

Economic Models of Optimism: What Does the Evidence Say?*

Modelos económicos del optimismo: ¿qué muestra la evidencia empírica?

Adrián Caballero
Universidad Autónoma de Madrid

Raúl López-Pérez
Instituto de Políticas y Bienes Públicos, CSIC**

Abstract

Recent economic models define a (non-Bayesian) optimist as someone whose beliefs and expectations are typically “too rosy”, in the sense that her priors and the evidence objectively available statistically warrant a more negative outlook. This paper reviews the existing empirical literature on optimism to assess the empirical relevance of those models. While there exists abundant and compelling evidence in favor of motivated inference, i.e., that preferences shape beliefs, the support for the most specific predictions of the economic models seems mixed, if not negative. We discuss open questions and opportunities for future research.

Keywords: *biases, inference, motivated beliefs, optimism, overconfidence, self-serving bias.*

JEL Classification: *D01, D80, D83, D84, D91.*

Resumen

En recientes modelos económicos, un optimista (no bayesiano) es aquel agente cuyas creencias y expectativas están frecuentemente «tintadas de rosa», dado que sus a priori y la evidencia objetivamente disponible implicarían una predicción estadística menos positiva. Este artículo repasa la literatura empírica sobre optimismo con vistas a evaluar la relevancia empírica de tales modelos. Si bien hay evidencia abundante y convincente de que la inferencia motivada existe, esto es, que las preferencias moldean las creencias, las predicciones más específicas de los modelos parecen tener apoyo limitado, si no nulo. Concluimos con una discusión de algunas cuestiones por resolver, así como de oportunidades de cara a futuras investigaciones.

Palabras clave: *sesgos, inferencia, creencias motivadas, optimismo, exceso de confianza, sesgo por interés personal.*

Clasificación JEL: *D01, D80, D83, D84, D91.*

* The authors gratefully acknowledge financial support from the Spanish Ministry of Economics, Industry and Competitiveness through the research project ECO2017-82449-P, and helpful research assistance by Roberto Lucas and Sergio Rubio.

** Institute of Public Goods and Policies (IPP), Spanish National Research Council (CSIC).

1. Introduction

People often have beliefs or expectations about many things: their material future, their abilities or personal characteristics, how well they perform relative to others, the future price of some asset, the world outlook, social causality, the lethality of some virus, etc. In this respect, abundant evidence from Psychology and Economics hints the existence of a positivity bias in people's inferences, a bias also called motivated inference, optimism, self-deception, wishful thinking, or asymmetric updating (Wicklund & Brehm, 1976; Kunda, 1990; Baumeister, 1998; Bénabou & Tirole, 2016; Epley & Gilovich, 2016). When people have some motivation to hold a certain belief, that is, updating often proceeds as if the good, confirmatory evidence received a relatively larger weight than the bad, negative news. This is at odds with Bayesian learning, e.g., Samuelson, 2004, the standard approach to inference in rational choice theory, which requires not to treat a piece of information differently conditional on whether we like it or not.¹

Understanding how and under what circumstances people infer in a motivated manner is an obviously important question for policy reasons. In effect, this bias has a potential negative side: many decisions like buying some good, exerting effort to get some future reward, investing on some asset, saving, or taking care of our health can be made sub-optimally if people have inflated beliefs or expectations. Kindleberger and Aliber (1978), for example, argue how irrationally optimistic expectations lead to overinvestment during economic booms, and Shiller (2000) includes optimism as one of the biases that can boost the confidence of investors, leading to large increases in asset prices and favoring the formation of financial bubbles. For another example, the survey by Case and Shiller (2003) to new homeowners revealed very optimistic expectations regarding future housing prices. In fact, 90 % of the surveyed believed that housing prices in their cities would keep increasing for the next 10 years, at an average estimated rate of between 9 and 15 %, depending on the city. An excess of optimism can also have detrimental consequences for workers. Using data from the American labor market, Spinnewijn (2015) shows that more than 80 % of the surveyed job seekers underestimated the time it would take to find a new job. Consequently, the job seekers devote little effort to search for a new job, save little for unemployment and consume these savings too quickly while they are unemployed. Regarding health, the study by Oster et al. (2013) with individuals at risk for Huntington disease, a degenerative neurological disorder associated to

¹ To prevent confusion, we stress that optimism is not the only bias which psychologists and other researchers have documented. Specifically, people may process information in a statistically "incorrect" manner for non-motivational reasons, e.g., not attending to some signal or forgetting it due to the mere effect of time. The literature about these "cold" biases is inextricably associated to the work by Daniel Kahneman and Amos Tversky, who developed the idea, introduced by Herbert Simon, that humans reason and solve problems by means of heuristics, i.e., intuitive rules –Simon, 1956; Tversky and Kahneman (1973, 1974); Kahneman and Tversky 2000. These rules, examples of which include availability and representativeness, are relatively efficient because they are easy to apply, but at the same time lead to systematic mistakes in our inferences. Benjamin (2019) extensively surveys the evidence on "cold" inference.

a genetic alteration, reveal that many of them underestimate their true risk, reject being tested and behave as if they do not have the disease, even when diagnosed people behave quite differently regarding decisions. For example, individuals that are certain that they have the genetic alteration are more likely to get married, to retire and to change their recreational activities and financial decisions than those who know not to carry the alteration. Interestingly, the behavior of the untested individuals mimics the behavior of the latter.²

A literature in Economics has proposed different models of motivated inference –e.g., Akerlof and Dickens (1982), Rabin (1994), Bénabou and Tirole (2002), Brunnermeier and Parker (2005), Möbius et al. (2014). In a nutshell, these models of non-Bayesian optimism (NBO henceforth) propose that motivated inference presents analogies with rational choice: people tend to hold the beliefs that maximize their utility, i.e., the beliefs they would hold if they could freely choose them, taking into account their costs and benefits. The NBO models differ mainly in the constraints that they posit people face when “choosing” their beliefs. First, papers like Akerlof and Dickens (1982) and Brunnermeier and Parker (2005) assume no restrictions: the wish begets the belief. Specifically, the optimal expectations conveniently balance the benefits from keeping pleasant beliefs and the advantages of belief accuracy in terms of better decision making. A second group of models posit the existence of cognitive costs of holding inaccurate beliefs as well as costs associated to information avoidance, e.g., Rabin (1994) and Bénabou and Tirole (2002). For example, Rabin (1994, p. 180) argues that departing from the “natural, intellectually honest set of beliefs” may require a lot of intellectual effort. It could imply also some opportunity costs in terms of refraining from pleasant activities in order to avoid challenging information. In addition, Bénabou and Tirole (2002) and Bénabou (2015) introduce sophisticated agents who somehow anticipate the existence of a positivity bias in their inferences and can hence limit the extent of such bias.

