RATIONALITY
& COMPETITION
CRC TRR 190

A Model of Ideological Thinking

Yves Le Yaouanq (LMU)

Discussion Paper No. 85

March 19, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

A model of ideological thinking

Yves Le Yaouanq [*]

Abstract

This paper develops a theory in which heterogeneity in political preferences produces a partisan disagreement about objective facts. A political decision involving both idiosyncratic preferences and scientific knowledge is considered. Voters form motivated beliefs in order to improve their subjective anticipation of the future political outcome. In equilibrium, they tend to deny the scientific arguments advocating the political orientations that run counter to their interests. Collective denial is the strongest in societies where contingent policy is the least likely to be implemented, either because of voters' intrinsic preferences or because of rigidities in the political process. The theory predicts that providing mixed evidence produces a temporary polarization of beliefs, but that disclosing unequivocal information eliminates the disagreement.

Keywords: beliefs, ideology, cognition, disagreement, polarization
JEL Codes: D72, D81, D83, D84, Z13

,

[*] Ludwig-Maximilians-Universität, Munich. E-mail: yves.leyaouanq@econ.lmu.de. A previous draft of this paper was circulated under the title "Political values and the polarization of beliefs". I am deeply indebted to Roland Bénabou, Christian Gollier and Jean Tirole for their guidance, advice and encouragement. I am also grateful to Florian Englmaier, Yann Kervinio, Eric Mengus, Elena Panova, Sébastien Pouget, Christoph Rheinberger, Andries Richter, Anton Suvorov, Nicolas Treich and Karine Van der Straeten, and seminar participants at TSE, Maastricht Behavioral and Experimental Economics Symposium, Toulouse TIGER Forum, EAERE and EEA for insightful comments and discussions. I thank the Department of Economics of Princeton University for its hospitality and I am grateful to the *Corps des Mines* and to the *Deutsche Forschungsgemeinschaft* (through CRC TRR 190) for financial support.

1 Introduction

Standard theories of information processing predict that disagreement about objective facts between laypeople should decrease as the knowledge produced by scientists becomes disseminated in the population. However, some socio-economic and scientific issues are still fervently debated in spite of a large consensus among experts. Climate change offers an important example. While the scientific community has become convinced of the causal influence of human activities on the climate, a substantial fraction of the population remains skeptical about the validity of the theory of anthropogenic climate change.[1] Strikingly, the disagreement seems to be mostly driven by individuals' political orientation: the controversy brings into opposition liberals, a majority of whom accepts the scientific evidence, and conservatives, who tend to reject it (Dunlap and McCright, 2008, 2011).[2] A partisan disagreement between the left and the right of the political spectrum is observed in a wide range of areas, such as risk perception (e.g., nuclear power, genetically modified organisms), economic issues (e.g., effects on the labor market of government intervention) or judicial policies (e.g., effectiveness of gun control policies).

While the existence of a partisan disagreement is consistent with a myriad of possible theories, several facts seem incompatible with the standard model of information processing, according to which individuals behave as dispassionate statisticians and try to form an accurate assessment of uncertain variables. First, some experimental evidence indicates that liberals and conservatives not only hold different opinions on scientific issues but also react differently to the disclosure of balanced information. In some cases, providing mixed evidence leads to polarizing the average opinions of both groups instead of to reducing the disagreement (Lord et al., 1979; Plous, 1991; Munro and Ditto, 1997). Second, some experimental manipulations that vary the perceived policy implications associated with the

[1] Surveys of climate scientists (Anderegg et al., 2010; Farnsworth and Lichter, 2012) find that the proportion of dissenters lies between 1% and 5%. In contrast, according to a Gallup survey, only 57% of Americans subscribe to the theory of human-induced climate change (Saad, 2014).

[2] In 2013, 78% of Democrats and 39% of Republicans agreed with the theory of anthropogenic climate change (Saad, 2014). Among Republicans, the skepticism is so strong that only 45% believe that the climate is warming (Pew Research Center, 2010).

signals without affecting their informativeness have been shown to influence the gap between liberals' and conservatives' beliefs about the underlying science (Feinberg and Willer, 2010; Braman et al., 2012; Campbell and Kay, 2014).

This paper aims at providing a unified explanation for the prevalence of partisan disagreement about scientific issues and the anomalous updating behavior observed experimentally. It argues that many of the above facts can be understood through the lens of motivated cognition, and analyzes the formation of ideologically motivated beliefs in a political economy context. The theory is based on well-documented psychological phenomena, and predicts that political preferences causally influence information processing. In contrast to existing explanations, disagreement about scientific issues does not result from heterogeneity in prior beliefs or private information but reflects an underlying conflict about the policy implications of the scientific arguments.

Section 2 describes the environment. A continuum of voters makes a binary decision $a \in \{0, 1\}$, e.g., regulating an industrial activity ($a = 0$) or not ($a = 1$), in a situation of binary uncertainty, e.g., whether the activity is polluting ($\omega = L$) or not ($\omega = R$). The laissez-faire policy produces a common net benefit equal to x_ω, which is positive in state R and negative in state L. In addition to this common payoff, voters have heterogeneous preferences regarding the political decision. Each citizen is characterized by a preference parameter v that describes their intrinsic ordering of the courses of political action a. For instance, some individuals oppose government intervention either because of their material interests or on ideological grounds, whereas the opposite part of the spectrum is likely to support regulation. In state ω, and given the political outcome a, the payoff to a voter of type v equals $(v + x_\omega)a$: the voter's preferences for political decisions are state independent if $|v|$ is large enough, but state dependent otherwise, in which case public information might influence voting behavior.

All voters hold the same prior beliefs and receive a common public signal correlated with ω. This assumption allows us to attribute any posterior disagreement to differences in information processing. The main assumption of this paper is that individuals have some "cognitive wiggle room" to interpret the public signal. In line with the experimental evidence on wishful thinking, they have the opportunity to discard the signal (at some cost) if doing so makes them more optimistic about their future payoff. As a result, for instance, individuals endowed with a large v, namely a strong aversion to governmental regulation, have an incentive to cast aside the signals advocating risk-prevention policies.

The first result, presented in Section 3, is the existence of a partisan disagreement driven by personal ideology in any equilibrium resulting from individually rational decisions. In equilibrium, an agent's interpretative strategy and posterior beliefs are pinned down by the agent's type v, leading politically opposed groups to disagree on the likelihood of the state ω even in the face of common information. The evidence that documents an environmental threat and suggests public preventative measures is rejected by voters who oppose regulation, but accepted by the rest of the political spectrum. Conversely, voters who have a vested interest in public regulation tend to deny the evidence that substantiates the costs of government intervention. Belief distortion arises on both sides of the political spectrum, and since updating behavior is monotonic in v, individuals with the strongest preferences over the possible courses of political action are the most likely to resort to motivated reasoning, a result that lines up well with the recent empirical findings of Ortoleva and Snowberg (2015).

This model makes specific predictions about the factors that favor the emergence of ideological denial, which would not play any role in a standard model of information processing without motivated beliefs. More precisely, the model shows that risk denial is strongest in societies where regulation is least likely to be implemented, either because a large and influential group of citizens oppose regulation on ideological grounds, or because the political system is affected by a status quo bias that makes new policies difficult to implement. In both cases, the low likelihood of an appropriate political response makes it optimal for all voters to deny the costs associated with unregulated risk. This result might, for instance, explain why skepticism vis-à-vis climate change has been particularly strong in the United States, where a large part of the electorate is intrinsically reluctant to let the government intervene in the economy, and where institutions are sometimes considered as "gridlocked," making the enactment of new policies difficult.

The second set of results, presented in Section 4, analyzes the conditions under which the model predicts a polarization of beliefs after the disclosure of common information. Even though individuals resort to motivated reasoning, they face constraints that limit their capacity to form preference-consistent opinions. The theory delivers several testable predictions regarding the type of information that leads to polarization and disagreement between politically opposed groups, some of which are substantiated by the experimental literature on disagreement about political issues. First, the opinions of politically opposed voters polarize if the common sequence of signals contains arguments in favor of both positions, something which enables voters to discredit preference-inconsistent signals and incorporate preference-consistent evidence in their beliefs. The experimental evidence on polarization relies on such mixed signals. For instance, in the experiment by Lord et al. (1979), participants' opinions polarize after they read press articles containing evidence both in favor and against the death penalty. In contrast, the theory does not predict any polarization if the groups receive clear-cut evidence in favor of one position: in this case, individuals are unable to find arguments to rationalize their preferred cognition, and Bayesian constraints force them to update in the right direction. Second, the disagreement vanishes if individuals receive unequivocal evidence, such as an infinite sequence of unbiased signals, which impedes their ability to self-deceive and eliminates any anomaly in updating. The model thus predicts that a partisan disagreement can be observed only temporarily, for issues for which plausible arguments are put forward by both camps.

The main contribution of this paper is to provide psychological foundations for the existence of partisan disagreement about objective facts, and to offer a theory that encompasses a wide range of experimental and empirical evidence. While motivated cognition has already been pointed to by several scholars as a possible explanation for partisan disagreement (for instance Sunstein, 2001), this paper is the first to give formal foundations for this explanation, thereby generating precise predictions about the nature of such disagreement. That liberals and conservatives are motivated and able to deny different types of signals is not assumed as a primitive proposition but derived from fundamental and measurable psychological ingredients (motivated cognition, wishful thinking and heterogeneous preferences over the political decisions) that have already been documented and incorporated into the economics literature (Bénabou and Tirole, 2006, 2011; Bénabou and Tirole, 2016). Moreover, while existing theories of disagreement and polarization (e.g., Rabin and Schrag, 1999) take heterogeneous prior beliefs as a starting point and only explain the divergence of beliefs between groups who already disagree, the present model predicts that liberals and conservatives react differently to the first piece of information that they encounter even if they start with common prior beliefs. This feature allows predicting the determinants of disagreement about novel and unfamiliar topics, such as nanotechnologies or geo-engineering. Lastly, in contrast with existing models, the theory makes predictions about the types of societies and topics for which inaccurate beliefs (here, driven by ideological denial) are the most likely to arise.

