Journal of Economic Perspectives—Volume 30, Number 3—Summer 2016—Pages 141–164

Mindful Economics: The Production, Consumption, and Value of Beliefs

Roland Bénabou and Jean Tirole

In the economic models of old, agents had backward-looking expectations, arising from simple extrapolation or error-correction rules. Then came the rational-expectations revolution in macroeconomics, and in microeconomics the spread and increasing refinements of modern game theory. Agents were now highly sophisticated information processors, who could not be systematically fooled. This approach reigned for several decades until the pendulum swung back with the rise of behavioral economics and its emphasis on "heuristics and biases" (as in Tversky and Kahneman 1974). Overconfidence, confirmation bias, distorted probability weights, and a host of other "wired-in" cognitive mistakes are now common assumptions in many areas of economics. Over the last decade or so, the pendulum has started to swing again toward some form of adaptiveness, or at least implicit purposefulness, in human cognition.

In this paper, we provide a perspective into the main ideas and findings emerging from the growing literature on *motivated beliefs and reasoning*. This perspective emphasizes that beliefs often fulfill important psychological and functional needs of the individual. Economically relevant examples include confidence in ones' abilities, moral self-esteem, hope and anxiety reduction, social identity, political ideology and religious faith. People thus hold certain beliefs in part because

■ *Roland Bénabou is Theodore A. Wells '29 Professor of Economics and Public Affairs, Princeton University, Princeton, New Jersey and a Senior Fellow, Canadian Institute for Advanced Research, Toronto, Canada. Jean Tirole is Chairman, Toulouse School of Economics, and President of the Executive Committee, Institute for Advanced Study in Toulouse, both in Toulouse, France.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.3.141

doi=10.1257/jep.30.3.141

they attach value to them, as a result of some (usually implicit) tradeoff between *accuracy* and *desirability*. Such beliefs will therefore be resistant to many forms of evidence, with individuals displaying non-Bayesian behaviors such as not wanting to know, wishful thinking, and reality denial. At the same time, motivated beliefs will respond to the costs, benefits, and stakes involved in maintaining different *self-views* and *world-views*. These tradeoffs can be influenced by experimenters, allowing for empirical tests, and by a person's social and economic environment, leading to the possibility of self-sustaining "social cognitions."[1]

At an individual level, overconfidence is perhaps the most common manifestation of the motivated-beliefs phenomenon. There is considerable evidence of overoptimistic tendencies on the part of consumers, investors, and top corporate executives (as discussed in a "Symposium on Overconfidence" in the Fall 2015 issue of this journal). While excessive overconfidence is quite dangerous, moderate amounts can be valuable: hope and confidence feel better than anguish and uncertainty, and they often also enhance an individual's ability to act successfully on their own behalf and interact productively with others. Using data from the Survey of Consumer Finances, Puri and Robinson (2007) thus find that more optimistic individuals work more, save more, expect to retire later, and are more likely to remarry after divorce. Alloy and Abrahamson (1979) and Korn et al. (2014) find that most psychologically "healthy" people display some degree of overoptimism and biased updating, while it is primarily depressed subjects who seem to be more objective. People thus find themselves motivated (often unconsciously) to achieve "positive" beliefs, and this typically occurs through a fundamental asymmetry in the *process* by which beliefs are revised in the face of new evidence: individuals update suitably when facing good news, but fail to properly account for bad news (Eil and Rao 2011; Möbius, Niederle, Niehaus, and Rosenblat 2011; Sharot and Garrett et al. 2016).

Although goal-directed, self-deception can nonetheless end up hurting the individual: since it is an informational game that people play with themselves, the outcome may be highly inefficient—a form of self-trap. When motivated thinking becomes a *social* phenomenon, consequences can be even more severe. Collectively shared belief distortions may amplify each other (an issue we shall address), so that entire firms, institutions, and polities end up locked in denial of unpleasant realities and blind to major risks: unsustainable fiscal imbalances or labor market policies, climate change, collapse of housing or financial markets, and so on. Case and Shiller (2003) surveyed the expectations of homeowners during the real-estate bubbles of 1988 and 2003. In both cases, 90 percent of respondents thought housing prices in their city would "increase over the next several years," with an average expected gain for their own property of 9 to 15 percent *per year* over the next ten years. In the political realm, examples of persistent ideological blind spots impeding reforms and of evidence-proof conspiracy theories are abundant.

[1] Parts of this paper draw substantially on Bénabou (2015), which also provides a more explicit treatment of the underlying formal framework.

We now turn to the sources, means, costs, and benefits of motivated cognition. In a sense, we propose to treat beliefs as regular economic goods and assets—which people *consume, invest in,* reap returns from, and *produce*, using the informational inputs they receive or have access to. We first highlight the theory's general principles, then turn to a number of empirical tests and specific applications.

Motivated Beliefs: Why and How?

Why?

For a standard economic agent, information is always valuable, whether the news is good or bad: more data helps make better choices, and if not, it can just be ignored. The value of information exactly equals the extent to which it improves decision-making, and it cannot be negative. Schelling (1988), in contrast, aptly described "the mind as a consuming organ," and indeed we are all familiar with beliefs that have a direct and powerful affective impact. These may be perceptions about ourselves, like self-esteem and self-disappointment (Smith 1759; Bénabou and Tirole 2002; Köszegi 2006), or about the broader environment we face and our prospects in it that evoke strong feelings of fear, anxiety, hope, excitement, and so on (Akerlof and Dickens 1982; Loewenstein 1987; Caplin and Leahy 2001; Brunnermeier and Parker 2005; Eliaz and Spiegler 2006; Bénabou and Tirole 2011). Such "consumable" beliefs can be represented as an argument directly entering the preferences of agents.

Subjective beliefs also often have an important instrumental value, enhancing "self-efficacy." First, confidence in one's ability and chances of success is a powerful motivator to undertake and persevere in long-term projects. This source of demand for "positive" thoughts is generally derived as arising from a self-control problem over effort or tempting consumptions (Carrillo and Mariotti 2000; Brocas and Carrillo 2001; Bénabou and Tirole 2002, 2004). Belief distortions can similarly serve as commitment devices in other settings involving a divergence between preferences that occur before or after a decision, as with an agent who fears "getting cold feet" when a risky decision becomes imminent (Epstein 2008; Eisenbach and Schmalz 2015) or succumbs to "excessive" empathy and generosity when confronted with human misery (Dillenberger and Sadowski 2012). Second, being convinced of one's strength, determination, talent, honesty, and even sincerity helps convince others. Trivers (2011) and Von Hippel and Trivers (2011) hypothesize that this signaling value is why humans may have evolved the capacity to self-deceive, which later on was coopted for other uses.

The framework sketched in the next section will incorporate both classes of motives underlying departures from objective cognition: *affective* (making oneself or one's future look better) and *functional* (helpful to achieve certain goals, internal or external). Religion, the number one form of valued beliefs, typically serves both purposes, simultaneously providing comfort/reassurance and self-discipline.

How?

A consumption or efficacy motive for holding certain self-views and world-views does not ensure that such views will arise and persist, given the constraints and feedback of reality. Because the activities of paying attention (or not), processing, encoding, and rehearsing data predate the stage where we retrieve and ultimately use these signals, however, they open the door to strategic manipulations of our own information, whether conscious or automatic, progressive or abrupt. The strategies of self-deception and dissonance-reduction used to protect valued beliefs are many and varied, but we can group them into three main types: *strategic ignorance, reality denial,* and *self-signaling*.

Strategic ignorance consists in avoiding information sources that *may* hold bad news, for fear that such news could demotivate us, induce distressing mental states, or both. For instance, many at-risk subjects refuse to be tested for Huntington's disease or HIV (Oster, Shoulson, and Dorsey 2013; Ganguly and Tasoff forthcoming) even though the test is free, accurate, and can be done anonymously.

Reality denial is the failure to update beliefs properly in response to bad news. When credible warning signs are received but the feared state of the world is not yet materially incontrovertible, these signals can be processed and encoded in a distorted or dampened manner. Thus, accumulating red flags may indicate an ever-rising probability of disease, or of a housing-market crash, yet agents find ways of not internalizing the data and rationalizing away the risks, as revealed by their unchanged life plans, failure to divest or diversify from risky investments, and so on.

