allegedly adverse consequences would not be felt until some in the future. Genetically modified food is an obvious example. People undoubtedly have different prior convictions on that subject, and for that reason, balanced information might well be subject to biased assimilation. If so, we would expect (on the basis of the studies described above) that such information would produce an increase, along predictable lines, in differences in willingness to pay. For this reason, we speculate (while emphasizing the value of empirical research) that the phenomena described here do have market analogues, at least where prompt feedback is absent. The models we discuss help explain when those analogues are likely to occur.

III. Signals, Priors, and Memory

We now turn to two explanations of the phenomena that we have described. The first involves the interaction between informational signals and people's antecedent beliefs. The intuition here is straightforward. Suppose that there is a report of corrupt behavior on the part of a high-level public official. If the official denies the report in the strongest and most detailed terms, his supporters, given their antecedent convictions, might well be convinced and hence dismiss the report, whereas his critics, given their different antecedent convictions, might believe that the denial is further indication that the report is true, and perhaps even a form of proof. The disparate reactions represent a form of rational updating on the basis of prior convictions. We believe that a mechanism of this kind, which we formalize here, helps to explain all of the findings that we have sketched.

The second explanation involves a memory boomerang. The basic idea is that the relevant issue, and the potentially competing interpretations, are not salient to observers until information is provided. After observers receive that information, it activates associated memories, and that activation ultimately produces polarization. Return to the case of the allegedly corrupt politician: Even if the issue is in the back of people's minds, it is not among their significant concerns, and for that very reason, beliefs may be weakly

held and polarization may be unlikely. But when the denial becomes salient, the issue does as well, and people put it in the context of their other beliefs and values. The nanotechnology findings might well be understood in these general terms. While there is no denial in that case, the balanced discussion increases the salience of the issue, and also puts it in a context that comes with established understandings about what is important, what is likely to be trustworthy, and what is likely to be fabricated.

There is a third explanation, which we merely note here. The explanation is that people's diverse reactions are *motivated*, in the sense that they reflect people's emotional commitments.[25] For example, some people's opposition to capital punishment is highly emotional, and the same is true for some people's support for it. Presentation of balanced information may trigger polarization simply because people are motivated to accept and discount different aspects of the presentation. Emotional investments may ensure that the balanced information divides people more sharply than they were before. Similarly, a denial may backfire with respect to those who are strongly motivated to hold onto their antecedent belief (and who may seek to reduce cognitive dissonance[26]). The denial may lead them to reassert that belief with even greater conviction. When motivations are at work, they will complement the two explanations that we highlight here.

A. Asymmetric Bayesianism: Signals and Priors

For our analysis, we present a stylized version of a setting in which two groups of people differ in their preexisting beliefs. Let us suggest, very abstractly, that people take the world to be either good or bad. It might be good in the sense that public officials are not corrupt, coworkers do not steal, current policies are sound, wars are fought only for the right reasons, or technologies are benign. It might be bad in the opposite senses. People begin with different priors on those topics.

We denote this by assuming that there is some underlying state of the world that is either 1 (good) or 0 (bad). We also assume that there is a leader and communicator, who first undertakes an action (going to war, introducing a product) and then can divulge information to the wider public. The leader's information may represent myriad, diverse facts but it can be distilled down to a single probability that the world is good, which is denoted π. The outside public has less information than the decision-maker, and we go so far as to assume that the only information about the state of the world comes from inference based on the decision-makers action and signal.[27] The public begins by assuming that π is distributed uniformly on the unit interval, but the model can readily be generalized.

---

[25] See, for example, the discussion of motivated reasoning in Jonathan Haidt, The Righteous Mind (2012).

[26] For a relevant study, see Leon Festinger et al., When Prophecy Fails (1956; reprinted 2011).

[27] The core results of the model would weaken if the leaders' informational monopoly became weaker, but the basic phenomena could still persist as long as the leader has some information that was only revealed through actions and signaling.

The leader's welfare from undertaking the action can be represented as $A + B \cdot \pi$, where A is the general preference for the action and B represents an extra benefit that occurs if the state of the world is good. In the Iraq context, A would represent the potential benefits or costs to the leader of going to war whether or not there were weapons of mass destruction, and B represents the benefits of going to war if there were weapons of mass destruction. In the product context, A would represent the costs of developing the product, and B represents the benefit if it is a success. The leader therefore only undertakes the action if $\pi > -A/B$, and we will only consider settings in which the action has taken place.

There are two groups in the population. Both agree that the underlying distribution of π is uniform on the interval [0, 1] and neither of which have any independent information about the state of the world. But the two groups differ in their assessment of $-A/B$. One group, which we refer to as optimists believe that $-\frac{A}{B} = \pi_o$, while the second group (pessimists) believe that $-\frac{A}{B} = \pi_P$, where $\pi_p < \pi_o$. In a war-related context, the optimists believe that the President had little interest in starting a war unless weapons of mass-destruction was real which means that $-\frac{A}{B}$ is high. Pessimists think that the President is interested in fighting whether or not there are weapons of mass destruction and hence $-\frac{A}{B}$ is low.

We do not model where these differences of opinion come from, and we assume that these differences persist, even if the two groups are aware of this belief heterogeneity. While agreeing to disagree is incompatible with standard Bayesian assumptions,[28] it is a fairly universal phenomenon, especially in political contexts.

Our assumption of uniform priors implies that before receiving any signal, optimists think that state of the world is good with probability $(1 + \pi_o)/2$. Pessimists believe that the state of the world is good with probability $(1 + \pi_p)/2$. Since optimists think that the leader would only have undertaken the action if the leader had strong information suggesting that the state of the world is good, they infer from the leaders' action that the state of the world is more likely to be good. We let θ denote the share of optimists in the population.

After the action stage, the leader can then transmit a message, suggesting that the state of the world is good. The connection between the leader and the message being sent is necessary for the results, which would differ substantially if the message was being sent by a totally neutral party. It is possible, of course, that the message is not being sent directly by the leader but by some intermediary who is relating information that the leader let loose.

[28] Robert Aumann, Agreeing to Disagree, 4 The Annals of Statistics (Institute of Mathematical Statistics) 1236 (1976).

We assume that the cost of sending the signal is $c(1-\pi)$, where $c'(1-\pi) > 0$. This assumption is crucial, for the model's core result, and it is worth discussing it at length. The most natural justification is that the principal's belief is itself a reflection of a collection of private observations—some of these which are easy to transmit and some of which are hard to transmit. For the principal to have a high value of π, then he must have received a bevy of positive signals, which makes it more likely that he has at least one signal that is cheap to transmit. . If π is high, then presumably the leader has a robust stock of real data showing that the state of the world is good. In that case, communicating simply involves releasing a bit of that information to the public, which is presumably pretty cheap. If π is low, then the leader will have no such stock of information, and as a real, signaling will require the manufacture of false information, which is presumably somewhat more costly.

A complementary view is that higher values of π reflect a clearer signal about the real world. That clarify makes it easier to fashion a plausible public signal than murky private signals. Even if the signal is basically false, then more accurate supportive information will presumably make the signal more plausible.

The returns to sending the signal depend on whether the state of the world will be revealed on its own. If the world is likely to be revealed anyway, then sending the signal does relatively little. If the world is unlikely to be revealed, then sending the signal will have more value.