Since economists and psychologists have gathered a substantial amount of evidence on motivated inference, a natural question arises: how well do the NBO theories explain that evidence? An exploration of the limits of the models is essential to eventually improve them, if necessary, and to better understand the phenomenon. With this in mind, this paper presents several predictions of the NBO models and compares them with relevant evidence, incidentally offering a (inevitably partial) review of the empirical literature. In contrast with previous surveys on motivated inference, therefore, the theoretical predictions organize the review, which leads to the

² While the prior discussion stresses the downsides of optimism, the bias seems to have a positive side as well. For instance, convenient ego-beliefs or beliefs about the future can bolster our self-esteem, confidence, and hope (think of religion), hence giving us peace of mind and serenity. In addition, Festinger (1957) proposes that choosing an option that is good in some respect but bad in another one triggers a painful cognitive dissonance. This can be prevented if beliefs are tuned accordingly, e.g., believing that the payoff-maximizing, self-interested choice is also the fair one. Consider also a third, more indirect example: Sometimes having strong beliefs on some issue motivates us to be bolder, taking more risks or undertaking difficult tasks; determination can be helpful as well if we want to convince others to join some collective project (e.g., von Hippel & Trivers, 2011).

analysis in detail of issues like, say, the interaction between accuracy and monetary incentives. As we will argue, NBO models (as any simplification of reality) face some issues, and the key one seems to be an inaccurate account of the restrictions or limits of motivated inference. In effect, the models predict that having an interest in believing that some event is true is sufficient to inflate such belief, provided that it is not “too” inaccurate (Rabin, 1994) or that the person is sufficiently naïve (Bénabou & Tirole, 2002). This contrasts with evidence showing that, as we will see, people are sometimes surprisingly accurate even if they have an interest in holding some biased belief, and in some cases hold non-Bayesian pessimistic beliefs, i.e., assign a “too” low probability to the most desirable events. In summary, economic models have mixed support, and their predictions are quite often hardly supported or even challenged by the experimental evidence.

The rest of the paper is organized as follows. The next section presents for illustrative purposes a toy model of NBO and discusses its empirical relevance. Section 3 concludes by mentioning open questions and potential future venues of research.

2. NBO: a toy model and a review of the evidence

The goal of this section is to evaluate the empirical performance of the NBO models, taking into account some of the literature on motivated inference. To organize the discussion, we first introduce a toy model of NBO, building on the theoretical framework from Akerlof and Dickens (1982) and Brunnermeier and Parker (2005). From such model we derive a series of implications. We then analyze whether the existing evidence supports each implication, based on our reading of the literature.

2.1. A simple framework

Let $S = \{\text{blue}, \text{red}\}$ denote the state space, where any state s fully specifies all relevant features of the decider’s environment. We describe the uncertainty involved by introducing the *objective* probability π that the state is red. The decider’s objective is to maximize a weighted sum of the expected utility of her material payoff plus the utility of holding the beliefs characterized by $\hat{\pi} \in [0, 1]$, which is interpreted as the decider’s subjective degree of confidence that the red state is true. Implicitly, the red state is a positive one in which, say, the decider (i) is successful in some task, (ii) has some “good” personal quality, e.g., physical attractiveness, intelligence, honesty, etc., or (iii) gets some material reward. We define the decider’s utility as:

$$U(a, s, \hat{\pi}) = u(a, s) + v(\hat{\pi}) \quad [1]$$

where a is the decider’s action. The first term represents the utility associated to the

material payoff, which is possibly determined by the state and the specific action chosen by the decider. The second term captures the utility derived directly from holding beliefs $\hat{\pi}$. Without loss of generality, we assume that $dv/(d\hat{\pi}) > 0$, so that believing that the actual state is red provides some utility. In Akerlof and Dickens (1982), for instance, a belief that some job is not hazardous can be relieving, leading to less stress and fear. Alternatively, an individual may feel comfortable when thinking that she will not suffer a health problem in the future or that the value of her assets will always increase.

In line with Akerlof and Dickens (1982), there are two periods. In the first one, the decider is aware of the objective probability π and “chooses” her beliefs $\hat{\pi}$. In the second period, she chooses an action based on her beliefs $\hat{\pi}$, and gets some pecuniary and non-pecuniary benefits, possibly including some utility $v(\hat{\pi})$ from beliefs $\hat{\pi}$. In general, we say that the decider is biased if $\hat{\pi} \neq \pi$, inflating (deflating) her beliefs if $\hat{\pi} > \pi$ ($\hat{\pi} < \pi$); we also employ the term “optimism” (“pessimism”) as a synonym of inflation (deflation). We assume that the decider in the second period ignores the possibility that she may have strategically chosen inflated beliefs in the first period. Thus, in the second period, her optimal action is:

$$a^*(\hat{\pi}) = \arg \max_a \hat{\pi}u(a, \text{red}) + (1 - \hat{\pi})u(a, \text{blue}) + v(\hat{\pi}) \quad [2]$$

Importantly, the decider anticipates in period 1 that her choice of beliefs $\hat{\pi}$ will possibly affect her action in period 2, perhaps leading to suboptimal behavior from the point of view of an accurate observer. In Akerlof and Dickens (1982), for instance, a worker that underestimates the risk associated to some job may feel less anxious, but at the same time he is more likely to suffer an accident. In choosing $\hat{\pi}$, therefore, the decider will look for a compromise, trading off the utility from $\hat{\pi}$ with the cost imposed by future wrong actions. The decider’s optimal belief is thus:

$$\hat{\pi}^*(\pi) = \arg \max_{\hat{\pi}} \pi u(a^*(\hat{\pi}), \text{red}) + (1 - \pi)u(a^*(\hat{\pi}), \text{blue}) + v(\hat{\pi}) \quad [3]$$

To repeat, note well that $\hat{\pi}$ determines the decider’s expected utility at period 1 in two ways. First, directly, since keeping beliefs $\hat{\pi}$ provides utility $v(\hat{\pi})$. Second, indirectly, since beliefs $\hat{\pi}$ may affect the decider’s future action. From this simple framework, let consider three possible scenarios. To start, if $v(\hat{\pi}) = 0$ for any $\hat{\pi}$, i.e. if beliefs do not provide utility by themselves, $\hat{\pi}^*$ is such that:

$$a^*(\hat{\pi}^*) = \arg \max_a \pi u(a, \text{red}) + (1 - \pi)u(a, \text{blue}) \quad [4]$$