Self-signaling refers to a set of strategies by which the agent manufactures "diagnostic" signals of the desired type, by making choices that he later interprets as impartial evidence concerning his own underlying preferences, abilities, or knowledge about the state of the world (Quattrone and Tversky 1984; Bodner and Prelec 2003; Bénabou and Tirole 2004, 2011). In the health domain, for instance, this corresponds to people who "push" themselves to overcome their symptoms, carrying out difficult or even dangerous activities not only for their own sake, but also as "proof" that everything is fine.

Three Telltale Markers

Three key features differentiate motivated thinking and cognitive tendencies from "mechanical failures" of inference due to bounded rationality or limited attention.

1. Endogenous directionality. In contrast to what are often referred to as "System I" biases, motivated beliefs are by definition directed toward some end, though generally not consciously so. As an example, consider the opposite predictions of *confirmation bias* and *self-enhancement* for how someone who is initially insecure about his skill, attractiveness, or health will respond to feedback about these qualities in himself. A "wired in" confirmation bias would lead him to read any ambiguous signals received as confirming and hardening his negative self-view. This type of response is quite rare and found primarily in clinically depressed individuals. The great majority of people, in contrast, find ways to interpret the same evidence

positively, and even clearly bad news as "not that bad," irrelevant, or biased—in line with a self-esteem maintenance motive (Alloy and Abrahamson 1979; Korn et al. 2014). Where confirmation bias typically arises is with respect to external facts, and in such cases it can often be understood as a form of motivated cognition: "This event or data is consistent with what I thought, and it shows I was right." Indeed, an important talent is the ability to analyze situations correctly from the outset. This generates a strong "demand for consistency" in judgments and choices (for evidence, see Falk and Zimmermann 2011), as a positive self-assessment on that dimension increases confidence that someone's personal investments will pay off, thus generating anticipatory utility and motivating people to undertake these projects in the first place

Another example of endogenous directionality is that, in contrast to the case of "built-in" overconfidence, agents will either overestimate or underestimate their own abilities depending on which distortion is advantageous in the situation they expect to face. In particular, when effort and talent are substitutes, rather than complements, building motivation to train requires attributing one's past successes to luck more than talent. Thus, a successful student or athlete may try to think of previous exams and competitions as having been easy compared to the next one that will require additional effort—a form of "defensive pessimism."[2]

2. *Neither naiveté nor lack of attention.* The concept of bounded rationality almost necessarily implies that more analytically sophisticated and better-educated individuals should be less prone to mistakes and biases. Such is indeed the case for the endowment effect, loss aversion, hyperbolic discounting, and even visual illusions (Frederick 2005). However, when it comes to rationalizing away contradictory evidence, compartmentalizing knowledge, and deluding oneself, more educated, attentive, and analytically able people often display greater propensities toward such behaviors. Thus in Kahan (2013) and Kahan, Peters, Dawson, and Slovic (2014), subjects who scored highest on the Cognitive Reflection Test (which measures deliberate and reflective versus intuitive and heuristic thinking) and highest on numeracy tests were less likely to display self-serving failures to update and rationalizations when facing ideologically neutral questions, but *more* likely to do so for ideologically charged issues such as man-made climate change or gun control. In large representative US surveys, Oliver and Wood (2014) similarly find that while education is negatively associated with belief in political conspiracies, political knowledge and interest are not. Ortoleva and Snowberg (2015a) find that overconfidence (about current and future inflation and unemployment) is uncorrelated with education or income and *increases* systematically with media exposure, age, and partisanship.

3. *Heat versus light.* Finally, in "motivated" there is also *emotion.* Challenging cherished beliefs directly—like a person's religion, identity, morality, or politics—evokes strong emotional and even physical responses of anger, outrage, and disgust.

[2] See Bénabou and Tirole (2002) in a context of self-motivation and Charness, Rustichini, and van de Ven (2013) in a context of strategic interactions, where experimental subjects who know they will face a competitive task become overconfident only when such beliefs confer a strategic advantage.

Such pushback is a clear "signature" of protected beliefs: not only would a Bayesian always welcome more data, but so would any naïve boundedly rational thinker.[3] Our emphasis on the interplay of emotions and information-processing is consistent with a similar trend under way in psychology and neuroscience, sometimes referred to as the "affective revolution" or "second cognitive revolution."

A Portable Paradigm

Our conceptual framework for analyzing motivated cognition, both individual and social, draws most closely on Bénabou and Tirole (2002) and Bénabou (2013), but more generally synthesizes a number of ideas common to the large literature on beliefs as direct and/or instrumental sources of utility (Akerlof and Dickens 1982; Loewenstein 1987; Carrillo and Mariotti 2000; Caplin and Leahy 2001; Brocas and Carrillo 2001; Brunnermeier and Parker 2005; Köszegi 2006, 2010). For a formal exposition of this approach, see Bénabou (2015).

A risk-neutral agent has a time horizon of three periods. Period 0 is when information may be sought or avoided, received, and ultimately processed into the beliefs carried into period 1. In period 1, the agent's actions and wellbeing will reflect these posteriors. In period 2, all uncertainty is resolved and final payoffs are received. These depend on the realized state of the world, the action taken at date 1, and possibly an initial endowment such as wealth, human or social capital, genes, or other factors. For simplicity, there are just two possible date-0 signals about the state of the world, L (low) and H (high), corresponding respectively to bad and good news (or, alternatively, bad news and no news, which then constitutes good news) about the return to effort.

A first reason why an individual (Self 0) may want to distort his or her own (Self 1's) beliefs away from what objective information indicates is enhancing *self-efficacy*. If the date-1 decision is subject to a temptation or self-control problem, Self 0 may want to bias Self 1's beliefs about the return to effort, like a parent telling their child that crime *never* pays and homework *always* does. The cost of "maintained optimism" as an internal commitment device is that decisions at date-1 will sometimes be costly mistakes even from an ex-ante point of view, such as attempting a task that is infeasible or even dangerous for the agent. When self-control is enough of a concern, however, some degree of "positive thinking" can be advantageous.

A second class of motives for self-deception is *affective*. The basic framework remains unchanged, except that instead of facing a self-control problem at time 1, the agent derives a direct flow of utility (or disutility) from the beliefs he holds during this period. These hedonic beliefs may be about his own fixed traits: seeing oneself as smart, attractive, and good is intrinsically more satisfying than the reverse. Alternatively, the beliefs can operate through *anticipatory utility*, meaning that the

[3] Sophisticated individuals who anticipate that they might be subject to "cognitive overload" may decline to receive information, but without any hostility.

individual experiences pleasant or aversive emotions from thinking about future (date-2) welfare: health or serious disease, successful marriage or messy divorce, riches or bankruptcy, eternal life or nothingness. Hope, fear, anxiety, and related emotions are important determinants of well-being, both as pure "mental consumptions" and through the psychosomatic, substance-abuse, and relational problems they induce. Under this broad class of preferences, a tradeoff clearly arises at date 0: one can react to bad news objectively, which leads to better decisions but having to live with grim prospects for some time (and possibly a long time), or adopt a more "defensive" cognitive response that makes life easier until the day of reckoning, when mistakes will have to be paid for.

Complementing these two sources of demand for "good" beliefs, the most frequent supply-side building block in the motivated-thinking paradigm is *selective (or differential) updating*, namely processing good and bad signals asymmetrically in term of *attention, interpretation, memory, or awareness*. While psychologically and neurally quite distinct (as discussed in the next section), these mechanisms are *formally equivalent* in terms of updating and behavioral consequences. To avoid repetition, we will often use the selective-recall interpretation, but it should *not* be taken literally. Realism then corresponds to appropriately coding L as L in memory, and denial to miscoding L as H, recalling it as an ambiguous mixture of the two, or (closest to standard information economics) forgetting the news entirely.

A more roundabout "belief-production" process, also based on imperfect accessibility of past states, is *self-signaling*: using our own behaviors as diagnostic of who we are, and conversely making choices with an eye toward our longer-run sense of identity. Because material actions are more easily codified, recalled, and documented than the exact mix of motives that caused them (evaluation of a hard tradeoff, momentary urges, feelings of guilt and pride), our past conduct can be informative about our "deep" preferences and predictive of later behaviors; yet at the same time, our choosing these self-signals makes future beliefs malleable.