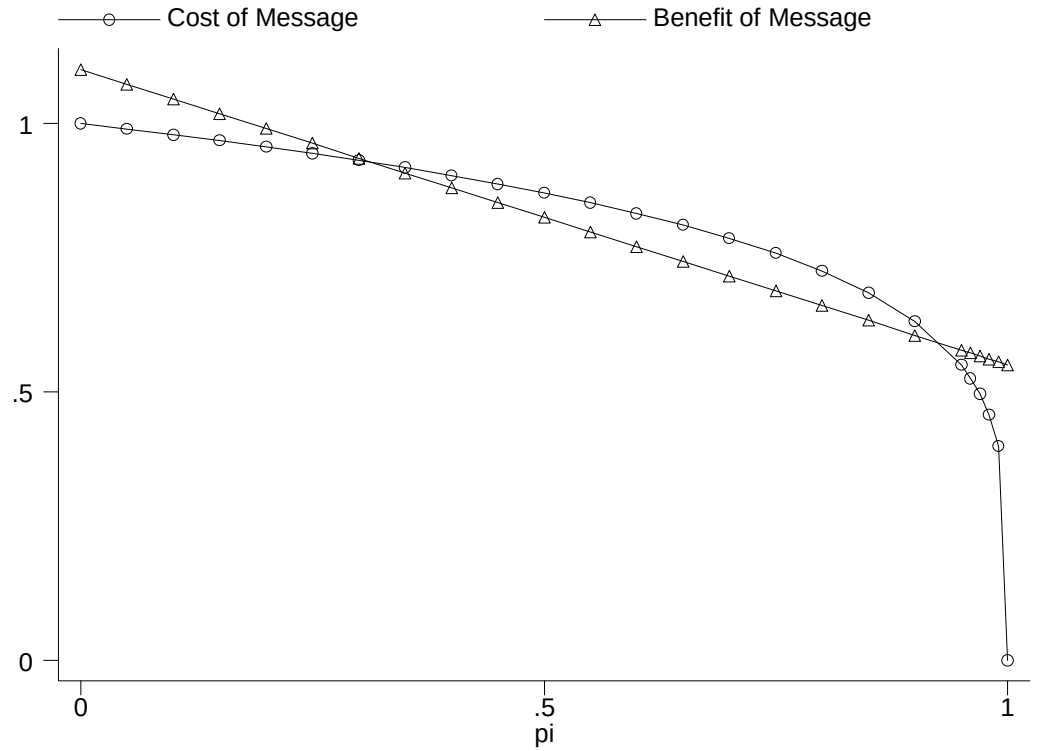
We assume that with probability $\varphi\pi$, the state of the world will be revealed to be good before any decision-making occurs. The state of the world cannot be revealed to be bad. We have assumed an asymmetry which seems natural in settings like the presence of weapons of mass destruction. If there were such weapons, then with some probability it is reasonable to expect that the public will learn that fact. If there are no weapons then it is hard to imagine compelling evidence that will force everyone to that conclusion.

With probability $1-\varphi\pi$, there will be no revelation, and the population will base their opinions on two pieces of information (1) that the communicator undertook the action to begin with and (2) that the communicator chose to send a signal. We have now made a second critical assumption that will also drive the model—revelation is more likely when the state of the world is positive. For our purposes, it would be enough to assume anything that generates a higher return from signaling is that probability of the good state is lower, and this revelation structure ensures that will happen.

Sticking with the revelation structure, we would also deliver similar results is the probability of revelation was highest when the state of the world was 0 or 1 but lowest when the probability is .5. It would also be possible to perturb the model so that the public becomes aware of the principal's own information, not the state of the world itself. That assumption would imply that if the decision-maker has higher quality information, in either direction, then it is more likely that this information will be revealed. .

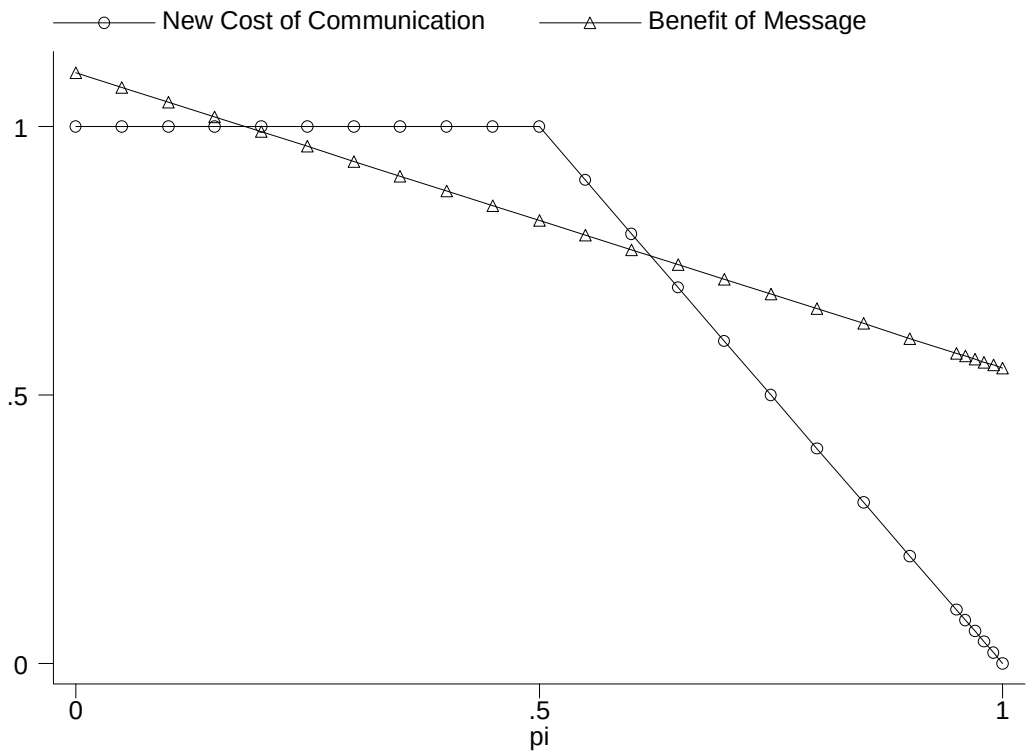
The signal will create a change in aggregative opinion that generates Δ worth of value to the communicator, which we will evaluate later, after describing the beliefs of the two groups conditional upon hearing the signal. As such, the communicator will send a signal if and only if $(1-\varphi\pi)\Delta \geq c(1-\pi)$.

The shape of the function c(.) determines the plausible range of behaviors for the communicator. We are particularly interested in the case where $c(1-\pi)$ is concave, and $(1-\varphi)\Delta \geq c(0)$, $\Delta \geq c(1)$, and there exists some value of π for which $(1-\varphi\pi)\Delta < c(1-\pi)$. The next figure plots the shapes of functions for which this might hold.



In this case, the cost function is $(1-\pi)^{.2}$ and $\Delta = 1.1$ and $\varphi = .5$. In this case, two types of communicators would send messages—those who have very good signals and those who have very bad signals. The communicators with very good signals will communicate because the cost is so low—since they have a lot of positive information about the state of the world. The communicators with very bad signals will communicate even though the cost is higher because the benefit of signaling is higher for lower levels of π, since they know that it is very unlikely that a good state will be revealed before decision-making needs to occur.

An alternative functional form is that $c(1-\pi) = \min(\overline{c}, c_1(1-\pi))$. This function basically says that there is a fixed cost of purely fabricating information ($\overline{c}$) and costs will not rise above that level no matter how low the value of π may be. The next figure shows this case where $\overline{c} = 1$ and $c_1 = 2$.