The theory relies on individuals' desire and capacity to forge illusions and repress inconvenient truths at the service of their emotional needs. While standard models predict that signals should be assessed according to their informativeness alone, evidence from several fields shows that individuals' updating behavior is influenced by their needs and desires, a phenomenon usually referred to as "motivated cognition" (Festinger, 1957; Kunda, 1990; Bénabou and Tirole, 2016). More precisely, in this model, individuals' reaction to information is affected by their desire to form optimistic beliefs about their future prospects, in line with the widespread evidence of unrealistic expectations about future life events (Weinstein, 1980) and wishful thinking (Caplin and Leahy, 2001; Mijovic-Prelec and Prelec, 2010; Mayraz, 2011). Motivated cognition has been incorporated into several economic models of belief formation (Brunnermeier and Parker, 2005; Köszegi, 2010; Bénabou and Tirole, 2011; Bridet and Schwardmann, 2017; Gottlieb, 2018, among others). The contribution of the present paper relative to this literature is to incorporate this psychological foundation into a model of voting, and to argue that wishful thinking provides a unifying explanation for the phenomena of partisan disagreement and polarization. The closest existing paper is Bénabou (2013), who analyzes collective reality denial on the part of individuals engaged in a joint project. Other models of wishful thinking in the political context applied to different issues are provided by Bénabou (2008) and Levy (2014).

Section 1.1 reviews the empirical evidence on partisan disagreement and polarization, with a particular emphasis on the evidence that speaks in favor of a theory involving motivated beliefs. Section 2 introduces the model. Section 3 proves the existence of an equilibrium in a general case, analyzes the factors that favor the emergence of collective denial, and discusses alternative theories for the existence of a partisan disagreement. Section 4 analyzes the conditions under which the model predicts a polarization of beliefs between different social groups, and discusses alternative theories of learning and polarization. Section 5 draws some conclusions and outlines avenues for future research. All proofs are in the Appendix.

1.1 Empirical evidence

The disagreement between laypeople and experts, and between laypeople, has been documented in many important areas, and has been the focus of numerous studies in the social sciences, in particular in the field of risk perception. While early explanations highlighted cognitive limitations and the use of inappropriate heuristics (Breyer, 1995; Slovic, 2000), recent theories and observations have drawn attention to the role of political preferences (see for instance the Cultural Theory of Douglas and Wildavsky, 1982; Douglas, 1994). Liberals and conservatives indeed consistently disagree about the scientific arguments pertaining to several important socio-economic issues, such as climate change (Dunlap and McCright, 2011), nuclear power (Jenkins-Smith et al., 2011), nanotechnologies (Kahan et al., 2009), gun control (Kahan, 2012), or stem cell research (Nisbet, 2005). Remarkably, the direction of the disagreement is consistent across these topics, as liberals always perceive greater objective risks, and thus greater benefits from regulation, than conservatives.

One remarkable experimental finding is that the provision of balanced information on a controversial issue does not always reduce the disagreement between politically opposed groups but sometimes aggravates it, as participants' beliefs become more extreme in their initial direction after the provision of mixed evidence. In a classic study, Lord et al. (1979) exposed proponents and opponents of capital punishment to an identical collection of research findings containing evidence both in favor and against the deterrent effect of the death penalty. Strikingly, the provision of balanced information not only polarized the individuals' attitudes toward the death penalty (which might be perfectly consistent with the use of Bayes' rule), it also strengthened the disparity of views regarding the objective effects of capital punishment: proponents became more convinced that capital punishment deters crime, whereas opponents became more convinced that the death penalty is ineffective. This finding contradicts some basic properties of Bayesian information processing, according to which the provision of common information should reduce the discrepancy between views.

Later experiments replicated this finding and extended it to several controversial social issues, such as nuclear power (Plous, 1991) and affirmative action (Munro and Ditto, 1997). In the case of climate change, the partisan gap has been growing between 2000 and 2010 while the scientific community was producing more evidence of the link between human activities and the climate (Dunlap and McCright, 2011).[3]

The correlation between political attitudes and scientific opinions, and the phenomenon of polarization, admit several plausible explanations, reviewed in subsections 3.5 and 4.4. However, a set of experimental results tend to corroborate the role of motivated reasoning and the causal effect of political attitudes on beliefs, which are at the core of the present paper. Whereas the canonical model of learning prescribes that posterior beliefs depend only on the prior and on the information contained in the signal, experimental evidence indicates that the formation of beliefs is affected by the perceived political consequences of the information. As an illustration, Feinberg and Willer (2010) show that people are more likely to believe in the conclusions of mainstream climate science when the information delivers potential solutions (emissions policies, technical innovations, etc.) than when the message insists on the dire consequences of untreated climate change.

Interestingly, and consistently with the theory developed in the paper, the partisan gap is itself affected by the perceived policy implications. In the experiment by Campbell and Kay (2014), participants read press ar-

[3] From 2003 to 2013, the fraction of liberals who subscribed to climate science rose from 68% to 78%, whereas it decreased from 52% to 39% among conservatives (Saad, 2013).

ticles that documented the scientific consensus on climate change and discussed potential solutions. In one treatment (“free-market”), the proposed solution was that the US become the world leader in green industries without any damage to its economy. In the other treatment (“governmental regulation”), the proposed solution was that the US implement mitigation policies. Politically conservative participants were more likely to believe in the theory of anthropogenic climate change if under the first treatment, namely, if the suggested policy implication fit their political preferences. In the same vein, the partisan gap was attenuated when geo-engineering was presented as a potential solution, which reduced skepticism among individuals who opposed regulation (Braman et al., 2012). Several experiments have shown that framing the information so as to minimize the ideological implications influences beliefs and attitudes. For instance, conservatives are more likely to endorse a Pigovian instrument presented as a “carbon offset” rather than as a “carbon tax” (Hardisty et al., 2010).

A second indication of the existence of motivated beliefs is that individuals seem prone to reject information that contradicts their ideology (Kahan et al., 2009). For instance, in the experiment by Nyhan and Reifler (2010), participants in the treatment group read a summary of the Duelfer report, whose main conclusion is that Iraq did not have an active program of weapons of mass destruction in 2003. The treatment worked in the expected way for liberals, who were less likely to believe in the existence of the program after reading the summary, but in the opposite way for conservatives. Conversely, liberals who believed that President Bush banned stem cell research do not revise their beliefs when they receive evidence to the contrary. On a related note, people form their opinions about climate change in light of their personal experience of the climate, but this perception is itself politicized: Democrats tend to believe that the temperatures in their area have recently been warmer than in the past, while Republicans believe the opposite (Goebbert et al., 2012; Akerlof et al., 2013). These observations substantiate the causal link from preferences to cognition, which is at the core of this paper.

2 Environment

Payoff structure The economy is composed of a continuum of agents of measure one who have to make a collective decision $a \in \{0, 1\}$: mitigating greenhouse gas emissions, instituting gun control, banning research on stem cells, etc. The decision $a = 1$ is interpreted as the status quo (laissez-faire policy), whereas $a = 0$ corresponds to a risk-prevention policy. The political decision involves some scientific uncertainty, summarized by a state variable $\omega \in \{L, R\}$, uniformly distributed. The payoff-relevant variable X equals $x_R > 0$ in state $\omega = R$ and $-x_L < 0$ in state $\omega = L$. Let

$$x_0 := \frac{1}{2}x_R - \frac{1}{2}x_L$$

be the (common) ex-ante expected value of X.

Once the decision is made, all citizens receive their payoff composed of a common term equal to Xa and of an idiosyncratic term equal to va. The parameter v captures an agent's material or ideological preference regarding the political decision itself: citizens endowed with $v > -x_0$ tend to support the status quo whereas those endowed with $v < -x_0$ have an intrinsic taste for regulation. The variable v is distributed on $\mathbb{R}$ according to an atomless and continuous pdf f. Table 1 summarizes the payoff matrix for an individual of type v conditional on the state and on the political decision.

	$\omega = R$	$\omega = L$
$a = 1$ (laissez-faire)	$v + x_R$	$v - x_L$
$a = 0$ (regulation)	0	0

Table 1 – Payoff matrix for a voter of type v

The state ω is thus irrelevant to social welfare if risk prevention policies are enacted. This assumption entails some loss of generality, since most regulation instruments (e.g., carbon tax, emissions market, command-and-control policies) reduce the externality but do not eliminate it entirely. This assumption is nevertheless maintained in order to simplify the mathematical expressions, and does not affect the main results. In addition, it can be considered as an approximation to the fact that the difference in payoffs between the states L and R is lower under regulation than under laissez faire.

Information At $t = 0$, all voters receive a public informative signal m about the state. The signal takes only two values: L ("bad news") or $\varnothing$ ("no news"). This assumption is inessential to the main results, and a symmetric signal structure would yield similar predictions (see Section 4). The availability of public information is measured by the parameter λ while its quality is measured by π: conditional on $\omega = L$, the signal L is sent with probability $\lambda\pi$; conditional on $\omega = R$, the signal L is sent with probability $\lambda(1-\pi)$. By Bayes' rule,

$$\begin{cases} X_L := & \mathbb{E}[X \mid m = L] = (1-\pi)x_R - \pi x_L \\ X_\varnothing := & \mathbb{E}[X \mid m = \varnothing] = \dfrac{1-\lambda(1-\pi)}{2-\lambda}x_R - \dfrac{1-\lambda\pi}{2-\lambda}x_L. \end{cases} \tag{1}$$

The parameters satisfy $\pi > 1/2$ and $\lambda > 0$, which implies that $X_L < x_0 < X_\varnothing$. The signal L is to be interpreted as evidence in favor of regulation, whereas $\varnothing$ gives support to laissez faire.

Voting The vote takes place at $t = 1$: each agent selects $a \in \{0, 1\}$. The probability with which a political decision is implemented is continuously increasing in the number of citizens who express this preference. For instance, if legislative power is allocated according to proportional representation, the likelihood with which regulatory policies are put in place increases with the number of representatives who support it. Formally, if ν is the fraction of citizens who choose $a = 1$, the probability of implementing $a = 1$ is equal to $\phi(\nu)$, where ϕ is a continuously differentiable function such that $\phi' > 0$.

Since there is a continuum of voters, individual voting decisions are inconsequential, for no citizen is ever pivotal. As usual for large elections, voters who care only about the political outcome have no strict incentive to vote. Consistently with theories of expressive voting in large elections, citizens derive some intrinsic utility from voting according to their true preference. Hence, a citizen of type v and whose subjective expectation equals $\mathbb{E}X$ chooses $a = 1$ if and only if $v + \mathbb{E}X \geq 0$.[4] The thresholds $-X_\varnothing$ and $-X_L$ separate the electorate into three parts: the left-wing fraction characterized by $v < -X_\varnothing$ chooses $a = 0$ in all states, the right-wing fraction defined by $v \geq -X_L$ chooses $a = 1$ in all states, while the remaining citizens have state-contingent political preferences and might vote for either policy depending on their beliefs.