Given our previous definition of a^* , it turns out that $\hat{\pi}^* = \pi$. That is, if beliefs are not a direct source of utility, the decider has no incentive to keep biased beliefs. More generally, this result applies if $v(\hat{\pi})$ is constant for any $\hat{\pi}$. In the second and third remaining scenarios, we assume otherwise. The second scenario is the simplest one to analyze and occurs when there are no material costs from choosing wrong,

inaccurate beliefs. More precisely, let $\hat{\pi}^b = \arg \max_{\hat{\pi}} v(\hat{\pi})$ denote the beliefs giving maximal “direct utility” within interval $[0, 1]$, and $Eu(a^*(\hat{\pi}))$ denote the expected utility derived from the *material* payoff, given the optimal action a^* for beliefs $\hat{\pi}$. If $Eu(a^*(\hat{\pi}^b)) = Eu(a^*(\pi))$, utility maximization can be conceived as a relatively simple two-step, backwards induction procedure: in period 1, the decider opts for the “most pleasing” beliefs $\hat{\pi}^b$, anticipating that she will later choose act $a^*(\hat{\pi}^b)$ in period 2. Since beliefs do not compromise future material payoffs, that is, she just opts for those beliefs that maximize $v(\hat{\pi})$.³ Particular sub-cases within this second scenario include (i) a^* maximizes $u(a, s)$ for any s , or (ii) $u(a_i, s) = u(a_j, s)$ for any pair i, j .

The third scenario appears when some beliefs provide more “direct utility” $v(\hat{\pi})$ than others but the decider can suffer a material loss for holding inaccurate beliefs. Here the analysis of NBO becomes more complex. Given some additional assumptions, however, the toy model still predicts some degree of optimistic bias regarding π . Recall first that $dv/d\hat{\pi} > 0$, as we assumed before. Assuming in addition that $Eu(a^*(\hat{\pi}))$ is continuous and twice differentiable and that it has a unique maximum, the problem’s first order condition for the decider in the first period is:

$$\frac{dEu(a^*(\hat{\pi}))}{d\hat{\pi}} + \frac{dv(\hat{\pi})}{d\hat{\pi}} = 0 \quad [5]$$

Recall from the first scenario that the first term is maximized for $\hat{\pi} = \pi$. For those beliefs, it turns hence out that $dEu(a^*(\hat{\pi}))/d\hat{\pi} = 0$ and that $d^2Eu(a^*(\hat{\pi}))/d\hat{\pi}^2 < 0$. Since the second term in [5] is strictly positive, optimal beliefs $\hat{\pi}^*$ are such that $dEu(a^*(\hat{\pi}))/d\hat{\pi} < 0$, which implies that $\hat{\pi}^* > \pi$ as long as $\pi < 1$. In other words, the decider finds optimal to form optimistic beliefs, inflating $\hat{\pi}$ up to the point in which the marginal “direct utility” of those beliefs is cancelled out by the marginal decrease in the expected utility from the material payoff.

2.2. Predictions and evidence

The first prediction of NBO is direct: a person’s beliefs should co-move with her preferences. If the decider maximizes utility by believing that some state s is the actual one, that is, she will tend to believe it. In contrast, she will be more critical towards that idea if it does not give utility.⁴

Prediction 1 (PI): Deciders inflate their beliefs. Specifically, the difference between the subjective and the objective probabilities is correlated with the decider’s preferences over the beliefs about the state space.

³ The discussion presumes for simplicity that both $\hat{\pi}^b$ and $a^*(\hat{\pi}^b)$ are unique. A multiplicity of equilibria exists otherwise.

⁴ This can be contrasted with a perhaps more “rational” account in which people believe S because they have (proper) reasons for that, e.g., evidence, and not because it is functional.

Discussion and evidence: For a well-known illustration of *PI*, consider a scenario where the decider's utility depends on her expected material payoff and some direct utility from beliefs. More precisely, π represents here the belief that the decider is a fair and honest person; intuitively she gets "ego-utility" from believing that she is such a person. In this type of scenarios, it is often safe to assume that holding inaccurate beliefs is not risky, that is, entails no material loss. As we have indicated before, the NBO maximization problem is hence straightforward: When people have to make a normative choice, intuitively, they first conceive the act that best serves their self-interest and *then* pick the beliefs about what is normative to make that choice seem correct or justifiable, inferring as a result that they are fair-minded people –i.e., the "good" state of the world. More specifically, consider a scenario with two acts a_{blue} and a_{red} and two states, blue and red, i.e., such that a_{blue} (a_{red}) is the right, normative act in the blue (red) state. Let π denote the objective probability of the red state. If choice a_{red} maximizes material utility, say, the prediction is that $\hat{\pi}^* = 1$, so that the decider can earn as much as possible and feel no guilt for acting in an unfair manner. If $1 > \pi$, e.g., if an impartial observer is not totally convinced that is the normative choice, a self-serving bias in fairness judgments is predicted.⁵

Apparently in line with *PI*, many studies on self-serving biases report behavior that can be rationalized by convenient beliefs about the fairness of the acts chosen.⁶ Since fairness judgments can often be conditioned on a myriad of reasons, the studies reflect this richness. In some cases, for instance, people seem to have self-serving beliefs about the applicable moral rules. In Shu and Gino (2012), participants read an honor code and were then asked to conduct a task and report their performance –they were allowed to cheat by over-reporting, thus getting a higher payoff. Those participants who cheated tended to forget the moral rules from the honor code even when monetary incentives were provided. In contrast, they were equally likely to remember information unrelated with moral. In the control condition in Babcock et al. (1995), in turn, subjects are assigned the role of plaintiff or defendant and matched in pairs. All of them work with the same real case, in which the plaintiff demands a compensation of \$100,000 after a traffic accident involving both parties. They are given real information regarding testimonies, police reports, and maps and informed that a real judge has been given the case materials to reach a decision about the compensation for the plaintiff. Each negotiator is then asked an incentivized, private estimate of the judge's decision, as well as an evaluation of what a fair settlement would be according to her own view. Afterwards, the negotiators bargain during 30 minutes over the amount that the defendant would pay to the plaintiff. If they do not reach an agreement over that time span, the plaintiff is awarded the compensation determined by the judge (for an amount unknown a priori by the

⁵ These occur when, in the words of Babcock and Loewenstein (1997, p. 110), people "...conflate what is fair with what benefits oneself". Or as defined by Dahl and Ransom (1999, p. 703): "A self-serving bias occurs when individuals subconsciously alter their fundamental views about what is fair or right in a way that benefits their interest."

⁶ In these studies, subjects' beliefs are often not elicited, though.

negotiators). In one treatment, subjects are informed of their roles only after reading the materials and estimating the judge's and fair settlements (but prior to negotiation). In line with *PI*, subjects in this veil-of-ignorance treatment are significantly more likely to settle and do so in less time. Further, they are significantly more likely to agree on the judge's and fair settlements and less likely to differ in the importance they attach to self-serving arguments.