To prove a more general theorem, still treating Δ as a fixed parameter rather than as the product of people's beliefs, we assume that c(.) is twice continuously differentiable, $c'(1-\pi) > 0$, $c"(1-\pi) < 0$, $lim_{\pi \to 1} c(1-\pi) = 0$, $lim_{\pi \to 0} c(1-\pi) = \overline{c}$, and $lim_{\pi \to 0} c'(1-\pi) = 0$. With these assumption, our first proposition follows:

Proposition 1: If $\overline{c} > \Delta > 0$, then there will exist a value of π dentoed π^{+}_{cross} between zero and one, and signaling will be optimal if and only if π is greater than π^{+}_{cross}. There exists a value of Δ between $\overline{c}$ and $\overline{c}/(1-\varphi)$, denoted $\Delta *$, if $0 > \Delta > \overline{c}$, then there will exist to values of π denoted π^{+}_{cross} and π^{-}_{cross}, with $\pi^{+}_{cross} > \pi^{-}_{cross}$, and signaling will be optimal if π is greater than π^{+}_{cross} or less than π^{-}_{cross} but not if π is between π^{+}_{cross} and π^{-}_{cross}. If Δ is greater than $\Delta *$, then signaling is always optimal. The value of π^{+}_{cross} is decreasing with Δ and increasing with φ and the value of π^{-}_{cross} is increasing with Δ and decreasing with φ.

This proposition illustrates that there are three regions of outcomes based on the value of Δ, given our assumptions about the shape of c(.). If Δ is sufficiently low, then the message will only be sent when the costs of the message are extremely low, which is when π is quite high. This low benefits region essentially implies something akin to truth-telling, since messages are only sent when it is quite likely that they will prove to be accurate.

When $\Delta > \overline{c}$, then it also becomes optimal to send the message when $\pi = 0$, because the rewards from sending the message have become high enough to justify the costs. The assumption that $lim_{\pi \to 0} c'(1-\pi) = 0$, means that the benefits of sending the message will initially fall with π more sharply than the costs, so as long as Δ, is close to $\overline{c}$, then it will not be optimal to send the message for some intermediate values of π. However, when π is sufficiently high, it will again be optimal to send the message.

As Δ rises, these two signaling regions expand, and at some value of Δ below $\overline{c}/(1-\varphi)$, it becomes optimal for everyone to send the signal. When the benefits of persuasion are low, then only truthful signals are sent. When the benefits take on an intermediate level, then there is mixture of truth-telling and highly misleading signaling. When the benefits are extremely high, then everyone signals.

We now endogenize Δ by assuming that observers believe that the signaler is trying only to increase confidence among optimists. This assumption simplifies the arithmetic considerably, and can be justified if listeners believe that the signaler has written off the pessimists as a lost cost. In this case, benefits equals v times the difference in beliefs associated with sending or not sending the signal, among optimists of $v(B_O^S - B_O^{NS})$, where B_0^S is the probability that optimists assign to the state of the world being one after receiving the signal and B_O^{NS} is the probability that optimists assign to the state of the world being one after not receiving the signal.

We make a further technical assumption that $1 - c'^{-1}\big(.5\varphi v^*(1-\pi_O)\big) < \pi_O$, where v^* is defined so that $c\left(c'^{-1}\big(.5\varphi v^*(1-\pi_O)\big)\right) = .5v^*(1-\pi_O)\left(1 - \varphi c' {-} 1.5\varphi v {*} 1 {-} \pi 0\right)$. This assumption enables us to rule out cases where optimists also start suspecting that the signal may have been an act of desperation. With this assumption Proposition 2 follows:

Proposition 2: If $v > v^*$, then observers believe that optimists of all types will signal and as such beliefs are unchanged before and after the signal. If $\frac{2c(1-\pi_P).}{(1-\varphi\pi_P)(1-\pi_O)} > v$, then both groups believe that the signal is more likely to have come if π is high, and that implies that both groups increase their assessment that the world is good. If $v^* > \frac{2c(1-\pi_P).}{(1-\varphi\pi_P)(1-\pi_O)} > v$, then optimists will become increasingly optimistic after hearing the signal. Pessimists may become more pessimistic, and sufficient conditions for their pessimism to increase are that v is close to π_p is close to zero, and $c'(.5) < \varphi\overline{c}$.

Proposition 2 delivers the main result of this section: *divergence is possible as long as different groups start with different assessments about the character of the decision-maker*. Optimists, who trust the decision-maker, are always likely to increase their optimism, because they assume that the signal reflects a low cost of signaling, and a high degree of accuracy. Our technical assumption essentially forced that result.

But pessimists may have a different reaction to the news. While this is possible under many different scenarios, it is sufficient for $c'(.5)$ and π_p both to be low. The assumption that π_p is low essentially means that they are sufficiently skeptical about the state of the world. The assumption that $c'(.5)$ is low ensures that the big decline in the slope of the c(.) function occurs for values of π above .5. This means that for high enough values of v, skeptics believe that the only decision-makers who don't send messages have values of π above .5, and that delivers the result. If the more hones decision-makers withhold signals (in the opinion of the skeptics), then they will become more skeptical after seeing the signal.

The model helps point towards situations where we should expect to see divergence. Groups need to differ in their fundamental assessments and skeptics need to think that signals are particularly likely to reflect desperation, where the cost of dishonesty is overwhelmed by the advantages of bolstering the base. Divergence is less likely if the groups are similar, or if it is fairly implausible that a negative signal reflects a dishonest politician.

We have not formally modeled the impact of different distributions of these prior probabilities, because we have assumed uniformity throughout. If the distributions, of both groups, were heavily weighted to higher values of π, then this would make divergence less likely. Even skeptics would be more likely to think that signals reflect good knowledge. If the distribution of both groups was weighted towards lower values of π, then divergence would be more likely. Optimists ignore the bottom tail of the distribution so the skewness would matter little to them. However, pessimists would put even more weight on the possibility that the message reflects a bad state of the world and cause them to further reduce their ex post beliefs.

.

B. Memory Boomerang

1. The Basic Model

In this section, we present a simple model that presents an alternative explanation for opinion divergence following new information revelation, which we call memory boomerang. In this model, we follow Mullainathan[29] and assume that individuals have forgotten many of their past experiences, but that an intervention may cause forgotten facts to be remembered. Even if an experimental intervention favors one view of the world (i.e. there were weapons of mass destruction in Iraq), the intervention may create a memory boomerang by causing the subject to recall forgotten evidence against that worldview that is far more compelling than the experimental intervention.[30]

[29] See Sendhil Mullainathan, A Memory-Based Model of Bounded Rationality, 117 Q.J. Ec. 735 (2002).

[30] For related findings, see Hardistey, supra note, finding that in the context of carbon emissions, the word "tax" has a different effect on Democrats, independents, and Republicans.

An extreme example of this phenomenon would occur if individuals take in evidence and keep only a brief summary judgment of that evidence at the top of their mind. Over time, that summary judgment may weaken, perhaps because people have forgotten why they held that opinion in the first place. But in an experiment, they are exposed to information that suddenly brings back all the evidence that had been dormant in their longer-term memory and the effect of recalling that lost information overwhelms the direct impact of the experiment. If the recalled information contradicts the experimental intervention, then we say that the experiment created a memory boomerang.

To formalize these ideas, we assume individuals enter an experiment and are exposed to a new piece of data about the state of the world. We assume that before the experiment, each subject has been exposed to a history of signals about an aspect of the state of the world, such as whether there were weapons of mass destruction in Iraq. We also assume that the new signal may jolt the memory of past facts or stories, and cause a past signal to be remembered. In Mullainathan's terminology,[31] the new information is "associated" with the past information and that makes it more likely for the past information to be remembered. Humans appear more likely to remember past events if they are similar to current events.

Data takes the form of signals—a stock of events during which the signal equaled either zero or one. For simplicity of exposition, we will use the term optimism to refer to the belief that the state of the world is one and pessimism to refer to the belief that the state of the world is zero, but that is an entirely arbitrary terminology.