Of course, the process of manipulating one's own attention, memory, or awareness must not be too transparent. There must be some opaqueness as to what exactly one is failing to update to, some ambiguity as to why certain actions are taken or not taken—such as crossing the street when seeing a beggar, starting or avoiding a fight, helping someone, or getting drunk. The thinnest of veils will often suffice however, as demonstrated by a number of experiments on "moral wiggle room" in which subjects seize upon or even seek out threadbare excuses for dishonesty, try to ignore or delegate their harming of others, and so on. (Konow 2000; Dana, Weber, and Kuang 2007; Gneezy, Saccardo, Serra-Garcia, van Veldhuizen 2014; Hamman, Loewenstein, and Weber 2010; Di Tella, Perez-Truglia, Babino, and Sigman 2015; Grossman and van der Weele forthcoming).

Naïveté is not needed for the key results (it just makes them stronger). A sophisticated individual knows that he has a tendency toward selective, self-serving attention, recall, and rationalizations. Such "metacognition" leads him to discount somewhat the "absence" of bad news at date 1, but as long as the sophisticated individual cannot fully reconstruct the censored or distorted original information, his

posterior will remain inflated. Where sophistication matters is in making possible different "cognitive styles" as alternative personal equilibria. An agent who is aware that his updating is very selective will discount the "news is good" states of awareness substantially, thereby making it safer to censor or misinterpret bad news when it occurs. Conversely, when someone tends to be "honest with himself," good news can be taken at face value, and this self-trust generates strong behavioral responses that make denial of bad news too dangerous.

First Implications and Evidence

From this framework, a number of predictions can be derived and confronted with data.

1. Information Avoidance and Asymmetric Updating

When asset-like beliefs are involved, people will tend to ignore, discount, rationalize away, or "put out of mind" news that conflicts with these ideas while welcoming data that supports them. Möbius et al. (2010) and Eil and Rao (2011) how show that subjects defend their beliefs concerning their IQ and (in the latter paper) also their attractiveness. The experimenters first elicit (in an incentive-compatible fashion) the prior distribution of the beliefs of every participant about being in each decile of the subject pool, then their updated beliefs following each of two rounds of objective feedback in which they learned whether they ranked above or below another, randomly drawn subject. Both studies find a statistically significant *good news/bad news asymmetry,* as predicted by our theory: subjects systematically under-update to negative signals, and are much closer to Bayesian updating for positive ones.[4]

Asymmetry also shows up in the demand for—or the avoidance of—information. In both studies, subjects' willingness-to-pay for learning their true IQ or/and beauty rank at the end of the experiment was positive for those who had arrived at "good" posteriors but *negative* for those who had arrived at "bad" ones, just as patients whose history and symptoms put them at high risk for some major disease often refuse to be tested. Similarly, investors studied in Karlsson, Loewenstein, and Seppi (2009) go online to look up the value of their portfolios much more on days when the market as a whole is up. Gottlieb (2014) formally shows how such conditional *informational preferences* arise from the general selective-recall model.

Wiswall and Zafar (2015) elicit college students' beliefs about their own future earnings and the average earnings in different majors. Then, they provide

[4] Bénabou (2013) shows how the model can generate strict under-updating (relative to Bayes' rule) to bad news and a lesser under-adjustment (possibly none) to good news. Gottlieb (2010) shows that the agents' (endogenous and motivated) failure to learn bad news persists even in an infinite-horizon setting, with signals or feedback received in every period.

the actual figures for each major, and elicit subjects' updated beliefs about their expected incomes. An underestimation of population earnings by $1,000 results in an upward revision in own earnings of $347 (significant at 1 percent), compared with a downward revision of just $159 for an overestimation (significant only at 10 percent).

Asymmetric responses to good and bad news, in turn, readily produce the so-called "Lake Wobegon" effect—that is, a distribution of posteriors where a very high fraction of people see themselves as above average. This holds true even for sophisticated agents, whose posterior beliefs must average back to the population mean, as Bayes' rule does not constrain skewness (Carrillo and Mariotti 2000; Bénabou and Tirole 2002).

2. The Role Of Memory and Other Neural Processes

Several complementary and *de facto* equivalent cognitive mechanisms can sustain motivated updating, but the simplest one is selective recall or accessibility of past signals, which is also relatively easy to test. The first experiment of that type in economics is Thompson and Loewenstein (1992), who show that: i) subjects assigned to represent different parties in a labor negotiation and given the same materials (from an actual case) recall, later on, more facts favoring their side than the other; and ii) these egocentric recall differences were associated with longer (hence costlier) delays during the negotiation phase.

More recently, Chew, Huang, and Zhao (2013) have subjects take four questions from an IQ test. Two months later they are shown the same four questions, plus two they had never seen, together with all the answers, and are *incentivized* to recall how they answered each one, or if they did not encounter it before. The probability of "remembering" having correctly answered a question which one actually failed is six times as high as the probability of the reverse error. The probability of not remembering one's answer, or whether one saw a question, is on average twice as high if the answer was wrong than if it was right. As for the questions they had never seen, 56 percent of subjects "remembered" answering them correctly versus 9 percent incorrectly. Furthermore, the three types of positive-attribution recall biases were highly correlated across subjects.

Work in neuroscience is starting to explore the deep mechanisms involved in differential recall and updating. Benoit and Anderson (2012) show that people are able to lower their later recall rates (for word pairs) by either blocking associations as they start to resurface or by focusing on different thoughts, and that different brain networks are involved in these two processes of *voluntary forgetting*. Sharot, Korn, and Dolan (2012) confirm the general finding of asymmetric updating to good and bad news and show that, while the "raw" data are well remembered by their subjects, distinct regions of the prefrontal cortex track and code for positively versus negatively valenced (more or less desirable) implied estimation errors. Furthermore, highly optimistic individuals consistently exhibit reduced tracking of negative estimation errors (which require updating in the direction of bad news).

3. Costs and Salience

Beliefs for which the individual cost of being wrong is small are more likely to be distorted by emotions, desires, and goals. An example often given is voting, as the cost of holding mistaken political opinions is usually said to be proportional to the probability of being pivotal, and hence extremely small (Caplan 2007). In reality, it need not be, due to social and self-signaling costs of political convictions. The more difficult question lies elsewhere, however. For the cognitive distortions of voters to have policy implications, it is necessary that a majority of them occur in the same direction. How such *ideological alignments* may occur and become dominant will be discussed in a later section, after extending the basic framework to social cognition.

4. Stakes-dependent Beliefs

Consider an agent with anticipatory utility who entered period 0 with some illiquid asset—housing, over-the-counter securities, specialized human or social capital, culture, or religion—that, at time 2, will be more valuable in state H than in state L. The incentive to self-deceive following bad news is clearly stronger, the greater is the amount of "sunk" capital with which the agent is initially endowed. This key implication of the motivated-cognition framework, which we term s*takes-dependent beliefs*, was first demonstrated in psychology by Kunda (1987).

Babcock, Loewenstein, Issachroff, and Camerer (1995) provide further evidence with an incentivized economic experiment. Pairs of subjects were given the same case file from a lawsuit over a traffic accident and were randomly assigned to be either the advocate for the plaintiff or for the defendant. They then bargained over a monetary settlement, with costs of delay. Based on the common materials they received, both sides also (independently) made incentivized predictions as to what outsiders would deem fair and how the judge ruled on the case. When roles were assigned *before* subjects saw the materials, they made highly divergent predictions of fairness and legal outcomes, and incompatible demands, leading to costly delays and breakdowns in bargaining. When roles were assigned *after* the information-processing stage, in contrast, there was far less asymmetry and delay.

In Mijovic-Prelec and Prelec (2010), subjects made incentivized predictions about a series of binary events, both before and after being (randomly) given stakes in the outcomes. When the stakes were such that their initial forecast corresponded to a low-payoff state, subjects showed a significant propensity to reverse their prediction. This is not just inconsistent with rational expectations, but also the *exact opposite* of confirmation bias. Mayraz (2011) has subjects randomly assigned to being "farmers" or "bakers" forecast the price at which they will later trade. Their predictions again vary systematically and optimistically with their positions, as well as with the size of the monetary stakes involved in facing favorable terms of trade. In the field, Di Tella, Galiant, and Schargrodsky (2007) document how land squatters randomly granted property rights adopted more "pro-market" beliefs (possibility of succeeding on one's own, importance of money for happiness), relative to their less-lucky neighbors.