Individuals are marginal in the vote. As a consequence, they have no incentives to form precise beliefs when they choose whether to deny the evidence or not. The theory concerns large elections where individuals vote for reasons distinct from instrumental concerns (social norms, the intrinsic utility of expressing one's opinion, social pressure, etc.) as documented in, for instance, Coate et al. (2008) and DellaVigna et al. (2016). Adapting the theory to a small group where each individual has a non-negligible probability of being pivotal would require taking into account this extra incentive as a limit for self-deception.

Formation of beliefs The driving force behind motivated reasoning is that voters form expectations about their future prospects and derive anticipatory feelings from it. At date 1, prior to the vote, an agent of type v contemplates the future political outcome and receives a flow of anticipatory utility equal to $s\mathbb{E}[Xa + va]$. Self-deception consists in creating some illusory optimism regarding future public decisions by distorting one's beliefs about X and a.

Several modeling strategies are possible to reflect the distortion in the cognitive process, which are all equivalent provided that they allow for asymmetric awareness of the signals L and $\varnothing$. The present paper follows the memory management model proposed by Bénabou and Tirole.[5] At date $t = 0$, the agent can influence the information recalled at $t = 1$. An agent who receives bad news ($m = L$) can repress this information and encode $\hat{m} = \varnothing$ at a cost $c > 0$. In contrast, an agent who does not receive any information ($m = \varnothing$) cannot forge a signal between periods 0 and 1 and

[4] The behavior of the agents who are indifferent between $a = 0$ and $a = 1$ does not matter, since they are marginal and have no impact on the political decision. To avoid discussing mixed strategies, the model is specified by assuming that they vote for $a = 1$ with probability 1.

[5] See Bénabou and Tirole (2002, 2011); Bénabou (2013).

always transmits $\hat{m} = \varnothing$.

An equilibrium cognitive strategy is a probability σ^* with which the agent truthfully transmits $\hat{m} = L$ conditional on $m = L$. This model is a metaphor for the diverse strategies that individuals can employ to bias their memory or awareness of the facts in their preferred direction: paying more attention to certain news than to others, rationalizing preference-inconsistent signals, etc.

In existing models involving anticipatory utility, decision makers trade off the pleasure of forming rosy beliefs about their future prospects against the costs of the suboptimal decisions that this distortion creates. In the setting of the present model, in contrast, each citizen individually has no influence on the political outcome. As a consequence, citizens do not take into account any benefit of remaining well informed when they choose their cognition, and the only force that counterbalances their desire to distort reality is the cost c associated with cognitive manipulations.

Meta-cognition The equilibrium concept requires that each agent's cognitive strategies result from an intra-personal information game. This has two implications: first, self 1 is sophisticated and does not take the message $\hat{m} = \varnothing$ at face value but infers a posterior probability for the messages L and $\varnothing$ as a function of self 0's equilibrium behavior σ^*. Relaxing this hypothesis and assuming that self 1 does not compute Bayes' rule after recollecting $\hat{m} = \varnothing$ reinforces the results of the paper since self-deception is less constrained in that case. The sophistication hypothesis constitutes a conservative benchmark under which natural properties of Bayesian updating (in particular, the law of iterated expectations) are preserved. A second implication is that self 0 finds it optimal to play the equilibrium action σ^* conditional on receiving the signal L: agents cannot commit to a strategy ex ante but react optimally ex post after seeing the information.

Timeline At date 0, all citizens learn their type v and the state of the world ω is realized. They receive the public message m. If $m = L$, they choose the probability with which they transmit it ($\hat{m} = L$) or conceal it ($\hat{m} = \varnothing$). At date 1, they form their recollection $\hat{m}$, update their beliefs about m and ω, derive their anticipatory utility $s\mathbb{E}[Xa + va]$, and vote.

The political outcome and the resulting payoffs are realized at date 2.

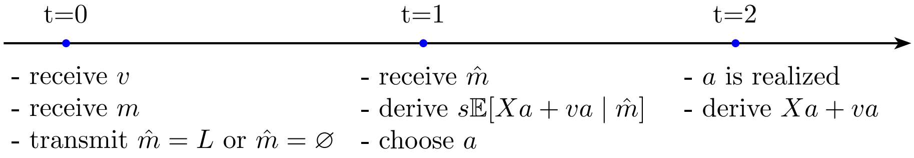

Figure 1 – Timeline

3 Partisan disagreement

3.1 Equilibrium concept

It will become apparent later that all equilibria are symmetric in the sense that all individuals endowed with the same v play the same equilibrium cognitive strategy $\sigma^*(v)$. A Perfect Bayesian Equilibrium of the game consists of a profile of strategies $\Sigma^* = \{\sigma^*(v)\}$ for $v \in \mathbb{R}$ and of a profile of voting decisions $\{a(v, \hat{m})\}$ for $v \in \mathbb{R}, \hat{m} \in \{L, \varnothing\}$ such that:

(i) For all $v \in \mathbb{R}$, $\sigma^*(v)$ belongs to

$$\arg\max_{\sigma \in [0,1]} \sigma s\mathbb{E}[Xa+va \mid \hat{m} = L, \Sigma^*] + (1-\sigma)\big(s\mathbb{E}[Xa+va \mid \hat{m} = \varnothing, \Sigma^*] - c\big).$$

(ii) For all $v \in \mathbb{R}$, $\hat{m} \in \{L, \varnothing\}$, $a(v, \hat{m})$ belongs to

$$\arg\max_{a \in \{0,1\}} \mathbb{E}[X \mid \hat{m}, \sigma^*(v)]a + va.$$

The expectations that depend on Σ^* are conditioned both on the agent's own equilibrium cognitive strategy and on the other citizens' strategies. The agents' expected payoffs depend on the whole strategy profile Σ^* in two ways. First, an agent's own cognitive strategy influences their posterior beliefs, due to the sophistication hypothesis. Second, other agents' cognitive strategies influence their vote, and thus the distribution of policy outcomes, which enters the anticipatory utility term.

The analysis of the equilibrium proceeds in two steps: first, solving for the individual best response holding fixed the other voters' behavior, and then finding a complete equilibrium by means of a fixed-point argument.

3.2 Intra-personal equilibrium

The behavior of other players matters only insofar as it influences the relative likelihood of the political outcomes $a = 1$ and $a = 0$. Let ν_L be the fraction of citizens who vote for $a = 1$ if the public message L has been sent; similarly, let $\nu_\varnothing$ be the fraction of citizens who vote for $a = 1$ if the public message $\varnothing$ has been sent. As shown in the Appendix, $\nu_L \leq \nu_\varnothing$ in all equilibria: regulatory policies receive more support when convincing evidence is presented.

Consider an individual of type v (fixed) who receives the signal L at date 0. Let us define

$$U_1 \left[m = L\right] := \mathbb{E}\left[Xa + va \mid m = L\right] = (v + X_L)\phi(\nu_L) \tag{2}$$

$$\text{and } U_1 \left[m = \varnothing\right] := \mathbb{E}\left[Xa + va \mid m = \varnothing\right] = (v + X_\varnothing)\phi(\nu_\varnothing). \tag{3}$$

The expression $U_1[m]$ represents the agent's anticipatory utility at date 1 conditional on being certain that the message m was sent.

Equations 2 and 3 yield

$$\begin{aligned} \mathcal{I}(v) &:= U_1[m = \varnothing] - U_1[m = L] \\ &= \underbrace{v(\phi(\nu_\varnothing) - \phi(\nu_L))}_{\text{idiosyncratic incentive}} + \underbrace{X_\varnothing\phi(\nu_\varnothing) - X_L\phi(\nu_L)}_{\text{common incentive}} \end{aligned} \tag{4}$$

where $\mathcal{I}(v)$ represents the individual's benefit from concealing the bad news L. Two forces influence this expression. The idiosyncratic incentive is related to the state-independent part of the payoffs. This term is nonnegative as long as $v \geq 0$: right-wing citizens benefit from encoding $\varnothing$ since it increases their perceived probability that the decision $a = 1$ will be chosen. The common incentive is related to the state-contingent part of the utility: it is always nonnegative. $\mathcal{I}(v)$ is nondecreasing in v: citizens with a stronger aversion to regulation have higher incentives to deny the risks. This observation drives the monotonicity of the cognitive strategy in v displayed in any equilibrium.

In the following, $\sigma^*(v)$ represents the agent's tentative equilibrium cognitive strategy, and σ denotes the choice variable following the reception of a signal equal to L. The analysis consists in finding the unique value of $\sigma^*(v)$ which is indeed a best response to the equilibrium strategy $\sigma^*(v)$ itself. Let

$$\begin{aligned}\mu(\sigma^*(v)) &\coloneqq \mathbb{P}[m = L \mid \hat{m} = \varnothing, \sigma^*(v)] \\ &= \frac{(1-\sigma^*(v))\lambda}{2-\lambda\sigma^*(v)}\end{aligned} \tag{5}$$

be the ex post probability attached to the public message L by an agent who recollects $\hat{m} = \varnothing$ given the equilibrium (habitual) cognitive strategy $\sigma^*(v)$.

If the agent truthfully encodes $\hat{m} = L$, self 1 puts probability 1 on the hypothesis $m = L$, which yields the utility

$$U_1[\hat{m} = L \mid \sigma^*(v)] = U_1[m = L]. \tag{6}$$

If, in contrast, the individual represses the signal and encodes $\hat{m} = \varnothing$, self 1 puts probability $\mu(\sigma^*(v))$ on the hypothesis $m = L$, and $1 - \mu(\sigma^*(v))$ on the hypothesis $m = \varnothing$, which yields

$$\begin{aligned}U_1[\hat{m} = \varnothing \mid \sigma^*(v)] &= \mu(\sigma^*(v))U_1[m = L] + [1 - \mu(\sigma^*(v))]U_1[m = \varnothing] \\ &= \frac{(1-\sigma^*(v))\lambda}{2-\lambda\sigma^*(v)}U_1[m = L] + \frac{2-\lambda}{2-\lambda\sigma^*(v)}U_1[m = \varnothing]\end{aligned} \tag{7}$$

by Equation 5. The agent's choice of the probability of truthful transmission σ maximizes the date 0 intertemporal utility conditional on $\sigma^*(v)$, given by

$$U_0[\sigma \mid \sigma^*(v)] = \sigma s U_1[\hat{m} = L \mid \sigma^*(v)] + (1-\sigma)\big(sU_1[\hat{m} = \varnothing \mid \sigma^*(v)] - c\big). \tag{8}$$

Substituting 6 and 7 into 8 yields

$$U_0[\sigma \mid \sigma^*(v)] = (1-\sigma)\big(\frac{2-\lambda}{2-\lambda\sigma^*(v)} s\mathcal{I}(v) - c\big) + sU_1[m = L].$$

A best response is a fixed point of the equation $\sigma^*(v) \in \arg\max_\sigma U_0[\sigma \mid \sigma^*(v)]$. Three cases arise:

- If $s\mathcal{I}(v) \leq c$, $\sigma^*(v) = 1$.