We assume that individuals believe that if the real state of the world is one, then the probability that any signal will equal one is $\frac{\lambda}{1+\lambda} > .5$. If the state of the world is zero, then a fraction $\frac{1}{1+\lambda} < .5$ of signals have a value of one. If an individual has a stock of N_1^R remembered signals with a value of one, and N_0^R remembered signals with a value of zero, then the individual believes that the state of the world is one with probability $\frac{1}{1+\lambda^{\left(N_0^R - N_1^R\right)}}$.

The experimenter provides the subject with a new signal. This experimenter-provided signal also has a value of one, but subjects don't believe that experimenter has any ulterior motives for providing this signal. They do think that the experimenter provided signal is provided by its own staochastic process, that would have generated a positive signal with probability $\frac{\lambda_E}{1+\lambda_E}$ if the state of the world is one, where $1 \leq \lambda_E$. If the state of the world is zero, then the experimenter provided signal would have been one with probability $\frac{1}{1+\lambda_E}$. The experimenter-provided signal may be more or less accurate than previous signals, but in order to get a memory boomerang, the experimenter-provided signal cannot be too accurate.

[31] See note supra.

The term λ_E reflects the accuracy or precision of the signal. As λ_E rises to infinity, then the signal itself essentially perfects discloses the state of the world. If $\lambda_E = 1$, then there is no new information embedded in the signal. The signal will always serve to activate memories, regardless of its precision. However, if this new signal is itself so accurate that it reveals the world, then the memory activation, and indeed all prior information, will have little impact on final beliefs.

If there was no other effect of the signal, then exposure to the experimenter-provided signals changes the posterior to $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}} > \frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}}$. The direct effect of the signal is to increase optimism or the posterior belief that the state of the world is one.

But we allow the signal to have a secondary effect—bringing back similar lost memory of a past signal. We consider two possible structures: random recollection and endogenous recollection. First, we assume that with probably m, the presence of related information cues some other fact, once buried deep in the brain, but now brought forward by this similar event or piece of information. The probability m determines whether a forgotten fact returns but not whether that fact favors optimism or pessimism.

With random recollection we assume that the remembered signal is particularly accessible, and that with probability $(1-\delta)a + \delta\frac{N_1^R}{N_0^R+N_1^R}$, this remembered signal is positive. As long as accessibility of memories is correlated with the sign of the signal, the bias towards accessibility will due to little to impact the posteriors. The δ parameter determines the extent to which recalled signals resemble the initial remembered stock of signals. If $\delta = 0$, then the remembered signal is positive with probability $a \geq \frac{\lambda}{1+\lambda} > .5$, to reflect the fact that the new positive signal may be more likely to bring back has a forgotten signal that is also positive. If $\delta = 1$, then the forgotten signal remembered after the intervention has the same probability of being recalled as the share of one-signals in the pre-intervention stock of signals.

After the experiment, with probability 1-m, the subject's belief is $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}$. With probability $m\left((1-\delta)a + \delta\frac{N_1^R}{N_0^R+N_1^R}\right)$, the post-intervention belief is $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R-1\right)}}$, and with probability $m\left(1-(1-\delta)a - \delta\frac{N_1^R}{N_0^R+N_1^R}\right)$, the post-intervention belief is $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R+1\right)}}$ which is less than $\frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}}$ as long as $\lambda > \lambda_E$. The process of recall makes it possible that the new information will end up reinforcing past beliefs and increase the divergence of beliefs across individuals, especially if the experimenter is providing information that is not itself seen as too important:

Proposition 3: As long as the new information is sufficiently uninformative (i.e. λ_E is low), then the information will on average cause the belief that the state of the

world is one to increase for individuals for whom $\frac{N_1^R}{N_0^R+N_1^R} > a + \frac{1}{\delta}\left(\frac{\lambda}{\lambda+1} - a\right)$ and to decrease for individuals for whom $\frac{N_1^R}{N_0^R+N_1^R} < a - \frac{1}{\delta}\left(a - \frac{1}{\lambda+1}\right)$.

When the new information revealed is weak, individuals who begin with a strong belief that the state of the world is one ($\frac{N_1^R}{N_0^R+N_1^R}$ is high), will increase that belief. There is both the direct effect of the new information, but even when the new signal has little information value it will trigger memories that will typically increase optimism for the initially optimistic, because initially optimistic people also have a relatively optimistic stock of forgotten memories. Recalling one of those memories will only reinforce the person's belief that the state of the world is one.

But for pessimists, who begin with a low value of $\frac{N_1^R}{N_0^R+N_1^R}$, the new information can actually create a memory boomerang that moves their beliefs in the opposite direction. As long as $\frac{1}{\lambda+1} > (1-\delta)a$, which requires a high value of δ—the probability that the newly remembered signal will look like the pre-experiment stock of remembered signals—then sufficiently pessimistic people who only get more pessimistic because of the new information. The correlation between remembered signals and the stock of pre-experiment signals creates the force driving divergence. As long as pessimists are more likely to remember a forgotten pessimistic signal, then any new information, even optimistic information that jogs past memories will create the possibility of generating more pessimism.

The model also suggests other comparative statics on when we should expect to see settings where the new information creates a memory-related backlash that pushes beliefs in the opposite direction. For the next proposition, we let N denote $N_0^R + N_1^R$ and Δ denote $N_0^R - N_1^R$:

Proposition 4: If $\frac{(\lambda+\lambda^\Delta)}{(1+\lambda)(1+\lambda^\Delta)} < \left((1-\delta)a + \delta\frac{N-\Delta}{2N}\right)$, then the intervention will increase the average belief that the state of the world is one. If $\frac{(\lambda+\lambda^\Delta)}{(1+\lambda)(1+\lambda^\Delta)} > \left((1-\delta)a + \delta\frac{N-\Delta}{2N}\right)$, then there exists a value of λ_E, denoted λ_E^*, between 1 and λ, at which the intervention will not change the expected belief that the state of the world equals one, where for values of of $\lambda_E > \lambda_E^*$, the intervention will, on average, increase optimism and for values of $\lambda_E < \lambda_E^*$, the intervention will increase pessimism. The value of λ_E^* is decreasing with a and N, holding Δ constant, and increasing with m. λ_E^* is increasing with δ if $a > \frac{N-\Delta}{2N}$.

Proposition 4 begins by requiring parameters values that rule out increased average optimism even when λ_E is close to zero. If the signal increases optimism when

it is uninformative, then it will increase optimism in all other cases as well. The proposition therefore concerns the case when there is enough initial pessimism (and other parameters are also aligned) so that an uninformative signal from the experimenter leads to more pessimism.

The proposition then emphasizes that in this case the optimism depends on the perceived precision of the new information from the experimenter. When the experimenter is giving out really good data, then this is likely to increase optimism and if the data is relatively meaningless, this will increase pessimism. There is a cutoff value of precision, denoted λ_E^*, that will determine whether the experiment generates added optimism. If other parameters reduce the value of λ_E^*, then this should be understood as suggesting that they will also increase optimism. Parameters that increase λ_E^* should be understood as increasing the likelihood that the new information will produce a memory boomerang. Higher values of λ_E don't change the memories that are awoken, but it does reduce their importance, since the new signal itself has a stronger impact on posterior beliefs.

Increases in a and N both decrease λ_E^* and make a memory boomerang less likely. Higher values of "a" make it more likely that the new signal will jog a positive memory and unsurprisingly that makes an increase in optimism more likely. Increases in the total number of signals, holding Δ constant, pulls the share of past signals that are positive or negative closer to zero, and that makes it harder to have a high probability of pulling a negative signal out of one's past memory. This suggests that individuals with less experience are more likely to experience a memory boomerang, since they are most likely to have a really skewed set of past forgotten signals.