5. Sunk-Cost Fallacy, Escalating Commitment, and the Hedonic Treadmill

A person who starts with enough of some illiquid or sunk asset, generating strong incentives to persuade oneself of its future value, experiences a type of endowment effect. Once persuaded, he will want to invest more in this capital, succumbing to a form of the sunk-cost fallacy that psychologists refer to as *escalating commitment*. Furthermore, although the agent is optimizing at every point in time given current preferences and beliefs, the ex-ante welfare implications of such ratcheting accumulation or specialization can be negative (Bénabou and Tirole 2011). The easiest way to understand this *hedonic-treadmill* result is to think of the case where self-signaling through personal actions is the way the agent tries to manipulate his own beliefs, and to recall that signaling usually involves a deadweight loss. More generally, censoring bad news or trying to offset it through identity-enhancing behaviors can prevent a deterioration of beliefs (like moral self-esteem) in bad states, but such censoring also reduces confidence that good states are really what they seem to be (creating self-doubt). When agents are sophisticated and beliefs enter preferences linearly, the two effects cancel out, leaving only the costs of generating, and then acting on, incorrect beliefs.[5]

Social and Organizational (Mis)Beliefs

Investigation reports following public-agency and corporate disasters commonly describe how willful blindness and reality denial spread within the organization, leading to systemic failures. A large literature in organizational psychology similarly emphasizes the key roles of moral self-deception and overoptimistic hubris in misconduct and financial fraud (Tenbrunsel and Messick 2004; Anand, Ashforth, and Mahendra 2004; Bazerman and Tenbrunsel 2011; Schrand and Zechman 2012). People engaging in reckless or dishonest behavior find ways to convince *themselves* that they are doing nothing wrong, so transgressions that typically start small gradually escalate through a series of rationalizations, which are then further insulated from reality by "echo chamber" group dynamics. For instance, the NASA (2003, vol. 1, pp. 196-199) investigations following the Challenger and Columbia space shuttle accidents found that:

> NASA appeared to be immersed in a culture of invincibility, in stark contradiction to post-accident reality. The Rogers Commission found a NASA blinded by its 'Can-Do' attitude … which bolstered administrators' belief in an achievable launch rate, the belief that they had an operational system, and an unwillingness to listen to outside experts … At every juncture, the Shuttle Program's

[5] The case of linear utility-from-beliefs is a useful benchmark. Clearly, if the functional is instead concave (respectively, convex) in beliefs, the agent will gain from achieving coarser (respectively, more dispersed) posteriors. The actual shape of self-esteem or anticipatory preferences is, ultimately, an empirical question.

> structure and processes, and therefore the managers in charge, resisted new information … [E]vidence that the design was not performing as expected was reinterpreted as acceptable and non-deviant, which diminished perceptions of risk throughout the agency … Engineers and managers incorporated worsening anomalies into the engineering experience base, which functioned as an elastic waistband, expanding to hold larger deviations from the original design.

Strikingly similar patterns recurred at companies like Enron and General Motors and, prior to the 2008 financial crisis, at major investment banks, the insurance company AIG, the Federal Reserve, and the Securities and Exchange Commission (Bénabou 2013, Appendix D).

How can such motivated thinking and reality denial become "contagious" and spread through an organization or some of its units? Consider a setting in which individuals are embedded in a firm, network, or other collective endeavor. To highlight the endogenous emergence of interdependence in how people *think* and perceive events, let us assume (without loss of generality) a simple, linear interaction structure: each agent's final payoff is a weighted average of his own action and the group's average action, all multiplied by a common (gross) return. In the good state of the world, both private and social (group-wide) net returns are positive. In the bad state, the (net) private return is always negative, but depending on the nature of spillovers, the public return could be positive or negative. This last factor turns out to be critical for how groups respond to bad news, and whether a collective failure to update represents beneficial *group morale* or harmful *group delusions*.

In the case of projects with no or little social downside, like team effort or mobilization for a good cause, blind perseverance in the face of bad news is individually suboptimal but constitutes a *public good*. The overoptimism of others thus makes the bad state more tolerable, and therefore each individual more willing to accept its reality: cognitive attitudes are thus *strategic substitutes* (they tend to dampen one another), and denial is self-limiting.[6]

The more interesting case is that of ventures with important downside risk, in which blind persistence can inflict further losses on others, such as capital and reputational losses, firm bankruptcy, layoffs, catastrophic accident, or prosecution. The more people fail to attend to bad news and continue doing "business as usual," the worse the bad state becomes, making it even harder to face the impending disaster. Perceptions of reality are now *strategic complements,* so delusions will spread.

This Mutually Assured Delusion (MAD) mechanism is rather perverse, as denial and reality avoidance become *contagious* when they are socially *harmful,* but *not* when they are *beneficial*. The underlying intuition is straightforward: we saw earlier how each individual tends to align their beliefs with the fixed stakes they have in different

[6] In a sufficiently asymmetric interaction structure, it can even be that some agent who can short-sell the project gains so much from others' denial of state L that he prefers it to H. In that case, he will have a tendency to believe in H rather than L. This strong cognitive substitutability can lead two (sets of) agents to take opposite sides of a bet on which state will realize, as in Brunnermeier and Parker (2005).

states of the world. In a group or network, these stakes now depend on what other people do (do they generate positive or negative spillovers?), and hence on what they *believe,* in those states. It follows that what is optimal for each agent to think *depends on what others think,* and vice versa. Furthermore, the nature and welfare consequences of these cognitive linkages depend quite simply on the sign of externalities in the network structure, rather than on any built-in nonlinearities in payoffs.

This "psychological multiplier" leads to the possibility of *multiple social cognitions:* fundamentally similar groups or organizations can operate either in a *realistic mode* where everyone faces the facts as they are, or in a *delusional mode* in which everyone engages in denial of bad news, which in turn makes those states even worse for everyone else. Bénabou (2013) shows that such "groupthink" is more likely: i) when codependency among group members is high, meaning that they share a largely common fate, with few exit options from the collateral damage inflicted by others' mistakes; and ii) when the adverse state of the world is relatively rare but, when it occurs, really bad—a so-called "black swan" event.

These ideas and results readily extend to *asymmetric networks* and organizations: an agent's propensity to realism or denial depends most on how the people whose decisions have the strongest impact on his fate respond to bad news themselves. Therefore, in a hierarchy, top management's (mis)perceptions of market prospects, legal liabilities, or odds of victory will tend to trickle down to middle echelons, and from there on to workers or troops.

We have emphasized in this section the importance of "bad beliefs" mechanisms of organization failure, which have so far received too little attention from economists relative to the standard "bad incentives" mechanisms. In practice, most failures have both channels at work, and how they feed into each other represents a promising avenue for future research.

Political Ideology

The study of political economy is also undergoing a pendulum swing of perspectives. An emphasis on the strategic choices of rational voters and pressure groups pursuing their material self-interest remains indispensable, but it is increasingly being complemented by "behavioral" considerations like the expression of identities and emotions, reference-dependent concerns such as fairness or loss aversion, biased attributions (like scapegoating), and ideological or wishful denials of reality. The ongoing political events and campaigns in the United States and a number of European countries should, if need be, dispel any remaining doubts about the relevance of such psychological factors in politics.

While each voter may choose to maintain beliefs which they value for affective or instrumental reasons, in order for this to have policy consequences these worldviews must somehow align within a country, while potentially diverging across borders. We now provide examples of how such complementarities in political beliefs can arise rather naturally, leading to the emergence and persistence of *dominant ideologies.*

Just-World Beliefs

A "just world" is one in which "people get what they deserve, and deserve what they get" (Lerner 1980). Do they? The World Values Survey reveals considerable differences in beliefs about the role of effort versus luck in life. In the United States, 60 percent of people believe that effort is key; in Western Europe, only 30 percent do on average, with major variations across countries. Moreover, these nationally dominant beliefs bear no relationship to the actual facts about social mobility or how much the poor are actually working, and yet they are strongly correlated with the share of social spending in GDP (Alesina, Glaeser, and Sacerdote 2001). At the individual level, similarly, voters' perceptions of the extent to which people control their own fate and ultimately get their just desserts are first-order determinants of attitudes toward inequality and redistribution, swamping the effects of own income and education (Fong 2001).