- If $s\mathcal{I}(v) \geq \frac{2c}{2-\lambda}$, $\sigma^*(v) = 0$.
- Otherwise, $\sigma^*(v)$ is (uniquely) defined by $s\frac{2-\lambda}{2-\lambda\sigma^*(v)}\mathcal{I}(v) = c$.

If $\mathcal{I}(v) \leq 0$, the agent prefers the state where $m = L$ to the state where $m = \varnothing$ and therefore never suppresses the signals equal to L. If $\mathcal{I}(v) > 0$, the agent's recall rate is a declining function of the anticipatory term s. In any case, there is a unique equilibrium characterized by the equilibrium strategy $\sigma^*(v)$. Since $\mathcal{I}(v)$ is a nondecreasing function of v, it is easy to see that $\sigma^*(v)$ is nonincreasing in v: the more an agent opposes regulation, the more that agent is likely to deny any news that advocates risk-prevention policies. Lemma 1 summarizes this result.

Lemma 1. *Given $(\nu_L, \nu_\varnothing)$, the best cognitive response $\sigma^*(v)$ of an agent of type v is unique, nonincreasing in v, and nonincreasing in s.*

3.3 Inter-personal equilibrium

The equilibrium concept requires that individual cognitive strategies are derived according to Lemma 1 and that the values of $\nu_\varnothing$ and ν_L are correctly anticipated by all players. Fix a profile of political outcomes $\nu = (\nu_\varnothing, \nu_L)$. Consider the profile of cognitive strategies $(\sigma^*(v,\nu))_{v\in\mathbb{R}}$ implied by the political outcomes ν according to Lemma 1, and the following function.

$$a(v, \hat{m}, \nu) = \begin{cases} 1 & \text{if } v + \mathbb{E}[X \mid \hat{m}, \sigma^*(v,\nu)] \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

This function describes the vote of individuals of type v who recollect the signal $\hat{m}$ given their cognitive equilibrium strategy $\sigma^*(v,\nu)$. The political outcomes resulting from these individual strategies are

$$g_\varnothing(\nu) = \int a(v, \varnothing, \nu)dF(v)$$

if $m = \varnothing$, and

$$g_L(\nu) = \int [\sigma^*(v,\nu)a(v,L,\nu) + (1 - \sigma^*(v,\nu)a(v,\varnothing,\nu)]dF(v)$$

if $m = L$. An equilibrium of the game is characterized by a fixed point of the mapping $\nu \to (g_{\varnothing}(\nu), g_L(\nu))$ and by the associated individual strategies $\{\sigma^*(v, \nu), a(v, \hat{m}, \nu)\}$. The existence of an equilibrium is verified in the Appendix as an application of Brouwer's fixed-point theorem.

Proposition 1. *There exists an equilibrium of the game. In any equilibrium, voters form partisan beliefs: there exist some thresholds*[6] $(v_- \leq v_+)$ *such that* $\sigma^*(v) = 1$ *if* $v \leq v_-$, $\sigma^*(v)$ *is a linearly decreasing function of* v *if* $v \in [v_-, v_+]$, *and* $\sigma^*(v) = 0$ *if* $v \geq v_+$.

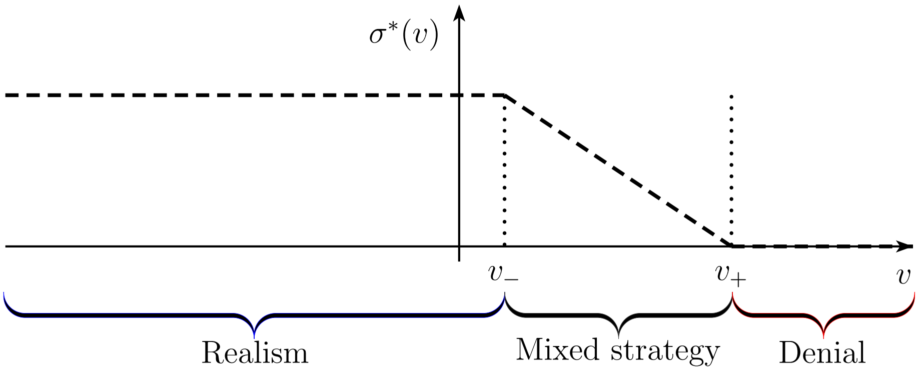

Figure 2 – Equilibrium cognition

The model therefore predicts a causal link from measurable political preference parameters to the contingent beliefs formed about policy-relevant scientific topics. The main prediction is that agents who have the greatest stake against government intervention are the most likely to deny the evidence that calls for regulation.

The fact that agents with a larger v are the most prone to wishful thinking, while agents with a low v form accurate beliefs, is an artifact of the asymmetric signal structure considered. In a model with a symmetric signal structure, the evidence that advocates against regulation would similarly be denied by the citizens at the left of the political spectrum, who have an intrinsic preference for the decision $a = 0$; the model predicts that these individuals are likely to form overly optimistic beliefs about the efficiency of government intervention, for instance in economic policies. Such

[6] The thresholds can take infinite values if σ^* is constant in v, which is the case when $\nu_{\varnothing} = \nu_L$ in equilibrium. For instance, it is possible that $\sigma^*(v) = 1$ for all $v \in \mathbb{R}$, which is captured by the case $v_- = v_+ = +\infty$.

a signal structure and its implications for disagreement and polarization in the society are considered in Section 4.

3.4 Determinants of collective denial

We now turn to analyzing the conditions that favor the emergence of collective denial in the society by studying the comparative static properties of the cognitive choices in equilibrium. The analysis of the general case is difficult due to the possibility of multiple equilibria. In this subsection we therefore consider a special case for the distribution of preferences under which it is possible to find some conditions that guarantee the uniqueness of the equilibrium.

We therefore restrict attention to a situation where the distribution of political preferences is split into two types. A fraction α of citizens are endowed with a preference parameter $v_R \geq -X_L$ and always oppose regulation; their ideology is strong enough to make information irrelevant to their vote. The remaining voters have moderate preferences $-X_\varnothing \leq v_M < -X_L$ and are therefore likely to follow public recommendations. The presence of unresponsive voters imposes a lower bound α on the share of ballots in favor of the status quo. Since the behavior of voters endowed with a type v_R does not vary with their beliefs, the analysis focuses on the moderates' cognition and on the resulting political outcome, as a function of: (i) the distribution of preferences in the society, as given by v_M and α; and (ii) the intensity of the political status quo bias, parametrized by the shape of ϕ. Since Proposition 1 requires a continuous distribution of preferences, the existence and characterization of an equilibrium in that particular case is established in the Appendix.

In the absence of wishful thinking, moderate voters would vote against regulation following the message $\varnothing$ and in favor of regulation following the message L. The presence of motivated reasoning affects this outcome, since the signal equal to L might not be truthfully encoded by the moderates. If moderate voters play a denial equilibrium, their beliefs are not updated relative to the prior beliefs, and therefore their vote depends on the sign of $v_M + x_0$. In the exposition, we restrict attention to the case where $v_M + x_0 \geq 0$. In this situation, moderates who engage in ideological denial oppose risk-prevention policies, whereas they would vote $a = 0$ following the message L if they were fully informed. However, all comparative statics results in this section remain true if $v_M + x_0 < 0$.

Uniqueness of the equilibrium Let us first analyze the conditions under which a realism equilibrium exists, namely an equilibrium where all moderate voters play $\sigma^*(v_M) = 1$. In such a candidate equilibrium, $\nu_L = \alpha$ and therefore the payoff to moderate voters conditional on $m = L$ equals $(v_M + X_L)\phi(\alpha)$. A moderate voter who deviates from this equilibrium and suppresses a signal equal to L ascribes probability 1 to the message $m = \varnothing$, and thus receives an anticipatory term equal to $(v_M + X_\varnothing)\phi(1)$ in that case. Realism is therefore an equilibrium if and only if

$$s[(v_M + X_\varnothing)\phi(1) - (v_M + X_L)\phi(\alpha)] \leq c. \tag{9}$$

Consider now a candidate denial equilibrium, namely an equilibrium where all moderate voters repress bad news ($\sigma^*(v_M) = 0$). All moderate voters vote against regulation, and thus $\nu_L = \nu_\varnothing = 1$. Denial is therefore an equilibrium if and only if

$$s(1 - \frac{\lambda}{2})(X_\varnothing - X_L)\phi(1) \geq c. \tag{10}$$

Equations 9 and 10 are mutually exclusive if and only if α is larger than some threshold α^*. Thus, the model might admit multiple equilibria if the number of moderates is large relative to the number of unresponsive voters, but not otherwise. The intuition is that the moderates' cognitive strategies are strategic complements: moderates who engage in denial vote against regulation, which therefore lowers the payoffs to moderates conditional on $m = L$; in turn, this reinforces the incentives of other moderates to self-deceive as well. This effect is the strongest when α is low, which might create multiple equilibria in that case.

Comparative statics The left-hand sides of Equations 9 and 10 are the basic incentives, in the realism equilibrium and in the denial equilibrium respectively, to deny the signals equal to L and distort one's beliefs into thinking that the scientific evidence is equivocal. These equations allow analyzing the comparative statics properties of the equilibrium, since the variations of the parameters that increase these expressions foster denial. The left-hand sides of Equations 9 and 10 are

(i) nondecreasing in v_M: moderate voters are therefore more likely to repress bad news when their own distaste for regulation becomes more intense, which increases the relative desirability of the state where $m = \varnothing$, in which regulation is less likely, relative to the state where $m = L$.

(ii) nondecreasing in α: moderate voters are less willing to accept the evidence when the number of opponents to regulation is large, which makes it more likely that laissez faire will be chosen even if risk prevention is warranted and, in turn, reinforces the incentives to understate the costs of unregulated risk.