In other occasions, a self-serving bias is manifested in the agents' selection of comparison groups, a point often relevant in fairness judgments. In a field study examining actual teacher contract negotiations in 500 school districts in Pennsylvania, indeed, Babcock et al. (1996) observe that unions tend to choose comparables (in other school districts) with higher salaries than those that school boards select. This has consequences, as the higher the difference between the proposed comparables, the higher the strike rate and hence the negotiation impasses.

In some studies, people apparently choose convenient beliefs about how deserving another person is. Di Tella et al. (2015) conduct a modified version of the dictator game, in which the receiver secretly decides the monetary value of the tokens allocated by the decider and receives an additional payoff if the lower value is chosen. Simultaneously, the decider chooses the allocation of 20 tokens, although the maximum number of tokens that deciders can keep for themselves varies among players (this information is known by the decider only). The main result was that those deciders who were allowed to keep a larger proportion of the tokens not only behaved more selfishly, but they were also more likely to believe that the receiver had behaved selfishly too, even when realistic beliefs were incentivized in some treatments. That is, deciders apparently form 'excusing' beliefs about receivers' behavior.⁷ In Barkan et al. (2012), for another example in this vein, participants were presented a hypothetical situation in which they work in the HR department of a firm and must decide whether to hire a candidate whose ethical behavior is depicted as questionable but advantageous for the firm. In one of the treatments, participants were asked in a previous task to describe an unethical thing they had done. Those participants were significantly less likely to hire the candidate and judged him as potentially less loyal to the firm and less honest than the participants in the control treatment (which had not been asked previously to recall their own past misconduct). It seems therefore that people recalling their past misconduct tend to reduce ethical dissonance by judging others' moral behavior even more harshly.

The previous studies strongly suggest that self-serving biases in fairness judgments exist; see also Gino et al. (2016) for a review of additional evidence. It must be noted however that not all of the existing evidence in this realm is in line with *PI*. For instance, Dahl and Ransom (1999) find very limited evidence of a bias among Mormons in their own determination of what constitutes income for the purpose of tithing –e.g., having received a sizable gift or inheritance does not affect individuals' views of whether gifts and inheritances should be tithed.

⁷ Yet Ging-Jehli et al. (2019) find no evidence of such kind of "strategic cynicism".

Honesty is not the only trait of intrinsic importance to the individual. Self-perceptions of one's intelligence and beauty, for instance, seem to affect ego-utility as well. In this sense, Moore and Healy (2008) distinguish three related types of overconfidence that have been commonly confounded. Two of them are particularly relevant in our context.⁸ In first place, NBO models predict *overestimation* of one's abilities and inflated self-perceptions in these regards, which fits well with the results of different calibration studies (Lichtenstein & Fischhoff, 1977; Budescu et al., 1997, Brenner et al., 1996). The structure of these studies is usually quite similar: the participants are presented a series of multiple-choice questions. In addition to choosing the right answer, they must estimate the probability that their answer is correct. If a participant's mean estimation is p , overestimation occurs when the percentage of actually correct answers is lower than $100 \cdot p$. Interestingly, the evidence suggests that this phenomenon is not universal and that it is conditioned to a large extent by the context (Gigerenzer et al., 1991). For instance, the degree of overestimation seems to increase with the difficulty of the task, which has been labeled the "hard-easy effect" –e.g., Brenner, 2003; Larrick et al., 2007.

As a second type of overconfidence in line with NBO, we have the phenomenon generally known as *over-placement* or *better-than-average effect*: when evaluating their own capacities and abilities in relative terms to others, a growing body of work from social psychology and economics suggests that people are often biased –consult Alicke and Govorun (2005) for a review. In Svenson (1981), for example, 93 % of U.S. students believe that they are both safer and better drivers than the median one. In the same vein, Cross (1977) find that more than 65 % of college teachers rank themselves in the top 25 % when asked about their teaching abilities relative to other professors.⁹ Note yet that the evidence is not conclusive; for instance, the experiment by Clark and Friesen (2009), in which the participants face real effort tasks and must infer their absolute and relative performance, finds little evidence of either over-placement or over-estimation. In this vein, one feature that seems to be at odds with NBO is that over-placement seems particularly true for skills or tasks that are relatively easy to master. In this line, a survey conducted by Kruger (1999) reported significant above-average effects in skills like driving, riding a bicycle or saving money, but that people systematically underestimated their ability relative to others in more complex tasks like playing chess or programming a computer. Finally, two interesting aspects of over-placement are that (i) people tend to see themselves as less biased than others (Pronin et al., 2004) and that (ii) it makes people more effective in persuading or deceiving others. In line with (ii), subjects in Schwarzmann and van der Weele (2019) first perform an intelligence test. Then

⁸ Moore and Healy (2008) also discuss over-precision, that is, statistically unwarranted overconfidence on the truthfulness of one's inferences, typically exemplified by "too" narrow confidence intervals.

⁹ As Benoit and Dubra (2011) argue, the evidence from some earlier social psychology studies on over-placement is not unequivocally at odds with Bayesian updating. Nevertheless, more recent studies in Economics like Eil and Rao (2011) and Möbius et al. (2014) eliminate these confounds; we review them below.

a random half of them are informed that they can earn a prize by convincing others of their superior performance in the test. It turns out that the privately elicited and incentivized beliefs of these subjects are significantly more confident than the beliefs of subjects in the control condition. Further, the more confident participants state a higher confidence during the face-to-face persuasion stage, and are significantly more likely to be rated as successful by their evaluators.

We turn now to another body of literature in line with NBO, this one suggesting that people tend to be optimistic about their future. In the survey conducted by Weinstein (1980), for example, people underestimate the chances of getting divorced, losing their jobs or suffering different health problems. In contrast, they overestimate the probability of living past eighty, owning a house or finding a good job. Far from anecdotal, the estimates suggest that about 80% of the population show this kind of bias; Sharot, 2011. In some studies, for another example, people have inflated prospects about their material prospects. In the experiment in Mayraz (2011), subjects are shown a chart of historical wheat prices and have to predict afterwards the price at some future time point. This is iterated 12 periods, one of which is randomly selected afterwards for compensation; the size of the bonus depends on the accuracy of the corresponding prediction. Additionally, any subject gets a payoff that increases (decreases) with the future price of wheat if her randomly-selected role is “Farmer” (“Baker”). If subjects get sufficient utility from being optimistic and truthfully report their predictions, NBO predicts that the average farmer should make significantly higher predictions than the average baker, as it is indeed observed. Bayesian learning, in contrast, predicts no difference, given random role allocation. In turn, Forsythe, Rietz and Ross (1999) conduct an experiment in which the participants trade with two assets, namely Blue and Green certificates. At the end of each period, Blue certificates liquidate for some monetary value with probability 0.3 (Blue state), while Green certificates expire worthless. The opposite occurs with probability 0.7 (Green state). Each sequence of three trading periods has the following structure: in the first period, no additional prize is provided. In the second (third) period, all participants get \$2 if the Blue (Green) state occurs. This prize is independent of the decision made by the traders. On average, the authors report that the price of Blue (Green) certificates increased when an additional prize was associated to the Blue (Green) state, suggesting an inflated probabilistic belief that the most favorable state occurs. In contrast, average prices tended to decrease in the subsequent periods with no state-contingent prize.