In the relevant parameter space, higher values of δ typically make memory boomerang more common. The stronger the correlation between the stock of remembered signals and forgotten signals jogged by the experiment, the more likely the experiment is to create a memory boomerang, because it is more likely that people who begin as pessimists will pull a lost negative signal from their memory.

2. Endogenous Memory and the Quest for Certainty

These results are relatively similar with an endogenous memory model where individuals have a cost of accessing past memories. In many real world cases, we access memories in order to have a more accurate opinion and make decisions with more accuracy. In our setting, individuals are not making decisions and they will recall information in order to buttress their current views. We assume that humans like to feel as if they know the state of the world and that there are cognitive costs to being uncertain.

Any new information will either reduce or increase uncertainty, which will affect the benefit of accessing past signals. If people are uncomfortable with uncertainty, then new information will increase the benefit of accessing past signals if that new information makes the state of the world less clear. For example, if a person walked in believing that there were weapons of mass destruction in Iraq, but was then exposed to

information disputing that view, then this person will become more uncertain about the state of the world. If the person values certainty either for instrumental reasons, or because ignorance is psychologically uncomfortable, then the person will have an increased incentive to wrack his brain to bring out forgotten data that restores his sense of surety.

It follows that if any new information decreases the current level of uncertainty by supporting people's current views, then that new information will in turn *decrease* the incentive to remember forgotten data. It is also possible that new information may affect the costs of accessing forgotten signals, but we do not model that here.

To keep matters comparable with the previous model, we assume that each individual has an individual specific cost φ_i of accessing a lost memory, where lower values of φ_i reflect a greater stock of accessible information. We also assume that individuals can choose to access different forms of information, so they can choose whether to access a memory supporting the view that the state of the world is good or bad. This memory process can also be seen as reflecting the costs of constructing an argument rather than accessing a memory.

To capture the benefits of certainty, we assume that the psychic benefit of being sure is modeled as $\beta|Posterior\ Belief - .5|$, as such individuals prefer having a probability of 1 or 0 to a probability of .5. Individuals recall information in this framework not to make decisions correctly, but rather so that they are confident in their opinions.

This model takes a different approach from a more standard setting where information is valued because it improves the accuracy of decision-making. Yet the taste for certainty can be connected to more standard models. Uncertainty is undesirable if an individual wants to avoid making a bad choice. That chance is greater if one's probability is close to one-half, and as a result, new information will still increase the incentive to recall if it pushes posteriors closer to one-half.

We will assume that individuals are able to choose what kind of memory to access. If they are leaning positive, but would like to be more confident, they will access a positive memory that reinforces their current views. If they are leaning negative, they will access a negative memory if they would like more certainty. In a more standard setting, individuals would expend effort trying to remember without stacking the deck towards either particular side. After all, remembering a past positive memory is far less informative if the individual set out to remember a positive memory than if the individual cleared his head and tried to think of the most relevant possible forgotten information, which could be either positive or negative. We will discuss the impact of less targeted recall at the end of this section.

We assume that individuals' pre-experiment stock of information reflects optimization decisions about certainty made before the experiment. As such before the experiment individuals were not willing to pay the cost of accessing past memories which

implies that $\varphi_i > \beta\left(\frac{1}{1+\lambda^{\left(N_0^R-1-N_1^R\right)}} - \frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}}\right)$ if $N_1^R > N_0^R$ and $\varphi_i > \beta\left(\frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}} - \frac{1}{1+\lambda^{\left(N_0^R+1-N_1^R\right)}}\right)$ if $N_1^R < N_0^R$. These conditions mean that it was not optimal before the experiment to invest in recalling another memory that confirms one's existing view

The following proposition reviews the possible outcomes after an experiment:

Proposition 5: If $N_1^R > N_0^R$ then the experiment will reinforce the optimism of already optimistic participants, there will be no new memory recall, and ex post beliefs will be more optimistic than ex ante beliefs.

If $N_0^R > N_1^R > N_0^R - .5\frac{Log(\lambda_E)}{Log(\lambda)}$, then experiment participants switch from pessimism to optimism after the experiment, but there will be no added memory recall.

If $N_0^R - .5\frac{Log(\lambda_E)}{Log(\lambda)} > N_1^R > N_0^R - \frac{Log(\lambda_E)}{Log(\lambda)}$, then experiment participants will switch from pessimism to optimism, and if $\frac{(\lambda-1)\lambda_E}{\left(\lambda_E\lambda^{\left(1+N_1^R-N_0^R\right)}+1\right)\left(\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}\right)} > \frac{\varphi_i}{\beta}$ there will be memory recall which supports the new evidence.

If $N_0^R - \frac{Log(\lambda_E)}{Log(\lambda)} > N_1^R$, then participants will remain pessimistic after the new information, the new information will increase the incentives to remember and lead to a pessimistic memory recall if $\frac{(\lambda-1)\lambda_E}{\left(\lambda_E\lambda^{\left(N_1^R-N_0^R\right)}+1\right)\left(\lambda_E+\lambda^{\left(N_0^R+1-N_1^R\right)}\right)} > \frac{\varphi_i}{\beta}$, which will lead pessimists to become more pessimistic after the experiment as long as $\lambda > \lambda_E$.

The proposition highlights that the impact of the information depends on the initial signals remembered by the participants, and their costs of recalling memories divided by their benefit of certainty. An optimistic signal will reassure only those individuals who begin as optimists. The new signal will increase their optimism and create no memory recall, since they have become even more confident in their views, and therefore less willing to expend effort to recall past experiences.

Individuals who are very mildly pessimistic will be made optimistic by the new signal. If these individuals have a prior close to .5, then they will be less uncertain after the new information, even though they have switched from pessimism to optimism, and they will not recall past experiences.

If their prior was slightly less close to .5, but still close enough to have their opinion reversed, then if the costs of recall are sufficiently low and if the benefits of certainty are sufficiently high, then it is possible that they will try to recall information that supports their changed opinion. In this case, memory will reinforce the new information. This effect would suggest that some people overreact to new information because it pushes them to drudge up supporting memories that make their new viewpoint more compelling.

Finally, individuals who begin as diehard pessimists will remain pessimists after the new information is provided to them. The new information will, however, make them less certain and increase the incentives to remember other pessimistic facts. If their costs of recall are low and the benefits of certainty are high, then they will recall a pessimistic memory. This recall will create a memory boomerang as long as the past memory has more statistical power than the new information, i.e. $\lambda > \lambda_E$.

If recall was not targeted, two things would change in the model. The expected benefit of search would decline, because it would not be clear if past memories would shift posteriors towards certainty. If the individual assumes, for example, that recalled memories are as likely to be positive as the share of positive signals in the current stock of recalled beliefs, then that individual would think that the new memory might create less, not more, certainty. The second change is that the ex post change would be less likely to support existing beliefs, because pessimists might recall optimistic memories, which is precluded in our model. These changes would make memory boomerangs less likely, although they would still leave this possibility open.

The endogenous memory model complements the random recollection model and provides similar results. Individuals can diverge if they dislike uncertainty, and if recollected memories are more compelling than new information. If the new information confirms pre-existing views, then certainty only increases, and there is no added incentive to recall forgotten knowledge. For this reason, optimistic people who hear an optimistic story will only get more optimistic. But if the new information counters pre-existing views, then certainty decreases, and some individuals will want to recall memories to become more sure, either to avoid the psychological discomfort of uncertainty of for more instrumental reasons. To get a memory boomerang, the information that is remembered will have to be more powerful than the knowledge that is proferred in the lab, but when that occurs, divergence can follow. Pessimistic people who hear optimistic stories become uncertain and have an incentive to remember. When they remember particularly powerful forgotten facts that support pessimism, then the experiment will actually make them more pessimistic, creating divergence in the population as a whole.