(iii) nondecreasing in $\phi(1)$ and in $\phi(\alpha)$, for the same reason. As a result, political systems that are more gridlocked, i.e., where the enactment of new policies is more difficult, are more likely to favor collective denial.

Proposition 2 summarizes and formalizes these observations by focusing on the region where the number of moderates is small, in order to guarantee the uniqueness of the equilibrium. The threshold α^* above which the equilibrium is unique, and the equilibrium cognitive strategy of the moderates in that region, are written $\alpha^*(v_M, \phi)$ and, respectively, $\sigma^*(v_M|\alpha, \phi)$, in order to emphasize the dependence on the values of the parameters. Proposition 2 establishes the uniqueness of the equilibrium and the comparative statics properties described above.

Proposition 2. *For any (v_M, ϕ) such that $-X_L > v_M > -x_0$, there exists a threshold $\alpha^*(v_M, \phi) \in [0, 1]$ such that, for any $\alpha > \alpha^*(v_M, \phi)$, there exists a unique equilibrium where moderates play the cognitive strategy $\sigma^*(v_M|\alpha, \phi)$. In addition,*

(i) collective denial is stronger in societies that have an intrinsic opposition to regulation:

- *if* $(v_M, \alpha_1, \alpha_2, \phi)$ *satisfy* $\alpha_1 \geq \alpha_2 > \alpha^*(v_M, \phi)$*, then*

$$\sigma^*(v_M|\alpha_1, \phi) \leq \sigma^*(v_M|\alpha_2, \phi).$$

- *if* $(v_M^1, v_M^2, \alpha, \phi)$ *satisfy* $-X_L > v_M^1 \geq v_M^2 > -x_0$ *and* $\alpha > \max(\alpha^*(v_M^1, \phi), \alpha^*(v_M^2, \phi))$*, then*

$$\sigma^*(v_M^1|\alpha, \phi) \leq \sigma^*(v_M^2|\alpha, \phi).$$

(ii) political status quo bias favors collective denial: if $(v_M, \alpha, \phi_1, \phi_2)$ *satisfy* $\phi_1(\nu) \geq \phi_2(\nu)$ *for all* $\nu \in [0, 1]$ *and* $\alpha > \max(\alpha^*(v_M, \phi_1), \alpha^*(v_M, \phi_2))$*, then*

$$\sigma^*(v_M|\alpha, \phi_1) \leq \sigma^*(v_M|\alpha, \phi_2).$$

Overall, the political factors that favor collective denial and (inefficient) opposition to regulation on the part of moderate voters are the following: (i) moderates' intrinsic preference for the laissez-faire policy; (ii) strong political obstacles to regulation that make political change difficult to implement, either due to a large number of ideological opponents, or to the existence of a status quo bias in the political system.

3.5 Alternative theories

This subsection reviews competing theoretical explanations for the existence of a partisan disagreement.

Reverse causal link One natural explanation for the correlation between beliefs and political attitudes is that beliefs about objective facts causally determine political orientation: liberals and conservatives do not resort to motivated reasoning but simply hold different prior beliefs or private information regarding scientific facts, which causes them to express different preferences regarding political decisions. Several facts are inconsistent with this explanation and tend to corroborate the reverse causal link, from political preferences to beliefs. First, this theory cannot account for the evidence reviewed above on non-standard updating behavior, in particular the fact that liberals and conservatives react differently to information about new issues on which they have little prior knowledge (see for instance Kahan et al., 2009, on nanotechnologies), and that the perceived policy implications of the signals affect information processing (Feinberg and Willer, 2010; Braman et al., 2012; Campbell and Kay, 2014). Second, voters' opinions on a range of scientific debates are remarkably correlated with each other: conservatives and liberals consistently disagree with each other about environmental threats (climate change, nuclear power), economic questions (efficiency of redistributive policies, sources of inequalities), and judiciary issues (deterrent effect of capital punishment, efficiency of gun control policies). This stylized fact is inconsistent with a model where individuals' preferences are determined by independent prior beliefs or private signals. In contrast, it can be accounted for by the theory developed in this paper under the assumption that people's preferences along these different dimensions (economic policies, judicial policy, regulation) exhibit some positive correlation, for instance because they reflect an underlying attitude towards individual autonomy and the role of government. Third, a theory in which beliefs determine political preferences would predict that beliefs are more precise among more highly educated groups who have access to more diverse sources of information; Kahan et al. (2012) shows instead that the disagreement between liberals and conservatives on climate science is higher among more educated individuals.[7] This fact is consistent with a motivated reasoning explanation under the assumption that a higher scientific education reduces the cost c of ideological thinking by making individuals more effective at manipulating the arguments and at finding preference-consistent sources.

Media bias Another important explanation for the gap between individuals' beliefs and experts' opinions is that the state of science is misrepresented by the media. Some existing theories can indeed account for a general misperception of scientific knowledge on the part of laypeople. Shapiro (2016) develops a model in which newspapers' incentives to build a reputation for truthfulness lead them to present mixed evidence even if the scientific diagnosis is unequivocal, which causes readers to misperceive the

[7] A Gallup survey also documents that skepticism regarding climate science is the highest among college educated Republicans (Newport and Dugan, 2015).

scientific consensus. Bramoullé and Orset (2017) analyze firms' incentives to "manufacture doubt" (Oreskes and Conway, 2010) and shape public perception in order to influence regulation. Stone (2016) develops a cheap-talk setting in which consumers are uncertain about the objectives of the media, which might have a vested interest in supporting one theory: as a consequence, it might be rational for consumers to update conservatively after receiving a signal. In these papers, the electorate is homogenous and the main focus is on the discrepancy between the beliefs held by scientists and those held by the population. The focus on disagreement and polarization in this paper therefore takes a complementary perspective to these theories. In particular, while the models of Shapiro (2016) and Bramoullé and Orset (2017) can explain the attitude of firms and media outlets regarding scientific knowledge, it does not capture demand-side effects in information processing as evidenced by the experimental literature in which the signals are controlled by an experimenter.

Another strand of the literature has also proposed an explanation based on a media bias for the persistence of disagreement in the population. According to this theory, media outlets slant their reports of scientific arguments, leading different readerships to receive different pieces of information. A variety of foundations for such a bias have been proposed, both from the demand-side perspective (Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006, 2010) and from the supply-side perspective (Baron, 2006; Larcinese et al., 2011). In addition, some empirical evidence documents that media reports are indeed biased on ideological grounds (Larcinese et al., 2011; Gentzkow et al., 2014), and that they influence consumers' attitudes (for instance DellaVigna and Kaplan, 2007). Explanations for disagreement based on frictions in the news industry predict that citizens have different beliefs because they are exposed to different sources of information. The present paper takes a complementary approach to this literature by analyzing the features of the political environment that favor belief distortion, allowing for predicting heterogeneity in beliefs between countries. The theory also predicts that liberals and conservatives react differently to the same pieces of information, which makes the model suitable for accounting for the laboratory evidence of anomalous updating of beliefs about political issues.

4 Polarization and disagreement

Besides explaining the existence of partisan disagreements, the theory of motivated reasoning can also account for the experimental evidence on anomalous updating behavior discussed in subsection 1.1. In this section we relate this theory to the existing evidence on polarization and we establish the following predictions: (*i*) beliefs of politically opposed groups polarize if they receive (common) mixed evidence; (*ii*) politically opposed groups disagree on the objective facts and form preference-consistent opinions if they receive mixed evidence; (*iii*) these phenomena are eliminated if individuals receive a sufficiently large amount of unbiased information, which prevents them from rationalizing their preferred opinion; (*iv*) unlike existing theories of polarization, beliefs are affected by individuals' preferences but not by the order in which the signals are received.

The usual interpretation of the polarization of beliefs is that people are prone to interpreting ambiguous evidence in light of their prior opinions: for instance, Lord et al. (1979) write (p. 2099), "... there is considerable evidence that people tend to interpret subsequent evidence so as to maintain their initial beliefs." The theory developed in this paper takes a different perspective and considers preferences (instead of prior beliefs) as the source of the assimilation bias. The dynamics of beliefs predicted by the model is therefore entirely driven by political attitudes, and does not feature any history-dependent bias.

We consider two individuals or groups with opposite political preferences. Group R is in favor of the laissez faire, whereas group L intrinsically prefers regulation. Both groups receive a common sequence of signals generated by the information structure introduced in Section 3, the only difference being that the signal is now symmetric and takes values in $\{L, R, \varnothing\}$: with probability λ, a message $m \in \{L, R\}$ is sent, in which case $m = \omega$ with probability π; with probability $1 - \lambda$, the message $\varnothing$ is sent.

A cognitive strategy is now a pair (σ_-, σ_+) of probabilities of transmitting the signals L and R respectively. We do not model the choice of a cognitive strategy and the voting behavior explicitly, but we assume that the individuals' cognitive choices follow the pattern obtained in Section 3: group R denies arguments in favor of regulation and accepts evidence against it, whereas group L plays the opposite strategy.

Beliefs are summarized by the random variable $\xi = \mathbb{P}[\omega = R]$. The groups' prior beliefs are written ξ_R^0 and ξ_L^0, and can be equal or different. Both groups receive a common sequence of public signals $M_n = \{m_1, \cdots, m_n\}$ and form their recollections $\hat{M}_n^i = \{\hat{m}_1^i, \cdots, \hat{m}_n^i\}$ according to their cognitive strategies. Their posterior beliefs are written $\xi_R(M_n)$ and $\xi_L(M_n)$, respectively.

The analysis consists in comparing the posterior beliefs of the two groups R and L, depending on the sequence M_n.

The first observation, provided without proof, is that beliefs follow a martingale process in spite of ideological thinking. The law of iterated expectations applies, due to the sophistication hypothesis: the expectation of an individual's beliefs following a sequence of i.i.d. signals is equal to their prior. There is therefore no systematic drift towards preference-consistent opinions. As shown below, this does not preclude ex post polarization and disagreement conditional on a sequence of signals.

Claim 1.

$$\mathbb{E}[\xi_i(M_n)] = \xi_i^0 \text{ for } i \in \{L, R\}$$

4.1 Polarization

We first focus on the conditions under which the opinions of the groups polarize. Definition 1 is adapted from Baliga et al. (2013) and Benoît and Dubra (2015): beliefs polarize if, given an initial disagreement between the two groups, their beliefs become more extreme in their original direction, after they both receive the same piece of information.