We note yet that negative evidence on optimism about financial prospects is offered by Heger and Papageorge (2018), where subjects sequentially completed several tasks, each one containing a number of questions –at the end of the session, one task was randomly selected for payment. In the Baseline task, subjects face several urns, each one containing a different proportion of white and black balls, which is common knowledge. Subjects report their probabilistic belief that a draw from each urn will be white, being paid for accuracy according to a quadratic scoring rule. To explore over-estimation of the probability of high-payoff outcomes, participants in the Payment task receive a side payment if a white ball is drawn from the urn.

Clearly, optimism implies higher average reported beliefs in this task, where white becomes payoff-favorable. Nevertheless, the data does not support this prediction. On the other hand, the authors also explore a Performance task in which white balls are not payoff-favorable, but ego-related. Specifically, each subject answers an IQ question and, if her answer is correct, one white ball is added to the urn. In this task, average beliefs are upwardly biased, in line with the evidence cited above about over-estimation. Further, the authors find that subjects who over/under-estimate the probability of white in the Payment task tend also to over/under-estimate their own performance in the Performance task. ■

Let us now turn on more specific predictions by the NBO models. In this vein, an implicit or explicit mechanism behind most models of NBO is asymmetric updating: Signal observations are over-weighted or under-weighted depending on the decider's goals, i.e., the target, optimal beliefs. While Akerlof and Dickens (1982) and Brunnermeier and Parker (2005) do not formalize how agents learn in their models (as neither does our toy model), Möbius et al. (2014) does analyze this point formally, building on the framework from Brunnermeier and Parker (2005). Möbius et al. assume that agents optimally choose the weight they give to bad and good news, predicting that (i) agents will be conservative, i.e., they will respond less to the evidence than a Bayesian, but at the same time (ii) they will be more conservative in response to bad news than good news. Benabou and Tirole (2002) propose an NBO model where agents display motivated recall, so that they recall better the evidence that supports their preferred beliefs. In their model, specifically, an individual that prefers to believe that she is good at certain task may repress or forget those pieces of information that question her ability. This manipulation is limited to some extent by cognitive costs. Further, their model allows for some sophistication in the sense that the individual can be aware up to a certain degree that her recalled sample is likely to be biased, which consequently may limit –but not eliminate– the effects of motivated recall.

Prediction 2 (P2; asymmetric updating/inference): Individuals treat as more informative the signals in line with their preferred beliefs than those that challenge them. Thus, individuals process “good” and “bad” news asymmetrically in order to reinforce their favorite beliefs.

Discussion and evidence: Neuroscientist and psychologists have recently gathered some supportive evidence for asymmetric updating. In Sharot et al. (2011), participants (N = 19) are presented some adverse event, such as being diagnosed with Alzheimer's disease or being robbed, and have 6 seconds to estimate their chances of facing that event in the future, not being paid for accuracy. Afterwards, subjects are shown in the screen during 2 seconds the actual frequency with which that event happens among individuals living in the same socio-cultural environment as them. This procedure is repeated for 80 events, randomly ordered. To assess whether participants use the information on the actual empirical frequency to update their predictions, the authors subsequently ask them to provide estimates of their likelihood of encountering the events considered before. They find evidence for

asymmetric updating in favor of good news.¹⁰ Interestingly, Sharot et al. (2011) obtain fMRI (functional Magnetic Resonance Imaging) data while subjects proceed with the experiment. This data shows that activity in a region of the frontal cortex (right IFG) identified as sensitive to negative estimation errors predicts the extent to which participants selectively update their beliefs, (relatively) dismissing the information conveyed by the empirical frequency when it was bad news.

Economists have explored as well the relevance of asymmetric updating. An interesting aspect of the experiments cited below is that, in contrast to our knowledge to the psychological studies, both signals and priors are strictly controlled by the experimenter. This makes the Bayesian benchmark readily computable, something unfeasible in the experiment just cited above by Sharot et al., 2011, for instance. Until recently, this was not very common in the literature, particularly when dealing with beliefs about aspects that are of every day concern –health, intelligence, beauty, academic performance, driving ability, etc. While these prior studies were tremendously suggestive, they failed to clarify the existence of biases, as the knowledge that subjects bring to the studies was hardly verifiable.¹¹

Recent studies also allow pinpointing some mechanisms through which these biases develop. In this regard, we must note that experimental attempts in Economics to test $P2$, i.e., asymmetric updating, have concluded with mixed results. While some works find a significant overweight of good news relative to bad news (Eil & Rao, 2011; Möbius et al., 2014; Charness & Dave, 2017), other scholars find no evidence of asymmetric updating (Grossman & Owens, 2012; Gotthard-Real, 2017; Buser et al., 2018; Barron, 2019) or even report the opposite phenomenon: that individuals overweight negative information relative to positive one (Ertac, 2011; Kuhnen, 2015; Coutts, 2019).

More specifically, Eil and Rao (2011) explore in two separate treatments how subjects update their beliefs about their ranking within a group of 10 people according to intelligence, as measured by score on an IQ test, or physical attractiveness, as rated by 5 subjects of the opposite sex after an “speed-dating” exercise. Subjects participated as well in a randomly-ordered control, where they updated their beliefs about their position in a “neutral” ranking, determined by a randomly assigned integer ranging from one to ten. In all treatments, subjects first revealed their prior probability distribution over the 10 possible ranks.¹² Afterwards, each subject received a series of three objective signals, i.e., whether he/she was ranked higher or lower than a randomly selected comparison subject, and had his/her (posterior) probabilistic beliefs elicited after each one. Accuracy was rewarded according to a

¹⁰ See Wiswall and Zafar (2015) for similar results regarding college students’ beliefs about their future earnings.

¹¹ For example, Kunda (1987) reports that, when given some evidence that caffeine drinking leads to negative health effects for women, female caffeine drinkers are less likely to rate it as valid than male ones. Yet this is not a clear signal that female drinkers are biased, as they could have gathered some disconfirming evidence in the course of their lives.

¹² These priors were non-incentivized. As the authors recognize, an issue here is that subjects might conceal their true subjective priors. For instance, they could reveal low probabilities for the top ranks, perhaps out of some sense of modesty. A Bayesian who acted in this manner would later under-respond to any negative signal.

quadratic scoring rule; consequently, a subject's expected payoff is maximized if she honestly reports her beliefs (assuming she is risk-neutral and understands the implications of the procedure). The authors find evidence for asymmetric updating in the IQ and beauty treatments (although the effect is significant only in the latter one), but not in the control. Further, agents' posteriors in the ego-relevant treatments are less predictable and less sensitive to signal strength after receiving negative feedback.