In sum, the new information can increase the incentive to remember for people with views that contradict the information, by introducing more uncertainty (which is assumed to be unpleasant), and the recollected memories more than offset the new information. By contrast, supporting information creates no incentive to remember and as such, supporting information has no added impact on beliefs.

IV. Surprising Validators

Our goal here is explanatory, not prescriptive, but both the empirical findings and the two accounts raise an obvious question: If the goal is to avoid polarization and to produce some kind of reasonable consensus, might anything be done?

The models offer some guidance, and the answer depends on the reason for polarization. If new information creates polarization by activating past memories, then there are real obstacles to eliminating the phenomenon. It is true, however, that a flood of new information should eventually overwhelm memories and the past. If people receive a great deal of information showing that nanotechnology is promising and not at all dangerous, their fears should be overcome, even if those fears are a product of memories of apparently analogous problems. It is hard to believe that we have stored away so much information that memories and associated judgments cannot be countered with sufficient new facts.

Nonetheless, a large volume of information may turn out to have no effect if the first model is at work. In that case, all that extra knowledge may serve only to convince pessimists that they are dealing with the most desperate of deceivers. In that case, the critical factor involves the perceived motives of the communicator. The polarization, in that model, disappears if the communicator is not thought to have ulterior motives and is therefore believed to be trustworthy.

Once, perhaps, mainstream news media may have played the role of honest providers of information, commonly thought to be relatively unbiased. To say the least, that perception is no longer universally held, and different ideological camps hold wildly different views about the reliability of different media groups. (Recall the discussion above about opposing reactions to the New York Times and Fox News.) This problem creates a need for alternative, widely credible sources of information.

In the cases that we have given, the communicator is deemed not to be trustworthy by some members of the target audience. This point is central to our first model. To see the relevance of communicator credibility, consider a striking finding.[32] When liberals and conservatives are asked for their private views about a generous welfare policy and a more stringent one, they react in the predictable ways, with liberals favoring the former and conservatives the latter. But things change dramatically when they are informed of the distribution of views within the House of Representatives. More specifically, conservatives end up disapproving of the more stringent policy, and favor the generous one, when they are told that 90 percent of House Republicans favor the generous policy. Liberals show the same willingness to abandon their private opinions, and thus end up favoring the stringent policy, when told that this is the position of 90 percent of House Democrats. Notably, the effect of learning about party views is as strong among those who are knowledgeable about welfare policy as it is among people who were not.[33] Also notably, both conservatives and liberals believe that their judgments are driven largely by the merits, and not by what they learn about the views of their preferred party – but in that belief, they are wrong.[34]

---

[32] *See* Geoffrey Cohen, Party over Policy: The Dominating Effect of Group Influence on Political Beliefs, 85 J. PERS. SOC. PSYCHOL. 808 (2003).
[33] Id. at 812.
[34] Id. at 811.

In these examples, political parties are themselves acting as surprising validators. The lesson is that if the Democratic Party, as such, favors a position conventionally associated with conservative politics, a large number of Democrats will end up favoring that position as well. And if the Republican Party, as such, favors a conventionally liberal position, a large number of Republicans will shift in that direction too.

Other evidence attests to the power of surprising validators. In the case of nanotechnology, for example, certain communicators are convincing to some members of the target audience while others are not. If a committed environmentalist suggests that nanotechnology is threatening to the environment, those who ordinarily dismiss environmentalists, and consider them to be hopelessly unreliable, are unlikely to be moved. But if exactly the same message is delivered by someone who is more credible to the target audience – say, a well-known skeptic about climate change, or someone whose social background and identity suggest general skepticism about environmentalism and government regulation – it will be far more credible, and it may well end up moving people's opinions.[35]

Or consider the influence of the *convert communicator*, who once believed the opposite of his current position – say, a former member of the National Rifle Association turned gun control advocate, or a former pacifist turned strong defender of a particular war. In one study, a reformed alcoholic was found to be substantially more persuasive than a teetotaler when extolling the importance of abstaining from alcohol.[36] Similarly, the *self-sacrificing validator* is surprising and effective because he is arguing against his apparent self-interest. This effect has been documented in legal contexts – as, for instance, when an attorney admits to a weakness in his case and wins jury votes by virtue of increased credibility.[37] The same effect has been found in political context -- where, for instance, a politician taking a pro-environmental stand turned out to be more persuasive when he was perceived by audience members as generally pro-business.[38]

In light of our discussion, the effect of surprising validators is straightforward to explain. Such validators have special credibility to precisely the people who would otherwise be inclined to dismiss them. If a longtime critic of an allegedly corrupt politician rises to his defense, contending that the corruption charges are baseless, then there is little reason for the kind of polarization that we have explored. Those who are antecedently inclined to believe the charges will be less likely to do so if they are dismissed by someone who does not support that politician. And if a pro-business

---

[35] Kahan, supra note.

[36] John Levine and Ronald Valle, The Convert as a Credible Communicator, 3 *Social Behavior and Personality* 81 (1975). For an interesting but different strategy, which we suspect involves overlapping mechanisms, see Charles Lord et al., Consider the Opposite: A Corrective Strategy for Social Judgment, 47 J Pers and Social Psych. 1231 (1984).

[37] K.O. Williams et al., The Effects of Stealing Thunder in Criminal and Civil Trials, 17 *Law and Human Behavior* 597 (1993).

[38] Alice Eagly et al., Causal Inferences about Communicators and Their Effects on Opinion Change, 36 *Journal of Personality and Social Psychology* 424 (1978).

speaker, typically critical of environmentalists, asserts that nanotechnology is indeed dangerous, the polarization that we have sketched is less likely to take place. Surprising validators are credible, and reduce rather than create polarization, because they counteract asymmetric Bayesianism.

Conclusion

Our principal goals here have been to unify a set of seemingly disparate findings, involving the polarizing effect of information, and to explain why and when they occur. If people have opposing antecedent convictions, they may well react differently to the identical information. Presentation of balanced information may turn out to intensify polarization, simply because different people will believe different parts of the presentation. Assurance that all is well – that a politician is not corrupt, that a war was fought for the right reasons, that a product is safe, that an alarming rumor is baseless – will be credible to some but not to others. As a result, such assurance might well increase polarization.

The same assurance will also make a question or problem salient and activate disparate memories and convictions, potentially producing division when it did not exist before. Surprising validators can be distinctly credible, and when asymmetric Bayesianism is the source of polarization, such validators can reduce or even eliminate the effects that we have described.

Proofs of Propositions

Proof of Proposition 1: The assumption that $c"(1-\pi) < 0$ and the fixed slope of $(1-\varphi\pi)\Delta$ ensures that if the function $c(1-\pi)$ crosses $(1-\varphi\pi)\Delta$ from above once (and hence has a steeper slope), at a value of π, denoted π^{+}_{cross}, it cannot cross $(1-\varphi\pi\Delta$ for any higher values of π and hence for all higher values of π signalling must be strictly optimal. Similarly if the function $c(1-\pi)$ crosses $(1-\varphi\pi)\Delta$ from below at a value of π, denoted π^{-}_{cross}, it cannot cross $(1-\varphi\pi)\Delta$ at any lower value of π, and hence for values of π, below π^{-}_{cross}, signaling must again be strictly optimal. As this implies that there are at most two crossing points, there are also at most three regions of behavior for the signaler.

If $\Delta = 0$, then signaling is never optimal and nothing is inferred from the absence of a signal. If $\Delta > 0$, but small less than $\overline{c}$, then $(1-\varphi)\Delta > c(0) = 0$ and so signaling is optimal for sufficiently high values of π, but $\Delta < \overline{c} = c(0)$ and hence signaling is not optimal for sufficiently low values of π. As $c(1-\pi)$ begins above $(1-\varphi\pi)\Delta$ and ends below $c(1-\pi)$ it can cross $c(1-\pi)$ only once. Hence observers must believe that there exists only a single cross point, denoted π^{+}_{cross}, and signaling occurs for values of π above that point, but not for values of π below that point.