Definition 1. Suppose that $0 < \xi_L^0 < \xi_R^0 < 1$. Beliefs *polarize* following the sequence M_n if $\xi_L(M_n) < \xi_L^0 < \xi_R^0 < \xi_R(M_n)$.

The information contained in the sequence M_n is summarized by the vector $(n_R, n_L, n_\varnothing)$: n_R is the number of signals equal to R, n_L is the number of signals equal to L, and $n_\varnothing$ is the number of signals equal to $\varnothing$. A member of group R converts the signals equal to L into $\hat{m} = \varnothing$ and therefore recollects n_R signals equal to $\hat{m} = R$ and $n_L + n_\varnothing$ signals equal to $\hat{m} = \varnothing$. Bayes' rule yields

$$\frac{\xi_R(M_n)}{1-\xi_R(M_n)} = \frac{\xi_R^0}{1-\xi_R^0}\left[\frac{\pi}{1-\pi}\right]^{n_R}\left[\frac{1-\lambda\pi}{1-\lambda(1-\pi)}\right]^{n_L+n_\varnothing}. \tag{11}$$

A member of group L recollects n_L signals equal to $\hat{m} = L$ and $n_R + n_\varnothing$ signals equal to $\hat{m} = \varnothing$. By Bayes' rule,

$$\frac{\xi_L(M_n)}{1-\xi_L(M_n)} = \frac{\xi_L^0}{1-\xi_L^0}\left[\frac{1-\pi}{\pi}\right]^{n_L}\left[\frac{1-\lambda(1-\pi)}{1-\lambda\pi}\right]^{n_R+n_\varnothing}. \tag{12}$$

Comparing Equations 11 and 12 provides the conditions under which beliefs polarize. Let

$$\alpha := \ln\left[\frac{\pi}{1-\pi}\right] \text{ and } \beta := \ln\left[\frac{1-\lambda(1-\pi)}{1-\lambda\pi}\right]$$

be the posterior log-likelihood ratios conditional on receiving, respectively, a message $\hat{m} \in \{L, R\}$ or $\hat{m} = \varnothing$. The assumptions $\lambda > 0$ and $\pi > 1/2$ imply that $\alpha > \beta > 0$.

Proposition 3. *Beliefs polarize following the sequence M_n if and only if*

$$\min\left[\frac{n_R}{n_L+n_\varnothing}, \frac{n_L}{n_R+n_\varnothing}\right] > \frac{\beta}{\alpha}. \tag{13}$$

Note that condition 13 depends on M_n and on the information parameters (π, λ) but not on the prior beliefs ξ_L^0 and ξ_R^0. Polarization occurs whenever n_L and n_R are both large compared to $n_\varnothing$, that is, whenever the information contained in M_n provides arguments in favor of both opinions. An immediate corollary is that beliefs polarize if the evidence if perfectly mixed ($n_R = n_L > 0, n_\varnothing = 0$) but not if the evidence is unequivocal ($n_R = 0$ or $n_L = 0$).

Opinions polarize if mixed arguments are provided to the two groups but not if the information unambiguously recommends one position: the divergence of opinions requires that arguments advocating both policy orientations are provided, so that each group can rationalize its preferred opinion. Consistently with this observation, experiments that document a polarization of beliefs (e.g. Lord et al., 1979; Miller et al., 1993) offer mixed evidence, for instance under the form of two essays advocating different policies.

4.2 Disagreement

In addition to predicting biased assimilation, the model also predicts the direction of disagreement between the two groups, starting from identical prior beliefs. This distinguishes the theory of motivated reasoning from existing explanations for the polarization of beliefs, which take disagreement as a primitive assumption and examine whether balanced information reinforces it. In contrast, the theory of motivated reasoning can relate the disagreement about objective facts to fundamental preference parameters, and thereby predict the direction of the disagreement in situations where both groups start with identical prior beliefs (see subsection 4.4 for a discussion).

To prove this point formally, let us assume that all individuals start from uninformed prior beliefs $\xi_R = \xi_L = \xi^0 = 1/2$. Proposition 4 compares the posterior beliefs $\xi_R(M_n)$ and $\xi_L(M_n)$ of the groups to each other, and to the posterior beliefs formed by an individual who is not subject to any updating distortion and correctly encodes all signals, written $\xi(M_n)$.

Proposition 4. *If $\xi_R^0 = \xi_L^0 = \xi^0 = 1/2$, then $\xi_R(M_n) > 1/2 > \xi_L(M_n)$ if and only if M_n satisfies condition 13. In that case, both groups are overconfident:*

$$\xi_R(M_n) > \xi(M_n) > \xi_L(M_n).$$

Under condition 13, starting from uninformed prior beliefs, individuals become convinced that the available evidence justifies their preferred policy orientation, and form beliefs that are too confident (i.e. further away from 1/2) relative to those of a dispassionate statistician who correctly encodes all signals.

4.3 Asymptotic learning

Lastly, we prove that ideologically-driven disagreement and polarization vanish when individuals are provided with a large amount of unbiased information. Proposition 5 examines the asymptotic properties of the learning process from the ex-ante point of view. Consider infinite sequences $\{M_n\}$

of i.i.d. signals. Proposition 5 analyzes the limit properties of posterior beliefs for a given state ω as n becomes asymptotically large.

Proposition 5. *Suppose that $\xi_i^0 \in (0,1)$ for $i = L, R$. Then:*

1. *If $\omega = R$, $\lim_{n\to+\infty} \xi_i(M_n) = 1$ almost surely for $i = L, R$.*
2. *If $\omega = L$, $\lim_{n\to+\infty} \xi_i(M_n) = 0$ almost surely for $i = L, R$.*

Proposition 5 establishes the consistency of posterior beliefs (parts 1 and 2). Disagreement and divergence of opinions can occur on the path but vanish when the groups are provided with a sufficiently large number of signals. Large samples convey overwhelming evidence in favor of the true hypothesis and the disagreement is therefore eliminated asymptotically due to the fact that self-deception is limited by Bayesian constraints.

The main conclusion of Propositions 3–5 is that motivated reasoning might produce a temporary disagreement and polarization on the path, as documented experimentally. However, a consensus about the objective facts can still be reached across the political spectrum if strong arguments are conveyed: the influence of ideological thinking on the formation of beliefs should therefore not be overestimated. In the case of climate change, many studies have shown that individuals vastly underestimate the scientific consensus, and that setting the record straight causes a significant increase in the acceptance of climate science and in the support for public action (Lewandowsky et al., 2013; Myers et al., 2015; van der Linden et al., 2015; van der Linden, 2016). Remarkably, this communication strategy is also effective with conservative citizens.

4.4 Alternative theories

Most existing explanations for the phenomenon of polarization consider an exogenous disagreement and propose a history-dependent updating rule. In contrast, in the present paper, disagreement arises endogenously through motivated reasoning and perpetuates itself through the polarization of beliefs even if the groups have the same prior opinions and receive the same piece of information. In particular, the updating process is history-independent, as in Bayes' rule, but depends upon political preferences.

Rabin and Schrag (1999) provide the first economic model of confirmatory bias. In their theory, individuals misinterpret the signals that contradict their current opinion and encode it, with some probability, as confirming rather than contradicting evidence. Fryer Jr. et al. (2017) provide a foundation for the updating rule used in Rabin and Schrag (1999) by assuming that individuals interpret mixed evidence as a function of their beliefs, and recall only their interpretation instead of the raw signals. In both those papers, the misperception can persist indefinitely and individuals can end up believing with certainty in a false hypothesis.

Other explanations involve departures from the expected utility framework or the common likelihood ratio paradigm. Baliga et al. (2013) provide a theory of polarization based on ambiguity aversion. In their model, individuals try to hedge against uncertainty when they make their predictions. Groups with different prior beliefs are (endogenously) averse to different directions of ambiguity, which is why they update in different directions. Andreoni and Mylovanov (2012) show that, if the state of nature is multidimensional, the disagreement can be reinforced by the provision of one-dimensional signals. Benoît and Dubra (2015) assume that individuals disagree about the likelihood ratios associated with the signals: if the groups have already been exposed to some information, their idiosyncratic interpretation of the signals conditions both their current opinion and their response to new information, which creates a spurious correlation between beliefs and updating patterns. Acemoglu et al. (2016) consider a related phenomenon, and prove that individuals might disagree forever absent the common likelihood ratio assumption.

The preference-based theory of polarization developed in the present paper differs from existing history-based explanations in several important ways, both in its applications and in its predictions:

(i) First, the theory requires that individuals have personal stakes in the outcome. It is therefore unable to explain polarization in situations where participants have to predict an event that does not affect their well-being, as in the experiment by Darley and Gross (1983) in which subjects are asked to assess a schoolgirl's academic skills.

(ii) Second, as noted above, the updating distortion is affected by preferences but not by prior beliefs: first impressions do not matter, but preferences do. The model thus predicts the first movement of beliefs if people have uninformed prior opinions and receive balanced evidence as the first piece of information. For instance, it explains why pro- and anti-regulation individuals disagree about the risks associated with nanotechnologies after receiving information but not prior to it, as documented by Kahan et al. (2009).

(iii) Third, the framework is entirely Bayesian and individuals' beliefs satisfy the law of iterated expectations, which is not the case in most of the papers mentioned above. The results show that this restriction on belief formation does not preclude polarization contingent on specific types of signals.

(iv) Fourth, the theory predicts a confirmatory bias if groups have already formed preference-consistent opinions, but not otherwise. To see why, imagine that both groups have identical prior beliefs $\xi_L^0 = \xi_R^0 = \frac{1}{2}$ and receive a preference-inconsistent signal: all members of group L receive a signal equal to R, whereas all members of group R receive a signal equal to L. Similarly to existing theories, the present model predicts that posterior beliefs move in the direction of the signal: $\xi_L(R) > \frac{1}{2} > \xi_R(L)$. Suppose now that all individuals receive a public message containing mixed evidence, for instance a sequence (R, L). Prior-based theories of polarization predict that each group interprets the new evidence as a confirmation of their prior opinion, which reinforces the disagreement: $\xi_L(R, R, L) > \xi_L(R) > \xi_R(L) > \xi_R(L, R, L)$. The model of Section 4 instead predicts that groups update according to their preference irrespective of their current opinion. This reduces (or even reverses) the discrepancy of views: $\xi_L(R, R, L) < \xi_L(R)$ and $\xi_R(L, R, L) > \xi_R(L)$.