Ertac (2011) considers a relatively similar design with a coarser ranking, as subjects must report beliefs over which third of the distribution they occupy. Each subject observes one signal indicating whether she is in the highest rank (the bottom one in some sessions). Beliefs are elicited both *ex ante* and *ex post*, and compensated by means of a quadratic scoring rule. She finds asymmetric updating but in favor of the *unfavorable* feedback, thus leading to overly pessimistic beliefs. The main exception are those individuals who have relatively optimistic priors. Further, subjects are relatively less pessimistic about their relative performance in a math quiz than in a verbal task that they found generally harder than the former. Hence, these findings are in line with the evidence cited above that people are more likely to be overconfident in domains that they perceive to be easy (e.g. Kruger, 1999; Moore & Cain, 2007).

In the web-based study by Möbius et al. (2014), in turn, subjects are ranked according to their score in an IQ test. Participants report beliefs about which half of the distribution they occupy, measured both before and after the test and then again after receiving each of four noisy signals. At each update, subjects are reminded of the full history of signals. The belief elicitation method used was the so-called crossover method, which is incentive compatible even for risk-averse agents. The authors report evidence that updating is asymmetric but also conservative, in other words, people respond less to feedback (of any sign) than Bayes' rule prescribes. These findings are compatible with the NBO theory that they propose. Further, the biases are substantially less pronounced in a follow-up where the conditions are as closely as possible to those in the baseline, except that ego is not at stake, as subjects updated their beliefs about the performance of a "robot" rather than their own performance.¹³

A distinctive feature of Grossman and Owens (2012) is that they explore learning about *absolute* performance. Participants take an incentivized ten-question quiz on logic and reasoning and estimate the entire probability distribution of the 11 potential scores, i.e., 0 to 10, before and after taking the quiz, and after receiving one noisy signal. Participants were not given feedback on prior estimates. One of the three belief elicitations, randomly selected, was compensated with a quadratic scoring rule.

¹³ Eil and Rao (2011) and Möbius (2014) check for asymmetric updating but also for the possibility of a confirmation bias, i.e., the tendency to place more weight on signals that confirm one's prior and underweight disconfirming signals; see Rabin and Schrag (1999) for a review. Neither of these studies finds evidence for such phenomenon. In Eil and Rao (2011), for example, subjects with negative priors also exhibited asymmetric updating, i.e., put relatively more weight on the positive, disconfirming signals.

Participants estimate their own score in treatment *Self* and the score of an anonymous participant in a separate condition called *Other*. This makes a difference with the studies cited before, where the controls feature events with objective probabilities, which may be interpreted very differently by subjects. Comparing a subject's expected score with her actual one, the authors find evidence of overconfident prior beliefs, particularly in the *Self* condition. However, non-Bayesian updating in the form of conservatism or asymmetric updating is not observed in the data, except for a minority of subjects. Taking into account the results from Eil and Rao (2011) and Möbius et al. (2014) cited before, the authors conjecture that asymmetric responsiveness to feedback might be more likely in beliefs about relative performance.

In the studies previously considered, the desired states are ego-relevant, e.g., a subject's self-esteem possibly depends on how well he/she performs in the IQ test relative to others, or in absolute terms. This is problematic in that people with different personal characteristics (e.g., intelligence) may update differently, and moreover receive different evidence—for example, a subject with a high IQ score is likely to receive good news, even if the signal is noisy. Under these circumstances, aggregate behavior may exhibit apparent asymmetries that are not present at the individual level; see Barron (2019) for a discussion of this point. In addition, one could wonder whether asymmetric updating extends to other domains. In this regard, several papers consider experiments in which subjects have a preference for some states simply because they give them a higher monetary reward. In Gotthard-Real (2017), there are two states (bad/good), priors are exogenously given and uniform, and a subject gets \$7 if the good state occurs (and zero otherwise). As in Möbius et al. (2014), each subject received four noisy signals, always with feedback, and reported her belief about the probability of the good state after each signal. Further, a randomly selected belief was compensated using a quadratic scoring rule. To diminish any suspicion from the subjects, the random selection of an event was always done mechanically (contrary to other studies). The author finds no evidence that subjects inflate the probability of the good state, relative to a Bayesian.

In Coutts (2019), subjects update their probabilistic beliefs about four different binary events, randomly presented. One refers to the mean temperature on a randomly selected day in New York City, another one is ego-relevant, i.e., relative performance on a math and verbal skills quiz, and the other two are neutral, as they involve rolling dice. In each scenario, a subset of the participants had a 50% chance of receiving \$80 if some target event occurred; hence they had a financial interest in this respect. In each scenario, further, subjects first stated their subjective priors, then received three noisy signals, and reported her belief about the probability of the corresponding good state after each signal—one belief, randomly selected across all scenarios, was compensated using the crossover method, as in Möbius et al. (2014). Coutts (2019) reports that updating is conservative, with many non-updates, and asymmetric. However, it is *negative* signals that are over-weighted, not positive ones. These phenomena are observed both in the valenced and neutral contexts. Any bias, therefore, seems not to be of a motivated nature. Since the analysis further reveals

that, after receiving two signals, subjects overweight a second positive (negative) signal if the first one was positive (negative) signal, the author suggests the potential role of a confirmation bias. As we saw before, however, neither Eil and Rao (2011) nor Möbius et al. (2014) find evidence for such phenomenon.

Barron (2019) considers asymmetric treatments, in which subjects get a (small) bonus payment if one state of the world (out of two) is realized. Priors of the two states are exogenously fixed, subjects observe a sequence of five independent but noisy signals (feedback about prior signals is always provided), and state their beliefs about the probability of the “desired” state upon receipt of each signal (the quadratic scoring rule is used for compensation). Subjects repeat the exercise five times with different priors in each case, chosen in random order from the set $\{1/6, 2/6, 3/6, 4/6, 5/6\}$. Barron (2019) finds no evidence of asymmetric overweighting of good news or alternatively confirmation bias; indeed, average updating behavior is well approximated by Bayes’ rule. This is also true at an individual level –this type of checking is relevant because, if some individuals are optimistic updaters and others pessimistic updaters, one could observe no asymmetry at the aggregate level if both groups were of a similar size and their bias of a similar magnitude.