In Bénabou and Tirole (2006), we describe how such diverse *politico-ideological equilibria* can emerge due to a natural complementarity between (self-)motivation concerns and marginal tax rates. When the safety net and redistribution are minimal, agents have strong incentives to maintain for themselves, and pass on to their children, beliefs that effort is more important than luck, as these will lead to working hard and persevering in the face of adversity. With high taxes and generous transfers, such beliefs are much less adaptive, so fewer people will maintain them. Thus, there can coexist: i) an "American Dream" equilibrium, with just-world beliefs about social mobility, and little redistribution; and ii) a "Euro-pessimistic" equilibrium, with more cynical beliefs and a large welfare state. In the latter, the poor are less (unjustly) stigmatized as lazy, while total effort (annual hours worked) and income are lower, than in the former. More generally, across all steady-states there is a negative correlation between just-world beliefs and the size and the welfare state, just as observed across countries.

Complementing this national-level evidence, Frank, Wertenbroch, and Maddux (2015) experimentally validate the role (and malleability) of just-world beliefs in determining distributional preferences. Using MBA participants from 30 countries, they find that: i) subjects' priors on the effort-versus-luck question predict their preferences toward redistribution of earnings from a task performed in the lab; ii) aggregating these beliefs at the national level yields a predictor of preferences (from national surveys) for performance pay versus redistributive pay; and iii) "priming" just-world beliefs so that they are more prominently in mind at the start of the experiment has a causal effect on individuals' choices over which pay system to impose in their session.

Statist and Laissez-Faire Ideologies

A similar international divergence is observed for beliefs in the merits of "the free enterprise system and free market economy." The average degree of agreement that this is "the best system on which to base the future of the world" was 61 percent in the 2005 World Public Opinion Survey. Countries near the top include China at 74 percent, the United States at 71 percent, and Germany at 65 percent. Those at

the bottom include Argentina at 42 percent, Russia at 43 percent, and France at 36 percent. Here again the objective facts belie these divergent worldviews; Germany and France, for instance, have very similar economic structures but an almost two to one divergence on this survey question. Yet again, these beliefs are highly predictive of the size of government, whether measured by the tax-to-GDP ratio or by indices of labor and product market regulation.

Bénabou (2008) documents this international divergence and shows how such ideological differences can be sustained. When people expect to be "living with" and paying for a large public sector, the psychological incentive is to view it as an important source of future benefits (anticipatory utility), which in turn makes one more willing to vote for it. Conversely, when people anticipate having to purchase these services in the market (as in the United States in the case of health insurance), the incentive is to think of the latter as efficient and see less need for public provision and funding. Individual voters' beliefs can thus again be mutually amplifying, leading to history-dependent dynamics and multiple steady-states: a "Statist" one featuring a large government and obstinate beliefs in its benevolent efficacy, a "Laissez-Faire" one with a small government and equally inflexible beliefs in the virtues of the invisible hand, or a "Realistic" one in which voters acknowledge both state and market failures.

Pandering Politicians

If voters have demands for rosy beliefs—say, painless solutions to economic and social problems, external scapegoats, or feel-good demonstrations of power—office-motivated politicians and profit-maximizing media will gladly oblige. As shown by Levy (2014), this feedback can lead to a "Soothing Politics" equilibrium, which features no reform even when needed, hence much pain down the line. This prospect increases voters' incentives to forget or rationalize bad news, and, in turn, their inattention and wishful thinking allow politicians whose interests are noncongruent with those of the electorate to indulge in the easy life of no reform. Conversely, a "*Realpolitik*" equilibrium can emerge in which voters remain aware of negative signals. In this case, politicians must follow up with reform, to avoid appearing noncongruent (lazy, incompetent, or captured by lobbies) and being voted out.

Overconfidence, Polarization, and Extremism

Within each polity there are also sharply divided beliefs, often with a tendency toward polarization rather than convergence. Here again, the perspective of voters bending their thought processes and worldviews to fit their needs and desires provides a useful explanatory framework. Agents whose material, social, human-capital, and cultural endowments give them different stakes in various states of the world (large or small role of effort; efficient or inefficient government; degree of trustworthiness of others) will process and interpret the same signals very differently. A greater divergence of beliefs even as more and more information becomes commonly available through the global media and internet is thus not really a puzzle.

This class of models also implies that people will seek interactions with those who think like them, and shun those whose words or actions provide signals and reminders that threaten valued "constructed realities" (Bénabou and Tirole 2011). There is an important role here for Cassandras, who speak unpleasant truths, to guard against the group falling prey to costly delusions (this also ensures that good news, or the absence of bad news, is genuine). Yet these truth-tellers will be cast away, or worse, once a bad state does occur, especially if some investments have already been sunk (Bénabou 2013). This *time-inconsistency* in attitudes toward dissent provides a new rationale for social commitment mechanisms such as constitutional rights to free speech, independence of the press, and so forth. Not only does the public presence of dissenting views help to ensure realism and confidence in the available information, but the anticipation that they will undermine wishful beliefs also lowers the return to engaging in motivated thinking in the first place.

Overconfidence and Ideology

The importance of overconfident beliefs in politics and their resistance to information, particularly among extremists, is documented in Ortoleva and Snowberg (2015a, b). A nationally representative sample of over 3,000 American adults was asked standard political-survey questions, and also to provide estimates and degrees of confidence for the current and next year's rates of inflation and unemployment. The study finds that: i) more overconfident agents have more extreme political views, and higher turnout rates in elections (thus cognitive distortions really matter); ii) overconfidence does not decrease with education, and it increases with both age and media exposure (polarization); and iii) it is found both on the Left and the Right of the political spectrum, though since 1980 more so on the Right. To explain these findings, Ortoleva and Snowberg (2015a) propose that voters suffer from heterogeneous degrees of "correlational neglect"—that is, they fail to take account that the observations they derive from their local environment, social network, and chosen information sources are largely redundant.

Such failures of Bayesian inference and biased information seeking and processing arise very naturally from a motivated-beliefs, identity-maintenance perspective. Thus, Le Yaouanq (2015) incorporates preference heterogeneity into the selective-awareness framework exposited above. He shows that the typical result of multiple dominant ideologies remains and, more interestingly, that: i) the more extremist agents in the political spectrum are the ones most prone to engage in reality denial; and ii) when agents can endogenously form networks within which political views will be exchanged or observed, *ideological homophily* will tend to prevail, making collective biases and polarization more likely (Della Vigna and Kaplan 2007; Gentzkow and Shapiro 2011).

Financial Bubbles and Crashes

The motivated-thinking framework also provides a psychologically grounded account for financial manias and crashes. Suppose that, following some initial good

news, investors have accumulated stocks of some financial asset that is relatively illiquid. Next, good or bad news may be received about fundamentals, and investors can then keep investing or stop. The liquidation price at date 2 will reflect the total supply accumulated up to that time and the realized demand for the asset. Downward-sloping demand makes investment decisions strategic substitutes, and thus contagion harder to sustain. Nonetheless, investors' *cognitive responses* to bad news can be *strategic complements* (they reinforce one another), giving rise to an "irrationally exuberant" buildup that further amplifies the coming crash. Indeed, when illiquid initial positions are sufficiently large, realism would require *recognizing*—in both senses of the term—early on major capital losses, made all the worse by the blind overinvestment of others. This capital-loss externality is the Mutually Assured Delusion (MAD) multiplier at work again, and when it dominates demand substitutability the market can be seized by periodic waves of contagious overoptimism, overheating, and meltdowns (Shiller 2005; Akerlof and Shiller 2009).

Cheng, Raina, and Xiong (2014) provide evidence supporting the relevance of such a mechanism in the real-estate-based financial bubble of 2003–2005. They examine the personal housing transactions of Wall Street "insiders"—a sample of 400 mid-level managers in the mortgage-securitization industry, who had a close-up view of the toxic subprime loans. Compared to sophisticated "outsiders"—lawyers and financial analysts not specializing in the real-estate sector—the insiders were *more* likely to buy a first, second, or larger house at the peak of the bubble, and slower to divest as housing prices started falling. As a result, they had a lower overall return on their own real-estate portfolios. The fact that insiders bought high and sold low goes against standard, rational moral-hazard explanations of the crisis, but it is very consistent with the mechanisms of escalating commitment and groupthink in which beliefs about future housing prices become badly distorted by personal and industry-wide stakes.