5 Conclusion

This paper has proposed that motivated reasoning can explain the existence and persistence of partisan disagreement about scientific issues. This theory takes wishful thinking and heterogeneous political attitudes as a primitive assumption, and shows that these ingredients can explain a wide range of empirical observations, including some experimental evidence of non-standard updating behavior. The main predictions of this model are that individuals might interpret scientific information in a heterogeneous manner depending on their stakes in the resulting political decisions, that collective denial is strongest in societies where appropriate political responses are least likely, and that this assimilation bias might predict a (temporary) polarization between politically opposed groups following the provision of balanced information.

The analysis can be extended in several directions. First, giving more precise foundations for the model of self-deception, and understanding the constraints that limit people's cognitive choices might be helpful for designing efficient communication strategies to limit the effects of motivated reasoning. Second, the theory assumes that all individuals receive a common public signal, and therefore remains silent about individuals' attitudes towards information sources. Understanding the consequences of motivated reasoning on individuals' preferences regarding the type of information received might be helpful for predicting their choices in important contexts, such as the choice of a media outlet, or even the formation of communication networks contingent on political attitudes. More generally, the interaction of consumers prone to cognitive distortions with sources of information motivated by idiosyncratic interests (the media, political parties, firms, etc.) raises interesting and important questions, that are left for future research.

Appendix: Proofs

A.1 Proofs of Section 3

A.1.1 Proof of Lemma 1

Lemma A.1. *For any intra-personal equilibrium strategy $\sigma^*(v)$,*

$$\mathbb{E}[X \mid \hat{m} = L, \sigma^*(v)] \leq \mathbb{E}[X \mid \hat{m} = \varnothing, \sigma^*(v)].$$

Proof. By Bayes' rule,

$$\mathbb{E}[X \mid \hat{m} = L, \sigma^*(v)] = X_L$$

whereas

$$\mathbb{E}[X \mid \hat{m} = \varnothing, \sigma^*(v)] = \mu(\sigma^*(v))X_L + [1 - \mu(\sigma^*(v))]X_\varnothing.$$

The result follows from $X_L \leq X_\varnothing$. □

Lemma A.2. *In any equilibrium, $\nu_\varnothing \geq \nu_L$.*

Proof. Consider a profile of equilibrium cognitive strategies $\Sigma^* = \{\sigma^*(v)\}$ and the associated votes:

$$a(v, \hat{m}) = \begin{cases} 1 & \text{if } v + \mathbb{E}[X \mid \hat{m}, \sigma^*(v)] \geq 0 \\ 0 & \text{otherwise .} \end{cases}$$

Lemma A.1 implies that $a(v, \varnothing) \geq a(v, L)$ for all v. The equilibrium political outcomes are given by

$$\nu_\varnothing = \int_{-\infty}^{+\infty} a(v, \varnothing) dF(v)$$

and

$$\nu_L = \int_{-\infty}^{+\infty} [\sigma^*(v)a(v, L) + (1 - \sigma^*(v))a(v, \varnothing)]dF(v)$$

which yields $\nu(\varnothing) \geq \nu_L$. □

To complete the proof of Lemma 1, note that $\sigma^*(v)$ is uniquely defined by

$$\sigma^*(v) = \begin{cases} 1 \text{ if } s\mathcal{I}(v) \leq c \\ \\ 0 \text{ if } s\mathcal{I}(v) \geq \dfrac{2c}{2-\lambda} \\ \\ \dfrac{2}{\lambda} - \dfrac{2-\lambda}{\lambda}\dfrac{s}{c}\mathcal{I}(v) \text{ otherwise.} \end{cases}$$

The monotonicity of $\sigma^*(v)$ in v follows from Equation 4.

A.1.2 Proof of Proposition 1

The function $g : \nu \to (g_{\varnothing}(\nu), g_L(\nu))$ maps the convex and compact set $\{(\alpha, \beta) \in [0,1]^2 \mid \alpha \geq \beta\}$ into itself, as shown by Lemma A.2. The next step is to prove that g is continuous in order to apply Brouwer's theorem and find a fixed point of g.

To prove the continuity of g, notice that the function $(v, \nu) \to \sigma^*(v, \nu)$ is continuous by the construction of lemma 1. As a consequence, the function $(v, \nu) \to v + \mathbb{E}[X \mid \hat{m}, \sigma^*(v, \nu)]$ is also continuous for $\hat{m} \in \{\varnothing, L\}$. Let us rewrite $g_{\varnothing}$ as

$$g_{\varnothing}(\nu) = \int_{v : v + \mathbb{E}[X \mid \hat{m} = \varnothing, \sigma^*(v,\nu)] \geq 0} f(v) dv$$

which shows that $g_{\varnothing}$ is continuous in ν (remember that f is continuous). A similar argument proves the result for g_L.

A.1.3 Proof of Proposition 2

The main text provides the conditions under which realism (Equation 9) and denial (Equation 10) are equilibrium strategies. Consider now a candidate mixed-strategy equilibrium where the cognitive strategy of the moderates equals σ^*. In this equilibrium, a fraction σ^* of the moderates transmits the signal equal to L and votes for regulation, whereas a fraction $1 - \sigma^*$ conceals the signal and votes for the status quo. Thus, $\nu_{\varnothing} = 1$ whereas $\nu_L = \alpha + (1 - \sigma^*)(1 - \alpha)$. Hence, σ^* is indeed an equilibrium strategy if and only if

$$s\frac{2-\lambda}{2-\lambda\sigma^*}[(v_M + X_{\varnothing})\phi(1) - (v_M + X_L)\phi(1 - \sigma^*(1-\alpha))] = c.$$

Consider the function

$$h(\sigma) = s\frac{2-\lambda}{2-\lambda\sigma}[(v_M + X_\varnothing)\phi(1) - (v_M + X_L)\phi(1-\sigma(1-\alpha))]$$

defined on $[0,1]$. This function is continuously differentiable in σ and its derivative is of the sign of (after some algebra)

$$\lambda[(v_M+X_\varnothing)\phi(1)-(v_M+X_L)\phi(1-\sigma(1-\alpha))]+(1-\alpha)(2-\lambda\sigma)(v_M+X_L)\phi'(1-\sigma(1-\alpha)).$$

If $\alpha = 1$, this expression is strictly positive for any $\sigma \in [0,1]$. Thus, by continuity, there exists a threshold $\alpha^*(v_M, \phi)$ such that, for any $\alpha > \alpha^*(v_M, \phi)$, the function h is strictly increasing in σ. As a result, if $\alpha > \alpha^*(v_M, \phi)$, there exists a unique equilibrium of the game in which the moderates' cognitive strategy is characterized by

$$\sigma^*(v_M|\alpha,\phi) \begin{cases} = 1 \text{ if } h(1) \leq c, \\ = 0 \text{ if } h(0) \geq c, \\ \text{ is the unique solution to } h(\sigma) = c \text{ otherwise.} \end{cases}$$

To prove the comparative statics properties, note that the benefit from self-deception in equilibrium, given by $h(\sigma^*)$, is

- nondecreasing in α, since ϕ is strictly increasing and $v_M + X_L < 0$;
- nondecreasing in v_M, since $\phi(1) \geq \phi(1-\sigma^*(1-\alpha))$;
- nondecreasing in $\phi(1)$ and in $\phi(1-\sigma^*(1-\alpha))$.

A.2 Proofs of Section 4

A.2.1 Proof of Proposition 3

Let us rewrite Equations 11 and 12 as

$$\ln\left[\frac{\xi_R(M_n)}{1-\xi_R(M_n)}\right] = \ln\left[\frac{\xi_R^0}{1-\xi_R^0}\right] + \alpha n_R - \beta(n_L + n_\varnothing) \qquad \text{(A.1)}$$

and

$$\ln\left[\frac{\xi_L(M_n)}{1-\xi_L(M_n)}\right] = \ln\left[\frac{\xi_L^0}{1-\xi_L^0}\right] - \alpha n_L + \beta(n_R + n_\varnothing). \tag{A.2}$$

The condition $\xi_R(M_n) > \xi_R^0$ is therefore equivalent to

$$\alpha n_R - \beta(n_L + n_\varnothing) > 0$$

whereas $\xi_L(M_n) < \xi_L^0$ is equivalent to

$$-\alpha n_L + \beta(n_R + n_\varnothing) < 0.$$

This proves the result.

A.2.2 Proof of Proposition 4

Given $\xi_R^0 = \xi_L^0 = 1/2$, A.1 shows that $\xi_R(M_n) > 1/2$ is equivalent to $n_R\alpha > (n_L + n_\varnothing)\beta$, whereas A.2 shows that $\xi_L(M_n) < 1/2$ is equivalent to $n_L\alpha > (n_R + n_\varnothing)\beta$. Therefore $\xi_R(M_n) > 1/2 > \xi_L(M_n)$ is equivalent to 13.

In addition, Bayes' rule yields

$$\ln\left[\frac{\xi(M_n)}{1-\xi(M_n)}\right] = \ln\left[\frac{\xi^0}{1-\xi^0}\right] + (n_R - n_L)\alpha.$$

Thus, using A.1 and A.2, condition 13 is sufficient for $\xi_R(M_n) > \xi(M_n) > \xi_L(M_n)$.

A.2.3 Proof of Proposition 5

To prove part 1 (part 2 is symmetric), suppose that $\omega = R$ and consider an agent belonging to the group R. Consider an infinite sequence of messages and, for any $n > 1$, the following random variables: $a_R(n)$ is equal to the number of signals equal to R in the sequence up to date n, $a_L(n)$ is equal to the number of signals equal to L, and $a_\varnothing(n)$ is equal to the number of signals equal to $\varnothing$. By the law of large numbers, with probability 1

$$\lim_{n\to+\infty} \frac{a_R(n)}{n} = \lambda\pi \text{ and } \lim_{n\to+\infty} \frac{a_\varnothing(n) + a_L(n)}{n} = 1 - \lambda\pi. \tag{A.3}$$

Consider an infinite sequence that satisfies property A.3. Let us rewrite Equation 11 as

$$\frac{1}{n}\ln\left[\frac{\xi_R(M_n)}{1-\xi_R(M_n)}\right] = \frac{1}{n}\ln\left[\frac{\xi_R}{1-\xi_R}\right] + \frac{a_R(n)}{n}\alpha - \frac{a_\varnothing(n)+a_L(n)}{n}\beta$$

which implies

$$\lim_{n\to+\infty}\frac{1}{n}\ln\left[\frac{\xi_R(M_n)}{1-\xi_R(M_n)}\right] = \lambda\pi\alpha - (1-\lambda\pi)\beta. \tag{A.4}$$

The assumptions $\pi > 1/2$ and $\lambda > 0$ imply that $\lambda\pi\alpha - (1-\lambda\pi)\beta > 0$. Thus, by Equation A.4,

$$\lim_{n\to+\infty}\frac{1}{n}\ln\frac{\xi_R(M_n)}{1-\xi_R(M_n)} > 0$$

which implies

$$\lim_{n\to+\infty}\xi_R(M_n) = 1.$$

This is true for any infinite sequence that satisfies A.3, namely almost surely.