If $\Delta > \overline{c}$, then observers must believe that signaling is optimal for both extremely low values of π and extremely high values of π. This can either mean that $(1-\varphi\pi)\Delta > c(1-\pi)$ for all values of π or that there exist three regions—where signaling is optimal at high and low values of π and not for intermediate levels of π.

Define $\hat{\pi}(\Delta)$ as the value of π that maximizes $c(1-\pi) - (1-\varphi\pi)\Delta$, which will satisfy: $\hat{\pi}(\Delta) = 1 - c'^{-1}(\varphi\Delta)$. There will exist three regions—as opposed to just one—if and only if $c(1-\hat{\pi}(\Delta)) > (1-\varphi\hat{\pi}(\Delta))\Delta$, When $\Delta = \overline{c}$, then the assumption that $lim_{\pi\to 0} c'(1-\pi) = 0$ guarantees that for sufficiently low values of π, $c(1-\pi)$ must lie above $(1-\varphi\pi)\Delta$, since $(1-\varphi\pi)\Delta$,= $c(1-\pi)$ at $\pi = 0$, but the slope of $(1-\varphi\pi)\Delta$, is steeper. This will continue to be the case for values of Δ that are sufficiently close to $\overline{c}$, and as such $c(1-\hat{\pi}(\Delta)) > (1-\varphi\hat{\pi}(\Delta))\Delta$. If Δ is greater than $\overline{c}/(1-\varphi)$, then $(1-\varphi\pi)\Delta$ is greater than $c(1-\pi)$ for all values of π and hence $c(1-\hat{\pi}(\Delta)) - (1-\varphi\hat{\pi}(\Delta))\Delta$, must be negative.

The derivative of $c(1-\hat{\pi}(\Delta)) - (1-\varphi\hat{\pi}(\Delta))\Delta$, with respect to Δ is $-(1-\varphi\hat{\pi}(\Delta))$ which is negative, so as $c(1-\hat{\pi}(\Delta)) - (1-\varphi\hat{\pi}(\Delta))\Delta$, is greater than zero, when $\Delta = \overline{c}$, and less than zero for $\Delta = \overline{c}/(1-\varphi)$, then there must exist a unique value of Δ denoted $\Delta*$ at which $c(1-\hat{\pi}(\Delta)) - (1-\varphi\hat{\pi}(\Delta))\Delta$, and for all values of Δ above $\Delta*$, signaling is always optimal, and for all values of Δ below $\Delta*$, there exist exactly three regions of behavior depending on π. If π is above π^{+}_{cross}, or below π^{-}_{cross}, then signaling is optimal. If π lies between π^{+}_{cross}, and π^{-}_{cross}, signaling is not optimal.

Differentiation gives us that π^{+}_{cross}, is decreasing with Δ and increasing with φ, while π^{-}_{cross}, is increasing with Δ and decreasing with φ. When Δ equals $\Delta*$, then π^{+}_{cross}, must equal π^{-}_{cross}.

Proof of Proposition 2: Following Proposition 1, if v is close to zero then Δ must also be close to zero, and both optimists and pessimists believe that a signal is a sign that π is close to one. If the cutoff value for π is denoted π^{+}_{cross}, then the signal causes an increase in beliefs among optimists equal to $.5(1-\pi_O)$ relative to not receiving the signal. In this case, $\Delta = .5v(1-\pi_O)$, which implies that π^{+}_{cross} satisfies $.5v(1-\varphi\pi cross+1-\pi 0=c(1-\pi cross+)$. Differentiation implies that the cutoff value is declining with v. If v rises to the point where $.5v(1-\pi_O) > \overline{c}$, then observers will believe that the signal could have come from a signaler with a sufficiently low value of π below π^{-}_{cross}, which is defined as in Proposition 1. When v rises to the point where $v > \frac{2c(1-\pi_P).}{(1-\varphi\pi_P)(1-\pi_O)}$, then pessimists will believe that the signal may have come from a signaler with a sufficiently low value of π.

Let $\hat{\pi}(v)$ denote the value of π that maximizes $.5(1-\varphi\pi)v(1-\pi_O) - c(1-\pi)$, and that satisfies $\hat{\pi}(v) = 1 - c'^{-1}\big(.5\varphi v(1-\pi_O)\big)$. As $.5(1-\varphi\pi)v(1-\pi_O) - c(1-\pi)$, is rising with v, and is negative for v close to zero and positive if $v > 2\overline{c}/(1-\varphi)(1-\pi 0$, then there must exist a unique value of v denoted v* . For values of v above v* signaling will be done by all types. The value of v* will satisfy $c\left(c'^{-1}\big(.5\varphi v^{*}(1-\pi 0=.5v*1-\pi 01-\varphi c'-1.5\varphi v*1-\pi 0$. The value of $\pi cross-$ must lie below $\pi v*$ since that is its limit as v goes to v^{*}. As we have assumed that $1 - c'^{-1}\big(.5\varphi v^{*}(1-\pi_O)\big) < \pi_O$, this implies that π^{-}_{cross} is below π_O. As such, the signal either has no impact on optimists' beliefs or it makes them even more optimistic. For values of v below v* and above $\frac{2c(1-\pi_P).}{(1-\varphi\pi_P)(1-\pi_O)}$ pessimists will think that the signal could have come either from a signaler with a very high or a very low value of π, and this can cause them to become even more pessimistic.

Pessimists ex post probability that the state is one $\frac{1-(\pi^{+}_{cross})^2+(\pi^{+}_{cross})^2-(\pi_p)^2}{2(1+\pi^{-}_{cross}-\pi^{+}_{cross}-\pi_p)}$ and this will be lower than their ex ante probability if and only if $\pi^{-}_{cross} + \pi^{+}_{cross} > 1 + \pi_p$. This will obviously not hold when π^{-}_{cross} is close to π_p. When π^{-}_{cross} is close to $\hat{\pi}(v^{*})$, then the condition becomes $2\hat{\pi}(v^{*}) > 1 + \pi_p$. and it is sufficient that π_p is low and $\hat{\pi}(v^{*}) > .5$, which will follow if $c'(.5) < \varphi\overline{c}$. If that condition holds, then the slope $.5\varphi v^{*}(1-\pi_O) > c'(.5)$, since $v^{*} > 2\overline{c}/(1-\pi_O)$, which means $\hat{\pi}(v^{*}) > .5$.

Proof of Proposition 3:

The average ex post belief is : $\frac{(1-m)\lambda_E}{\lambda_E+\lambda^{N_0^R-N_1^R}} + m\left(\left((1-\delta)a+\delta\frac{N_1^R}{N_0^R+N_1^R}\right)\frac{\lambda_E}{\lambda_E+\lambda^{N_0^R-N_1^{R-1}}}+\right.$ $1-1-\delta a-\delta N1RN0R+N1R\lambda E\lambda E+\lambda 1+N0R-N1R$ which equals $\frac{(1-m)}{1+\lambda^{N_0^R-N_1^R}}+$ $m\left(\left((1-\delta)a+\frac{\delta N_1^R}{N_0^R+N_1^R}\right)\frac{1}{1+\lambda^{N_0^R-N_1^{R-1}}}+\left(1-(1-\delta)a-\frac{\delta N_1^R}{N_0^R+N_1^R}\right)\frac{1}{1+\lambda^{1+N_0^R-N_1^R}}\right)$ when $\lambda_E = 1$.
This is greater than $\frac{1}{1+\lambda^{N_0^R-N_1^R}}$ if and only if.

$$(1-\delta)a+\frac{\delta N_1^R}{N_0^R+N_1^R} > \frac{\lambda^{N_1^R-N_0^R}+\lambda}{\left(\lambda^{N_1^R-N_0^R}+1\right)(\lambda+1)}$$

The right hand side is decreasing in $N_1^R - N_0^R$ and runs from $\lambda/(\lambda+1)$ (when $N_1^R - N_0^R$ goes to negative infinity) to $1/(\lambda+1)$ (when $N_1^R - N_0^R$ equals positive infinity). Hence as long as $\frac{N_1^R}{N_0^R+N_1^R} > \frac{1}{\delta}\left(\frac{\lambda}{\lambda+1}-(1-\delta)a\right)$ then the average belief that the state is positive must rise after the signal.