Identity and Morality

Psychologists, sociologists, and more recently economists (starting with Akerlof and Kranton 2000) have emphasized the central role that identity plays in determining social behavior. Identity is in essence a set of beliefs: about one's character, preferences, moral or religious values, abilities, and prospects (personal identity); and about where one belongs—within a family, firm, network, culture, or nation (group identity). Identity pertains to beliefs that people *value*—and therefore defend. For example, beliefs in an afterlife clearly affect anticipatory utility, just-world beliefs provide both motivation to act and a sense that life is somewhat predictable, and trust in others makes us more optimistic about the society we live in.

Personal Identity

There are obvious reasons why people would want to be seen by others as honest and prosocial. But the desire to think of *oneself* as a moral person—and

the concomitant monitoring and judgment of one's behavior—is more subtle to explain.

A first benefit of maintaining such self-respect is to help resist short-run temptations to act opportunistically (cheating, defecting) or rashly (impulses toward sex, anger, or violence) which are likely to have detrimental long-run consequences for one's social relationships. Another adaptive benefit is that deceiving others is not that easy (we are "programmed" to blush or more broadly to give ourselves away), so that believing one's own "line" helps one profess it more credibly in public. In an experiment with incentive-compatible belief elicitations, Schwardman and van der Weele (2016) find that after performing a cognitively challenging task, subjects were 50 percent more overconfident in their relative performance, and less responsive to objective feedback on it, if informed in advance that they could later earn money by convincing others, face-to-face, of having scored high. Furthermore, this belief-management strategy was effective, as subjects in this condition did receive better assessments (also incentivized) from evaluators.

On the other hand, any form of reality distortion or wasteful signaling has costs as well: for instance, forcing us to act more generously than we really would like to, even in anonymous settings. This explains why, as demonstrated by a series of "moral wiggle room" experiments, people often seek excuses and situational ambiguities that avoid putting their moral identities to an explicit and necessarily costly test (in this issue, Gino, Norton, and Weber discuss the evidence on this point). Di Tella et al. (2015) actually elicit the beliefs generated in the process using a variant of the "trust game."[7] When given the opportunity to unconditionally confiscate all of the trustee's earnings, without *any* knowledge of how he or she had actually behaved, trustors became significantly more likely to predict (with incentives) that the trustee had chosen a "corrupt" action and then "punish" them by taking their money.

Social Identity

Many core identities relate to "belonging," such as identifying with a town, ethnic or cultural group, profession, religion, political party, or public cause. A first explanation may be that we derive material benefits from being part of a community. Berman and Iannaccone (2006) thus note that a number of religious groups provide "club goods" such as insurance against economic or health shocks, help in finding a spouse or job, and assistance with raising children. However, it is not at all obvious why public-goods clubs would have to rely on supernatural beliefs (which are absent from these models), as opposed to membership fees in money or in kind, reputational enforcement, or other signaling devices.[8] It is also not clear why religious beliefs and identities should be shared and sometimes violently defended across the world, when the above public goods are in most cases extremely local.

[7] In the canonical trust game, one player chooses how much of their endowment to entrust to another one. This investment gets multiplied (say, by three), then the "trustee" decides how much to return to the "trustor."

[8] On the choice between costly beliefs and costly rituals as signals, see Levy and Razin (2014).

A second explanation for a general "desire to belong" is evolutionary: humans are a social species, deriving an intrinsic satisfaction from interactions with similar others and deep anxiety from isolation—emotional incentives that serve to promote the fitness benefits of community. Even today, the quasi-automatic nature of the in-group/out-group phenomenon, in which shared identities and beliefs form almost immediately on the basis of meaningless and random group assignments, speaks to our need not to be left in a social no-man's land.

Beliefs as Identity Capital

Although economic treatments of identity generally describe it as a set of beliefs, in practice they often model it as preferences, or meta-preferences over utility functions (Akerlof and Kranton 2000; Rabin 1994; Shayo 2010). Our approach is cognitive, in that it explicitly models identity as beliefs about one's deep values, emphasizing the self-inference process through which beliefs operate and the underlying needs that identity serves.

The individual starts with a stock of identity-relevant capital that will affect future welfare. This stock can be viewed as fixed (gender, race) or augmentable (friends, professional accomplishments, wealth, religious faith). Critically, the agent is uncertain about how much this capital will contribute to his own welfare over the long run. Thus, an immigrant may at times be unsure of how attached he is to his original culture relative to the benefits of integration; a professional, of whether more accomplishments and wealth or more time with the family will make for a happier life; and a religious person, of the true strength of his faith. As explained earlier for "stakes-dependent" beliefs more generally, a positive view of the future returns to these stocks directly raises anticipatory utility, and may also enhance self-regulation.

This self-reputational approach to identity sheds light on many otherwise puzzling aspects of behavior, particularly in the moral domain. First, identity-enhancing behaviors are more likely when objective information about deep preferences is scarce (like true generosity, loyalty, or faith), and they are easily affected by minor manipulations of salience such as cues, reminders, and semi-transparent excuses that can be used to muddle personal responsibility. Second, whereas challenges to a weakly held identity (prior belief) may elicit acquiescence, challenges to a strongly held identity generally elicit forceful counterreactions. Another form of history-dependence is "hedonic treadmill" effects: when agents are endowed with sufficient identity capital, they will tend to keep investing even in the face of negative net returns. In some cases, a person's conflicting identities can actually generate self-destructive behaviors: rejecting education and integration into the mainstream labor market, unwillingness to adapt to societal or economic change, going to fight in a faraway land, or suicide bombings. In a milder example of this phenomenon, Burzstyn et al. (2014) found that 25 percent of male experimental subjects in Pakistan chose to forego a participation bonus equivalent to one-fifth of a day's wage when receiving it required anonymously checking a box indicating gratitude toward the US government for the funds.

A beliefs-based model also naturally generates *sacred values* and mental *taboos* (not just social ones), characterized by a strong aversion to even *thinking* about violations: the mere contemplation of tradeoffs between some "higher" principles and self-interest suffices to cast lasting doubt on one's identity (Fiske and Tetlock 1997; Bénabou and Tirole 2011). This creates the potential for significant reality distortion in realms where identity concerns loom large, and can also account for the desire to ban "repugnant markets" where *others* might too visibly engage in certain taboo transactions: prostitution, organ sales, payments for adoption or surrogate pregnancy, and so on.

Ostracism is another natural implication. Since the preferences and prospects of similar individuals are often correlated, "deviant" behavior by peers—violating norms and taboos, fraternizing with outsiders—conveys bad news about the value of existing social assets (anticipatory-utility motive) or that of future investments in them (imperfect self-control motive). On the other hand, if the morally dubious action was one's own, it is good behavior by others that becomes threatening, as it takes away potential excuses involving situational factors. In *both* cases, ostracizing the deviators suppresses the undesirable *reminders* created by their presence. Thus, depending on who acted more (im)morally, the same person or group will act pro-socially and shun free riders, or act selfishly and shun moral exemplars (for empirical evidence, see Monin 2007, Herrmann, Thöni, and Gächter 2008; on the demand for belief consonance and the "tyranny of small differences," see also Golman et al. in this volume).

Identity investments also help explain inefficiencies in bargaining, contracting, and the functioning of organizations. The failure to reach efficient Coasian deals—leading to legal trials, divorces, strikes, scapegoating of minorities in hard times, wars, and so on—is usually explained by economists through informational asymmetries about gains from trade and outside options. Evidence is accumulating, however, that belief distortions also play a key role in those phenomena, with field studies such as Bewley (1999) and Krueger and Mas (2004) complementing the previously mentioned experiments of Babcock et al. (1995). Pride, dignity concerns, and wishful thinking commonly lead people or groups to walk away from "reasonable" offers, try to shift blame for failure onto others, and destroy surplus and seek refuge in political utopias, resulting in costly impasses and conflicts. (For a model of bargaining with motivated beliefs, see Bénabou and Tirole 2009.)

Conclusion

The basic utility function based on consumption and leisure (or even social payoffs) was always recognized as a simplification—defensible in many cases, less so in others. The theory of motivated cognition broadens the purposefulness of human behavior along a variety of dimensions. Some beliefs and emotions are affectively more pleasant than others, like hope and confidence over fear and anxiety. People receive utility from having a positive self-image, and from thinking of themselves as

belonging to groups. Optimistic beliefs can also be valuable motivators to overcome self-control problems, as well as helpful in strategic interactions.