We skip the proof of $\lim_{n\to+\infty}\xi_L(M_n) = 1$, which relies on similar arguments.

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2016). Fragility of asymptotic agreement under Bayesian learning. Theoretical Economics 11(1), 187–225.

Akerlof, K., E. Maibach, D. Fitzgerald, A. Cedeno, and A. Neuman (2013). Do people ”personally experience” global warming, and if so how, and does it matter? Global Environmental Change 23(1), 81–91.

Anderegg, W., J. Prall, J. Harold, and S. Schneider (2010). Expert credibility in climate change. Proceedings of the National Academy of Sciences 107(27), 12107–12109.

Andreoni, J. and T. Mylovanov (2012). Diverging opinions. American Economic Journal: Microeconomics 4(1), 209–232.

Baliga, S., E. Hanany, and P. Klibanoff (2013). Polarization and ambiguity. American Economic Review 103(7), 3071–3083.

Baron, D. P. (2006). Persistent media bias. Journal of Public Economics 90(1), 1–36.

Bénabou, R. (2008). Ideology. Journal of the European Economic Association 6(2-3), 321–352.

Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets. Review of Economic Studies 80, 429–462.

Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. Quarterly Journal of Economics 117(3), 871–915.

Bénabou, R. and J. Tirole (2006). Belief in a just world and redistributive politics. Quarterly Journal of Economics 121(2), 699–746.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. Quarterly Journal of Economics 126(2), 805–855.

Bénabou, R. and J. Tirole (2016). Mindful economics: The production, consumption, and value of beliefs. Journal of Economic Perspectives 30(3), 141–164.

Benoît, J.-P. and J. Dubra (2015). A theory of rational attitude polarization. Working paper.

Braman, D., D. Kahan, H. Jenkins-Smith, T. Tarantola, and C. Silva (2012). Geoengineering and the science communication environment: A cross-cultural experiment. GW Law Faculty Publications and Other Works. Paper 199.

Bramoullé, Y. and C. Orset (2017). Manufacturing doubt. Working paper.

Breyer, S. (1995). Breaking the vicious circle: Toward effective risk regulation. Harvard University Press.

Bridet, L. and P. Schwardmann (2017). Selling dreams: Endogenous optimism in lending markets. Working paper.

Brunnermeier, M. and J. Parker (2005). Optimal expectations. American Economic Review 95(4), 1092–1118.

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

Caplin, A. and J. Leahy (2001). Psychological expected utility theory and anticipatory feelings. Quarterly Journal of Economics 116(1), 55–79.

Coate, S., M. Conlin, and A. Moro (2008). The performance of pivotal-voter models in small-scale elections: Evidence from texas liquor referenda. Journal of Public Economics 92(3), 582–596.

Darley, J. M. and P. H. Gross (1983). A hypothesis-confirming bias in labeling effects. Journal of Personality and Social Psychology 44(1), 20.

DellaVigna, S. and E. Kaplan (2007). The Fox News effect: Media bias and voting. Quarterly Journal of Economics 122(3), 1187–1234.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2016). Voting to tell others. Review of Economic Studies 84(1), 143–181.

Douglas, M. (1994). Risk and blame: Essays in cultural theory. Routledge.

Douglas, M. and A. Wildavsky (1982). Risk and culture: An essay on the selection of technological and environmental dangers. University of California Press.

Dunlap, R. and A. McCright (2008). A widening gap: Republican and Democratic views on climate change. Environment 50(5), 26–35.

Dunlap, R. and A. McCright (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001-2010. The Sociological Quarterly 52, 155–194.

Farnsworth, S. and R. Lichter (2012). The structure of scientific opinion on climate change. International Journal of Public Opinion Research 24(1), 93–103.

Feinberg, M. and R. Willer (2010). Apocalypse soon? Dire messages reduce belief in global warming by contradicting just-world beliefs. Psychological science 22(1), 34–38.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford University Press.

Fryer Jr., R. G., P. Harms, and M. O. Jackson (2017). Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. Working paper.

Gentzkow, M. and J. M. Shapiro (2006). Media bias and reputation. Journal of Political Economy 114(2), 280–316.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? Evidence from US daily newspapers. Econometrica 78(1), 35–71.

Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2014). Competition and ideological diversity: Historical evidence from US newspapers. American Economic Review 104(10), 3073–3114.

Goebbert, K., H. Jenkins-Smith, K. Klockow, M. Nowlin, and C. Silva (2012). Weather, climate, and worldviews: The sources and consequences of public perceptions of changes in local weather patterns. Weather, Climate, and Society 4(2), 132–144.

Gottlieb, D. (2018). Will you never learn? Self-deception and biases in information processing. Working paper.

Hardisty, D., E. Johnson, and E. Weber (2010). A dirty word or a dirty world? Attribute framing, political affiliation and query theory. Psychological Science 21(1), 86–92.

Jenkins-Smith, H. C., C. L. Silva, M. C. Nowlin, and G. DeLozier (2011). Reversing nuclear opposition: Evolving public acceptance of a permanent nuclear waste disposal facility. Risk Analysis 31(4), 629–644.

Kahan, D. M. (2012). Cultural cognition as a conception of the cultural theory of risk. In S. Roeser, R. Hillerbrand, P. Sandin, and M. Peterson (Eds.), Handbook of Risk Theory: Epistemology, Decision Theory, Ethics and Social Implications of Risk. Springer.

Kahan, D. M., D. Braman, P. Slovic, J. Gastil, and G. Cohen (2009). Cultural cognition of the risks and benefits of nanotechnology. Nature Nanotechnology 4(2), 87–90.

Kahan, D. M., E. Peters, M. Wittlin, P. Slovic, L. L. Ouellette, D. Braman, and G. Mandel (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. Nature climate change 2(10), 732–735.

Köszegi, B. (2010). Utility from anticipation and personal equilibrium. Economic Theory 44, 415–444.

Kunda, Z. (1990). The case for motivated reasoning. Psychological bulletin 108(3), 480.

Larcinese, V., R. Puglisi, and J. M. Snyder (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers. Journal of Public Economics 95(9), 1178–1189.

Levy, R. (2014). Soothing politics. Journal of Public Economics 120, 126–133.

Lewandowsky, S., G. E. Gignac, and S. Vaughan (2013). The pivotal role of perceived scientific consensus in acceptance of science. Nature Climate Change 3(4), 399–404.

Lord, C. G., L. Ross, and M. R. Lepper (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology 37(11), 2098–2109.

Mayraz, G. (2011). Priors and desires - a model of optimism, pessimism, and cognitive dissonance. Working paper.

Mijovic-Prelec, D. and D. Prelec (2010). Self-deception as self-signalling: A model and experimental evidence. Philosophical Transaction of the Royal Society 365, 227–240.

Miller, A. G., J. W. McHoskey, C. M. Bane, and T. G. Dowd (1993). The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. Journal of Personality and Social Psychology 64(4), 561.

Mullainathan, S. and A. Shleifer (2005). The market for news. American Economic Review 95(4), 1031–1053.

Munro, G. D. and P. H. Ditto (1997). Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. Personality and Social Psychology Bulletin 23(6), 636–653.

Myers, T. A., E. Maibach, E. Peters, and A. Leiserowitz (2015). Simple messages help set the record straight about scientific agreement on human-caused climate change: The results of two experiments. PloS one 10(3), e0120985.

Newport, F. and A. Dugan (2015). College-educated Republicans most skeptical of global warming. http://www.gallup.com/poll/182159/college-educated-republicans-skeptical-global-warming.aspx?g_source=CATEGORY_CLIMATE_CHANGE&g_medium=topic&g_campaign=tiles. Accessed: 2016-05-10.

Nisbet, M. (2005). The competition for worldviews: Values, information, and public support for stem cell research. International Journal of Public Opinion Research 17(1), 90–112.

Nyhan, B. and J. Reifler (2010). When corrections fail: The persistence of political misperceptions. Political Behavior 32(2), 303–330.

Oreskes, N. and E. Conway (2010). Merchants of Doubt. Bloomsbury Press.

Ortoleva, P. and E. Snowberg (2015). Overconfidence in political behavior. American Economic Review 105(2), 504–35.

Pew Research Center (2010). Wide partisan divide over global warming. http://www.pewresearch.org/2010/10/27/wide-partisan-divide-over-global-warming/. Accessed: 2016-05-10.

Plous, S. (1991). Biases in the assimilation of technological breakdowns: Do accidents make us safer ? Journal of Applied Social Psychology 21(13), 1058–1082.

Rabin, M. and J. L. Schrag (1999). First impressions matter: A model of confirmatory bias. Quarterly Journal of Economics 114(1), 37–82.

Saad, L. (2013). Republican skepticism toward global warming eases. http://www.gallup.com/poll/161714/republican-skepticism-global-warming-eases.aspx. Accessed: 2016-05-10.

Saad, L. (2014). A steady 57 percent in US blame humans for global warming. http://www.gallup.com/poll/167972/steady-blame-humans-global-warming.aspx. Accessed: 2016-05-10.

Shapiro, J. M. (2016). Special interests and the media: Theory and an application to climate change. Journal of Public Economics 144, 91–108.

Slovic, P. (2000). The Perception of Risk. Routledge.

Stone, D. F. (2016). A few bad apples: Communication in the presence of strategic ideologues. Southern Economic Journal 83(2), 487–500.

Sunstein, C. (2001). Republic.com. Princeton University Press.

van der Linden, S. L. (2016). A conceptual critique of the cultural cognition thesis. Science Communication 38(1), 128–138.

van der Linden, S. L., A. A. Leiserowitz, G. D. Feinberg, and E. W. Maibach (2015). The scientific consensus on climate change as a gateway belief: Experimental evidence. PloS one 10(2), e0118489.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. Journal of Personality and Social Psychology 39(5), 806.