In Charness and Dave (2017), there are two possible states, equally likely a priori. In one state, a container has seven black balls and three white balls, while these numbers are reversed in the other case. In each of 10 rounds, the actual state is first determined and then participants observe a sequence of six random draws from the container; after each draw, they state the posterior that each state has been realized, with a payoff for accuracy. At the end of each round, the true state is revealed and (in the Strategic condition) each subject is randomly matched with another one to play a 2×2 game; which payoff matrix is in force depends on which state has manifested. Subjects in the “Odd” role receive the same equilibrium payoff in both matrices, while those in the “Even” role receive a higher equilibrium payoff in one of the two matrices. The Even role, therefore, has an interest that one of the states occurs. On the other hand, in the “Non-strategic” treatment, each state is associated to a fixed payoff, with an identical ordinal structure for both roles as in the Strategic condition. The regression results indicate that Odd players learn differently than Even ones, as the latter give less weight to negative signals, i.e., those confirming the unfavorable state. To our knowledge, this seems to be the only economic study offering detailed evidence of asymmetric updating when subjects have a financial interest for some state.

In Buser et al. (2018), participants perform three tasks. The first one is announced to be designed so as to measure the individual’s IQ (Raven’s matrices). In the second, participants must form anagrams of different words. Subjects are informed that the task assesses their verbal ability. In the last task, they observe several matrices filled with numbers; in each one, they must select the two numbers that add up to 10. Participants are told that this third task measures their mathematical ability. In each task, groups of 8 participants are randomly selected. After completing a task, each participant is asked for her (incentivized) belief that she is in the top-performing

half in her group. Then, the participant observes a total of six informative signals about her relative performance, and her beliefs are elicited after each observation. At the aggregate level, Buser et al. (2018) find evidence for conservatism but not for asymmetric updating. Further, individuals tend to be more conservative in tasks that they perceive as more ego-relevant. While an individual's conservatism is correlated across tasks, finally, asymmetric updating was less stable. ■

We move to a slightly different issue. The framework model presented in *Section 2.1* captures in a simple way the reasons why an individual may prefer keeping biased, optimistic beliefs rather than realistic expectations. The idea of asymmetric updating, further, characterizes learning by optimistic agents. However, the question of which specific cognitive mechanism generates the optimistic beliefs remains unanswered. In this sense, Bénabou and Tirole (2012) discuss three non-exclusive alternatives. The first one assumes that, through evolution, a cognitive bias has persisted so that humans systematically overweight favorable information and underweight adverse one. The authors note, however, that this hypothesis is limited as long as people are not always overoptimistic and, in fact, some of them can be quite pessimistic. Second, people may prefer to ignore new information if that could challenge their favored beliefs. Thus, they could attend only those sources of information that they expect to reinforce their preferred beliefs. Finally, selective memory (denial, repression, etc.) could help manage those signals that are unavoidable and that may challenge the individual's preferred beliefs. This latter possibility leads naturally to the next prediction.

Prediction 3 (P3; Bénabou and Tirole, 2012): Asymmetric updating operates through biased recall and selective attention.

Discussion and evidence: there is some suggestive evidence in line with P3, but we believe that it is hardly conclusive. For instance, Thompson and Loewenstein (1992) explore labor negotiations, and find that subjects representing opposite sides later remember, from the same case file, more facts favoring their position than going the other way. The more divergent their recalls, moreover, the longer and costlier is the delay to agreement in the bargaining phase –see also Loewenstein et al., 1993. Recall also the evidence from Shu and Gino (2012) cited in the discussion of P1, where those participants who cheated tended to forget the moral rules from the honor code. In Chew et al. (2019), in turn, subjects (N = 1143) answer in an incentivized manner four questions from a Raven's IQ test. Some months later they are shown the same four questions, plus two which are new but similar, together with all the correct answers, and are asked to recall for each of the six questions whether they (a) answered it correctly, (b) incorrectly, (c) never saw it, or (d) just do not remember. For each question, subjects receive a prize for being accurate, suffer a loss if they recall incorrectly, and get nothing if they cannot remember. Subjects often make errors, which show some systematic patterns. First, the error by which one inaccurately "remembers" having correctly answered one question is significantly more likely than the reverse error, that is, a wrong memory of an incorrect answer. Second, there

is significant evidence of positive delusion. i.e., fabricating a positive event that did not actually happen. In effect, subjects had never seen questions 5 and 6, but more than 56% of them “remembered” answering any of them correctly, versus less than 6% incorrectly. Third, positive amnesia, i.e., forgetting a past negative event, was also very significant. In effect, the probability of not remembering one’s answer, or whether one saw a question, is on average twice as high for those individuals who did a question incorrectly, relative to those who answered it correctly.¹⁴

On the negative side, the above cited study by Sharot et al. (2011) reports that asymmetric updating in their data cannot be explained by differential memory for desirable compared to undesirable information. In effect, after the scanning session, participants had to recall the (previously presented) actual probability of each of the 80 events occurring to an average person in the same socio-cultural environment. The errors thus committed did not depend on whether the actual probability was better or worse than initially expected by the participants, i.e., whether it was bad or good news. In most of the studies discussed in relation to *P2*, finally, subjects receive feedback so that biased recall should play no role. Still, some of them, e.g., Eil and Rao (2011) and Möbius (2014), find a positivity bias. This is perhaps a signal that NBO does not require that subjects “forget” or “misinterpret” signals. ■

For the following *Prediction 4*, consider two scenarios *S1* and *S2*. In *S1*, there is a risk of a substantial material loss for having mistaken beliefs. In *S2*, the expected loss is small or there is no such risk, e.g., in a lab scenario, belief elicitation is incentivized in *S1* but not in *S2*. Economic models of NBO predict that people will infer in a more Bayesian manner in *S1*, so as to reduce mistakes.¹⁵

Prediction 4 (P4): The correlation between beliefs and preferences will be weaker if the expected material loss for holding inaccurate beliefs is high.

Discussion and evidence: several of the papers cited above have systematically analyzed whether the size of the expected loss reduces the degree of NBO. Coutts (2019) runs sessions with different accuracy payments, i.e., low (\$3), medium (\$10), or high (\$20). In addition, participants can either get a nil or high prize (\$80) if some target event *E* occurs. According to *P4*, subjects have no incentive to distort their beliefs about the probability of *E* in case the prize is \$0, provided that they get no other utility from the occurrence of *E*, e.g., if *E* is not ego-relevant. In contrast, distortion should be maximal if the prize is high and the accuracy payment low, i.e., \$3. In spite of this, the author concludes that neither prizes nor accuracy payments alter updating. Mayraz (2011) varies the size of the accuracy bonus from £1 to £5 and reports that

¹⁴ This is highly suggestive, but is it asymmetric updating? Maybe people use the following heuristic: “If the answer sounds intuitive and hence familiar with my preconceptions, I answered it correctly”. Familiarity, not convenience, might explain recall here. On the other hand, is the observed pattern of recall optimal? Over-estimating the rate of correct answers might bolster ego-utility, but why should subjects commit the (costly) mistake of saying that they saw one question that in fact they did not?