In such situations, people will tend to manipulate their collection and processing of information in ways that depart from strict Bayesian inference, trading off the affective or functional value of belief distortions against the costly mistakes they also induce. It may seem that they are just displaying limited cognitive abilities due to some the biases discussed in the large behavioral-economics and bounded-rationality literatures. Instead, the theory of motivated beliefs emphasizes that many observed departures from standard rationality are not hard-wired or mechanical but can instead be understood within a broadened context of goal-directed (but not necessarily efficient) individual behavior. This, in turn, leads to novel views of risk-taking, prosociality, identity, organizations, financial crises, and politics.

■ *We are grateful to S. Nageeb Ali, Rafael DiTella, Gordon Hanson, Ulrike Malmendier, Enrico Moretti, and Timothy Taylor for valuable comments. Bénabou gratefully acknowledges financial support from the Canadian Institute for Advanced Research, and Bénabou and Tirole from the European Research Council (Grants No. FP7/2007-2013-249429 and No. 669217).*

References

Akerlof, George, and William T. Dickens. 1982. "The Economic Consequences of Cognitive Dissonance." *American Economic Review* 72(3): 307–19.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115(3): 716–753.

Akerlof, George A., and Robert J. Shiller. 2009. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism.* Princeton University Press.

Alesina, Alberto, Edward Glaeser, and Bruce Sacerdote. 2001. "Why Doesn't the United States Have a European-Type Welfare State?" *Brookings Papers on Economic Activity*, no. 2, pp. 187–277.

Alloy, L. B., and L. Y. Abrahamson. 1979. "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?" *Journal of Experimental Psychology: General* 108(4): 441–85.

Anand, Vikas, Blake E. Ashforth, and Joshi Mahendra. 2004. "Business as Usual: The Acceptance and Perpetuation of Corruption in Organizations." *Academy of Management Perpectives* 18(2): 39–53.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review* 85(5): 1337–43.

Bazerman, Max H., and Ann E. Tenbrunsel. 2011. *Blind Spots: Why We Fail to Do What's Right and What to Do About It.* Princeton University Press.

Bénabou, Roland. 2008. "Ideology." *Journal of the European Economic Association* 6(2–3): 321–52.

Bénabou, Roland. 2013. "Groupthink: Collective Delusions in Organizations and Markets." *Review of Economic Studies* 80(2): 429–62.

Bénabou, Roland. 2015. "The Economics of Motivated Beliefs." Jean-Jacques Laffont Lecture, *Revue d'economie politique* 125(5): 665–85.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117(3): 871–915.

Bénabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy* 112(4): 848–87.

Bénabou, Roland, and Jean Tirole. 2006. "Belief in a Just World and Redistributive Politics." *Quarterly Journal of Economics* 121(2): 699–746.

Bénabou, Roland, and Jean Tirole. 2009. "Over My Dead Body: Bargaining and the Price of Dignity." *American Economic Review* 99(2): 459–65.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126(2): 805–55.

Benoit, Roland G., and Michael C. Anderson. 2012. "Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories." *Neuron* 76(2): 450–60.

Berman, Eli, and Laurence Iannaccone. 2006. "Religious Extremism: The Good, the Bad, and the Deadly." *Public Choice* 128(1): 109–129.

Bewley, Truman F. 1999. *Why Wages Don't Fall During a Recession*. Cambridge, MA: Harvard University Press

Bodner, Ronit, and Drazen Prelec. 2003. "Self-Signaling and Diagnostic Utility in Everyday Decision Making." In *The Psychology of Economic Decisions*, vol. 1, edited by Isabelle Brocas and Juan D. Carrillo. Oxford University Press.

Brocas, Isabelle, and Juan D. Carrillo. 2001. "Rush and Procrastination under Hyperbolic Discounting and Interdependent Activities." *Journal of Risk and Uncertainty* 22(2): 141–64.

Brunnermeier, Markus K., and Jonathan A. Parker. 2005. "Optimal Expectations." *American Economic Review* 95(4): 1092–1118.

Bursztyn, Leonardo, Michael Callen, Bruno Fermanx, Saad Gulzar, Syed Ali Hasanain, and Noam Yuchtman. 2014. "Identifying Ideology: Experimental Evidence on Anti-Americanism in Pakistan." NBER Working Paper 20153.

Caplan, Bryan. 2007. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton University Press.

Caplin, Andrew, and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics* 116(1): 55–80.

Carrillo, Juan D., and Thomas Mariotti. 2000. "Strategic Ignorance as a Self-Disciplining Device." *Review of Economic Studies* 67(3): 529–44.

Case, Karl E., and Robert J. Shiller. 2003 "Is There a Bubble in the Housing Market?" *Brookings Papers on Economic Activity* no. 2, pp. 299–342.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven. 2013. "Self Confidence and Strategic Behavior." CESifo Working Paper 4517, December.

Cheng, Ing-Haw, Sahil Raina, and Wei Xiong. 2014. "Wall Street and the Housing Bubble." *American Economic Review* 104(9): 2797–2829.

Chew, Soo Hong, Wei Huang, and Xiaojian Zhao. 2013. "Selective Memory and Motivated Delusion: Theory and Experiment." Unpublished paper, National University of Singapore, February.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wriggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33(1): 67–80.

Dillenberger, David, and Philipp Sadowski. 2012. "Ashamed to Be Selfish." *Theoretical Economics* 7(1): 99–124.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics* 122(3): 1187–1234.

Di Tella, Rafael, Sebastian Galiant, and Ernesto Schargrodsky. 2007. "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters." *Quarterly Journal of Economics* 122(1): 209–41.

Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman. 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism." *American Economic Review* 105(11): 3416–42.

Eliaz, Kfir, and Ran Spiegler. 2006. "Can Anticipatory Feelings Explain Anomalous Choices of Information Sources?" *Games and Economic Behavior* 56(1): 87–104.

Eil, David, and Justin M. Rao. 2011. "The Good News–Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3(2): 114–38.

Eisenbach, Thomas M., and Martin C. Schmalz. 2015. "Anxiety, Overconfidence, and Excessive Risk-Taking." Staff Report 811, Federal Reserve Bank of New York.

Epstein, Larry G. 2008. "Living with Risk." *Review of Economic Studies* 75(4): 1121–41.

Falk, Armin, and Florian Zimmermann. 2011. "Preferences for Consistency." IZA Discussion Paper 5840, Institute for the Study of Labor.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford University Press.

Fiske, Alan Page, and Philip E. Tetlock. 1997. "Taboo Trade-offs: Reaction to Transactions that Transgress the Spheres of Justice." *Political Psychology* 18(2): 255–97.

Fong, Christina M. 2001. "Social Preferences, Self-Interest, and the Demand for Redistribution." *Journal of Public Economics* 82(2): 225–46.

Frank, Douglas H., Klaus Wertenbroch, and William W. Maddux. 2015. "Performance Pay or Redistribution? Cultural Differences in Just-World Beliefs and Preferences for Wage Inequality." *Organizational Behavior and Human Decision Processes* 130: 160–70.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19(4): 25–42.

Ganguly, Ananda, and Joshua Tasoff. Forthcoming. "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future." *Management Science.*

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126(4): 1799–1839.

Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen. 2014. "Motivated Self-Deception and Unethical Behavior." Unpublished paper.

Gottlieb, Daniel. 2010. "Will You Never Learn? Self-Deception and Biases in Information Processing." Unpublished paper, Wharton School, University of Pennsylvania.

Gottlieb, Daniel. 2014. "Imperfect Memory and Choice under Risk." *Games and Economic Behavior* 85: 127–58.

Grossman, Zachary, and Joël van der Weele. Forthcoming. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association.*

Hamman, John R., George Loewenstein, and Roberto A. Weber. 2010. "Self-Interest through Delegation: An Additional Rationale for the Principal–Agent Relationship." *American Economic Review* 100(4): 1826–46.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment across Societies." *Science*, March 7, 319(5868): 1362–67.

Kahan, Dan M. 2013. "Ideology, Motivated Reasoning, and Cognitive Reflection." *Judgment and Decision Making* 8(4): 407–24.

Kahan, Dan M., Ellen Peters, Erica Cantrell Dawson, and Paul Slovic. 2014. "Motivated Numeracy and Enlightened Self-Government." Cultural Cognition Project Working Paper 116, Yale Law School.