Conversely, as long as $\frac{N_1^R}{N_0^R+N_1^R} < \frac{1}{\delta}\left(\frac{1}{\lambda+1}-(1-\delta)a\right)$, then the average assessment must fall after the new information is revealed. By continuity these conditionals also hold for levels of $\frac{(1-m)\lambda_E}{\lambda_E+\lambda^{N_0^R-N_1^R}}$ that are close enough to zero.

Proof of Proposition 4: Using the formulas derived in the text, the expected belief that the state of the world is one will equal $\frac{(1-m)\lambda_E}{\lambda_E+\lambda^{\Delta}} + m\left(\left((1-\delta)a+\delta\frac{N-\Delta}{2N}\right)\frac{\lambda_E}{\lambda_E+\lambda^{(\Delta-1)}}+\right.$ $\left.\left(1-(1-\delta)a-\delta\frac{N-\Delta}{2N}\right)\frac{\lambda_E}{\lambda_E+\lambda^{(\Delta+1)}}\right)$, which is strictly increasing in λ_E. When $\lambda_E = 1$, then this expected belief is less than pre-intervention beliefs ($\frac{1}{1+\lambda^{\Delta}}$), if and only if $\frac{(\lambda+\lambda^{\Delta})}{(1+\lambda)(1+\lambda^{\Delta})} > \left((1-\delta)a+\delta\frac{N-\Delta}{2N}\right)$. If $\frac{(\lambda+\lambda^{\Delta})}{(1+\lambda)(1+\lambda^{\Delta})} < \left((1-\delta)a+\delta\frac{N-\Delta}{2N}\right)$, then for all values of λ_E greater than one, the intervention raises the average belief that the state of the world is one, because it increases the belief when $\lambda_E = 1$, and the average ex post beliefs increase in λ_E. When $\lambda_E = \lambda$, then $\frac{\lambda_E}{\lambda_E+\lambda^{(\Delta+1)}}$ equals the ex-ante belief and beliefs most always become more positive. As long as $\frac{(\lambda+\lambda^{\Delta})}{(1+\lambda)(1+\lambda^{\Delta})} > \left((1-\delta)a+\delta\frac{N-\Delta}{2N}\right)$, then average ex post beliefs are below ex ante beliefs when $\lambda_E = 1$ and above ex ante beliefs when $\lambda_E = \lambda$, and increasing with λ_E, and as such there must exist a value of λ_E, denoted λ_E^*, before 1 and λ, at which the intervention will not change the expected belief that the state of the world equals one. Monotonicity ensures that for values of of $\lambda_E > \lambda_E^*$, the

intervention will, on average, increase the expected belief that the state of the world is one and for values of $\lambda_E<\lambda_E^*$, the intervention will reduce the expected belief that the state of the world is one.

The value of λ_E^* satisfies

$$\left(\frac{1}{\lambda_E^*+\lambda^{\Delta}}+\frac{m(\lambda^{(\Delta+1)}-\lambda^{\Delta})}{(\lambda_E^*+\lambda^{\Delta})(\lambda_E^*+\lambda^{(\Delta+1)})}+\left((1-\delta)a+.5\delta-\frac{\delta\Delta}{2N}\right)\left(\frac{m(\lambda^{(\Delta+1)}-\lambda^{(\Delta-1)})}{(\lambda_E^*+\lambda^{(\Delta-1)})(\lambda_E^*+\lambda^{(\Delta+1)})}\right)\right)\lambda_E^*(1+\lambda^{\Delta})=1,$$

This can be written $h(\lambda_E^*,Z)=1$, where Z represents the vector of other parameters. For all parameters, it is sufficient to show that h(.,.) is increasing in the variable, to show that λ_E^* is decreasing in the variable (and that h(.,.) is decreasing in the variable to show that λ_E^* is increasing in the variable). As such λ_E^* is decreasing with a and N, holding Δ constant, and increasing with m. λ_E^* is increasing with δ if and only if $a>\frac{N-\Delta}{2N}$. If $\Delta>0$, then this condition must hold. If $\Delta=0$, then $\frac{(\lambda+\lambda^{\Delta})}{(1+\lambda)(1+\lambda^{\Delta})}=.5$ which is less than $(1-\delta)a+.5\delta$, so the condition $\frac{(\lambda+\lambda^{\Delta})}{(1+\lambda)(1+\lambda^{\Delta})}>\left((1-\delta)a+\delta\frac{N-\Delta}{2N}\right)$ is violated. The maximum value of $\frac{(\lambda+\lambda^{\Delta})}{(1+\lambda)(1+\lambda^{\Delta})}$ is $\frac{\lambda}{1+\lambda}$ so as long as $a>\frac{\lambda}{1+\lambda}$, then the starting assumption is violated as well.

Proof of Proposition 5: After the experiment presents the evidence indicating that the state of the world is good, then the initially optimistic group will have even less incentive to recall a past memory, as $\frac{1}{1+\lambda^{\left(N_0^R-1-N_1^R\right)}}-\frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}}>\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-1-N_1^R\right)}}-\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}$, so the ex post belief is $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}>\frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}}$

If $N_0^R>N_1^R>N_0^R-\frac{Log(\lambda_E)}{Log(\lambda)}$, then participants were pessimistic before the experiment, but after the experiment, their posterior will equal $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}>.5$. They will recall an added positive memory after the experiment if $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-1-N_1^R\right)}}-\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}>\frac{\varphi_i}{\beta}$, and it must also be true that $\frac{\varphi_i}{\beta}>\frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}}-\frac{1}{1+\lambda^{\left(N_0^R+1-N_1^R\right)}}$ for it to have been optimal for them not to recall more ex ante. If $N_1^R>N_0^R-.5\frac{Log(\lambda_E)}{Log(\lambda)}$, then this is impossible, but if $N_0^R-.5\frac{Log(\lambda_E)}{Log(\lambda)}>N_1^R>N_0^R-\frac{Log(\lambda_E)}{Log(\lambda)}$, then if $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-1-N_1^R\right)}}-\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}>\frac{\varphi_i}{\beta}$, added memory recall will be optimal and this will only strengthen the impact of the positive signal.

If $N_1^R<N_0^R-\frac{Log(\lambda_E)}{Log(\lambda)}$, then pessimists will remain pessimistic after the new information, but they will be less sure in their pessimism. Their incentive to recall a memory will increase as long as

$\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}-\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R+1-N_1^R\right)}}>\frac{1}{1+\lambda^{\left(N_0^R-N_1^R\right)}}-\frac{1}{1+\lambda^{\left(N_0^R+1-N_1^R\right)}}$ which will always hold in this region. A new memory recall is optimal if $\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R-N_1^R\right)}}-\frac{\lambda_E}{\lambda_E+\lambda^{\left(N_0^R+1-N_1^R\right)}}>\frac{\varphi_i}{\beta}$ and that will cause pessimists to be even more pessimistic ex post as long as $\lambda>\lambda_E$.