Karlsson, Niklas, George Loewenstein, and Duane Seppi. 2009. "The Ostrich Effect: Selective Avoidance of Information." *Journal of Risk and Uncertainty* 38(2): 95–115.

Konow James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90(2): 1072–91.

Korn, C. W., T. Sharot, H. Walter, H. R. Heekeren, and R. J. Dolan. 2014. "Depression Is Related to an Absence of Optimistically Biased Belief Updating about Future Life Events." *Psychological Medicine* 44: 579–92.

Köszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4(4): 673–707.

Köszegi, Botond. 2010. "Utility from Anticipation and Personal Equilibrium." *Economic Theory* 44(3): 415–44.

Krueger, Alan B., and Alexandre Mas. 2004. "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires." *Journal of Political Economy* 112(2): 253–89.

Kunda, Ziva. 1987. "Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories." *Journal of Personality and Social Psychology* 53(4): 636–47.

Lerner, Melvin J. 1980. *The Belief in a Just World: A Fundamental Delusion.* New York, NY: Plenum Press.

Levy, Raphaël. 2014. "Soothing Politics." *Journal of Public Economics* 120: 126–33.

Levy, Gilat, and Ronny Razin. 2014. "Rituals or Good Works: Social Signaling in Religious Organizations." *Journal of the European Economic Association* 12(5): 1317–60.

Le Yaouanq, Yves. 2015. "Political Values and the Polarization of Beliefs." Unpublished paper, Toulouse School of Economics, October.

Loewenstein, George. 1987. "Anticipation and the Valuation of Delayed Consumption." *Economic Journal* 97(387): 666–84.

Mayraz, Guy. 2011. "Wishful Thinking." Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1955644.

Merkle, C., and M. Weber. 2011. "True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence." *Organizational Behavior and Human Decision Processes* 116(2): 262–71.

Mijović-Prelec, Danica, and Dražen Prelec. 2010. "Self-Deception as Self-Signaling: A Model and Experimental Evidence." *Philosophical Transactions of the Royal Society,* B, 365: 227–40.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya Rosenblat. 2011. "Managing Self-Confidence: Theory and Experimental Evidence." NBER Working Paper 17014.

Monin, Benoit. 2007. "Holier Than Me? Threatening Social Comparison in the Moral Domain." *International Review of Social Psychology* 20(1): 53–68.

NASA. 2003. *Report of Columbia Accident Investigation Board.* http://www.nasa.gov/columbia/home/CAIB_Vol1.html.

Oliver, J. Eric, and Thomas J. Wood. 2014. "Conspiracy Theories and the Paranoid Style(s) of Mass Opinion." *American Journal of Political Science* 58(4): 952–66.

Ortoleva, Pietro, and Erik Snowberg. 2015a. "Overconfidence in Political Behavior." *American Economic Review* 105(2): 504–35.

Ortoleva, Pietro, And Erik Snowbert. 2015b.

"Are Conservatives Overconfident?" *European Journal of Political Economy* 40(Part B): 333–44.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review* 103(2): 804–830.

Puri, Manju, and David T. Robinson. 2007. "Optimism and Economic Choice." *Journal of Financial Economics* 86(1): 71–99.

Quattrone, George A., and Amos Tversky. 1984. "Causal versus Diagnostic Contingencies: On Self-Deception and on the Voter's Illusion." *Journal of Personality and Social Psychology* 46(2): 237–48.

Rabin, Matthew. 1994. "Cognitive Dissonance and Social Change." *Journal of Economic Behavior and Organization* 23(2): 177–94.

Schelling, T. C. 1988. "The Mind as a Consuming Organ." Chap. 15 in *Decision Making: Descriptive, Normative, and Prescriptive Interactions,* edited by David E. Bell, Howard Raiffa, and Amos Tversky. Cambridge University Press.

Schrand, Catherine M., and Sarah L. C. Zechman. 2012. "Executive Overconfidence and the Slippery Slope to Financial Misreporting." *Journal of Accounting and Economics* 53(1–2): 311–29.

Schwardman, Peter, and Joël van der Weele. 2016. "Deception and Self-Deception." Tinbergen Institute Discussion Papers 16-012/I.

Sharot, Tali, Christoph W. Korn, and Raymond Dolan. 2012. "How Unrealistic Optimism is Maintained in the Face of Reality." *Nature Neuroscience* 14(11): 1475–79.

Sharot, Tali, and Neil Garrett. 2016. "Forming Beliefs: Why Valence Matters." *Trends in Cognitive Science* 20(1): 25–33.

Shayo, Moses. 2009. "A Model of Social Identity with an Application to Political Economy: Nation, Class and Redistribution." *American Political Science Review* 103(2): 147–74.

Shiller, Robert J. 2005. *Irrational Exuberance,* 2nd Edition. Princeton University Press.

Smith, Adam. 1759. *The Theory of Moral Sentiments.*

Tenbrunsel, Ann E., and David M. Messick. 2004. "Ethical Fading: The Role of Self-Deception in Unethical Behavior." *Social Justice Research* 17(2): 223–62.

Thompson, Leigh, and George Loewenstein. 1992. "Egocentric Interpretations of Fairness and Interpersonal Conflict." *Organizational Behavior and Human Decision Processes* 51(2): 176–197.

Trivers, Robert. 2011. The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life. Basic Books.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science,* New Series, 185(4157): 1124–31.

Wiswall, Matthew, and Basit Zafar. 2015. "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital* 9(2): 117–169

Van den Steen, Eric. 2004. "Rational Overoptimism (and Other Biases)." *American Economic Review* 94(4): 1141–51.

Van den Steen, Eric. 2010. "On the Origins of Shared Beliefs (and Corporate Culture)." *RAND Journal of Economics* 41(4): 617–48.

Von Hippel, William, and Robert Trivers. 2011. "The Evolution and Psychology of Self-Deception." *Behavioral and Brain Sciences* 34(1): 1–56.

This article has been cited by:

1. Fuhai Hong, Wei Huang, Xiaojian Zhao. 2018. Sunk Cost as a Self-Management Device. *Management Science* . [Crossref]
2. Carsten Herrmann-Pillath. 2018. Beyond dualities in behavioural economics: what can G. H. Mead's conceptions of self and reflexivity contribute to the current debate?. *Journal of Economic Methodology* **22**, 1-15. [Crossref]
3. Milan Zafirovski. 2018. Rational Choice Theory or Pretense? The Claims, Equivalences, and Analogies of the "Economic Approach to Human Behavior". *Sociological Spectrum* **38**:3, 194-222. [Crossref]
4. Maël Lebreton, Shari Langdon, Matthijs J. Slieker, Jip S. Nooitgedacht, Anna E. Goudriaan, Damiaan Denys, Ruth J. van Holst, Judy Luigjes. 2018. Two sides of the same coin: Monetary incentives concurrently improve and bias confidence judgments. *Science Advances* **4**:5, eaaq0668. [Crossref]
5. Charles F. Manski. 2018. Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise. *NBER Macroeconomics Annual* **32**:1, 411-471. [Crossref]
6. Bruno Larue. 2018. Economic Integration Reconsidered*. *Canadian Journal of Agricultural Economics/ Revue canadienne d'agroeconomie* **66**:1, 5-25. [Crossref]
7. Scott M Swinton. 2018. Why Should I Believe Your Applied Economics?. *American Journal of Agricultural Economics* **100**:2, 381-391. [Crossref]
8. George Loewenstein, Andras Molnar. 2018. The renaissance of belief-based utility in economics. *Nature Human Behaviour* **2**:3, 166-167. [Crossref]
9. Benjamin Handel, Joshua Schwartzstein. 2018. Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care?. *Journal of Economic Perspectives* **32**:1, 155-178. [Abstract] [View PDF article] [PDF with links]
10. Raj Aggarwal, John W. Goodell. 2018. Sovereign wealth fund governance and national culture. *International Business Review* **27**:1, 78-92. [Crossref]
11. David Wuepper, Travis J. Lybbert. 2017. Perceived Self-Efficacy, Poverty, and Economic Development. *Annual Review of Resource Economics* **9**:1, 383-404. [Crossref]
12. Richard G. Lipsey, Yew-Kwang Ng. 2017. Concluding Comments to the Debate. *Pacific Economic Review* **22**:2, 213-228. [Crossref]
13. Jean-François Bonnefon. Delusions 149-175. [Crossref]