¹⁵ Something similar should be expected if individuals get some intrinsic utility from being accurate in *S1*, but not in *S2*. The literature on motivated inference has to our knowledge rarely considered this possibility, though.

the magnitude of the bias does not depend on the scale of the bonus (neither on the size of the prize associated to the desirable event). Similarly, Ertac (2011) studies the effect of rewarding accurate beliefs and finds no significant difference across compensated and non-compensated sessions in terms of the distribution of priors and the absolute value of the bias. If people suffer from NBO, in summary, the cost of such bias does not seem to set limits on its size, at least given the parameterizations that have been considered so far.

More indirect evidence is also at odds with *P4*. In the surveys conducted by Case and Shiller (2003), respondents showed an exaggerated optimism regarding the future evolution of house prices. Also, people tend to be too optimistic regarding their future health (Weinstein, 1987; Caponecchia, 2010). A survey conducted by Weinstein et al. (2005), for instance, find that smokers underestimate their risk of developing lung cancer and that they tend to see themselves as having a lower risk than the average smoker. Given the huge losses potentially involved in all these scenarios, one would expect people to verge more on the Bayesian side than the wishful one. ■

The following prediction is a trivial corollary from *P1*. Consider the simple framework presented in *Section 2.1* and assume that there are two deciders whose utility has the functional form defined in [1]. For one decider, we assume that $dv/(d\hat{\pi}) > 0$, so believing that the state is red gives her some utility. For the other one, however, $dv/(d\hat{\pi}) < 0$, so she gets utility from thinking that the true state is blue. Assuming that both deciders have the same priors, the toy model predicts that the deciders' optimal beliefs will differ. Specifically, the first (second) decider will inflate (deflate) the probability of the state Red.

Prediction 5 (P5): If two deciders have sufficiently opposed preferences, their beliefs will be opposed as well, even if they have access to the same set of information.

Discussion and evidence: There are to our knowledge few controlled studies on this issue. The existing evidence seems mixed. On the positive side, recall Mayraz (2011), discussed after *P1*, where Bakers and Farmers had opposed preferences and expressed different beliefs given the same information –a time series of “wheat” prices. Regarding negative evidence, Bar-Hillel and Budescu (1995) ask subjects each week over a four-week period to estimate the probability that the Dow Jones Industrial Average (DJIA) would change by more than 20 points in a week. Half of the subjects were eligible for a cash prize if the DJIA changed by more than 20 points, while the other half were eligible if it changed by less than 20 points. Obviously, each group had opposing interests in what regards the state of the world, and according to NBO this should translate into some differences in their probability judgments. Yet no significant differences were found in any of the four weeks of their study. Finally, some indirect evidence comes from comparing valenced and neutral treatments, i.e., treatments with and without a desired state of the world, respectively. Most of the papers cited after *P2* consider both types of treatments and it must be noted that few studies report a difference across treatments –e.g., Eil

and Rao (2011) and Möbius et al. (2014). Barron (2019), for a negative example, compares belief updating by two groups of individuals who share the same prior belief and receive an equally informative signal, but for one group the signal goes in line with their interests while for others the signal has no positive or negative value. His results show that there are no significant differences across groups. ■

As we have noticed before, few NBO models formalize in detail the learning process, with exceptions like Möbius et al. (2014). Suppose however that agents can choose whether or not to process a signal that has been observed (i.e., to not pay attention to it, explain it away, or not think about it); if the agent gets anticipatory utility from putting high probability on the good state, then agents may selectively ignore bad news. If people systematically choose to under-update to negative news, in other words, a natural response could be a biased search for information, i.e., avoiding those sources that typically bring negative news. This seems rational given that the bad news are hardly taken into account. The next prediction is then a straightforward one.

Prediction 6 (P6; information avoidance): Deciders avoid evidence or data that might challenge their favorite beliefs, even if they are aware that such information exists and it is costless to obtain. In contrast, they actively search for evidence that is expected to be consonant with their favorite beliefs.

Discussion and evidence: In each treatment of Eil and Rao (2011), subjects are asked at the end of the experiment their willingness-to-pay (WTP) to learn whether they are in the top half of the IQ or beauty ranking. Their WTP was positive for those who believed to be in the top ranks, whereas those who had arrived at “bad” posteriors required a subsidy. In the control, WTP was not affected by beliefs or signal history. In a previous experiment by Brock and Balloun (1967), participants are presented several sound recordings, some of them with speeches supporting or challenging the link between smoking and lung cancer. The recordings were masked by static noise that the participants can remove by repeatedly pressing a button. Interestingly, the smokers pressed the button significantly more than nonsmokers during the speeches that questioned the link between smoking and lung cancer, but less in the speeches that supported that link. The range of situations in which similar patterns of behavior have been reported is quite heterogeneous. In financial contexts, for example, Karlsson et al. (2009) observe that investors check the evolution of their portfolios more frequently in rising markets than when they are falling. This kind of behavior has been often called “ostrich effect”. Also, some people at risk of suffering some health problems seem reluctant to get diagnosed. For example, in the survey conducted by Kellerman et al. (2002) with populations at high risk for HIV, the main reasons reported by individuals refusing to get tested were denial of risk and fear of being HIV-positive. Similarly, in their study with individuals at risk for Huntington disease (HD), Oster et al. (2013) find that some of the individuals avoid to get tested. Moreover, these individuals tend to be overoptimistic about their probability of having HD and behave as if they do not have HD. In the study by

Ganguly and Tasoff (2016), some participants were reluctant to get tested for the HSV-1 and the HSV-2 (herpes simplex virus) and were even willing to renounce to some payoff in order to avoid the test. Further, the avoidance rate was three times higher for the HSV-2, which was perceived as worse than the HSV-1.

Other works provide evidence of information avoidance in problems in which one's actions may have an impact on others and in which moral principles play a role. In those situations, moral rules may constrict individual action. However, if individuals are not sure about the consequences of their actions on others, some of them may choose the most convenient action without feeling their moral principles compromised. This is often known as the “moral wiggle room”. In Dana et al. (2007), for instance, subjects play a binary dictator game with (decider, receiver) allocations A: (6, 1) and B: (5, 5). In an additional treatment labeled *hidden information*, the decider does not know a priori the receiver's payoffs in options A and B, knowing only that they can respectively equal \$1 and \$5 or \$5 and \$1, as determined by a coin flip before the session –deciders could choose to reveal the true payoff scheme at no cost before choosing. The idea of information avoidance implies that deciders will believe that option A leads to the socially optimal allocation (6, 5), having no reason to reveal the true payoffs, since they are sure that A is the ‘good’ option. Indeed, only 26 % of the deciders chose option A in the baseline, whereas 44 % of them preferred to remain ignorant in *hidden information*, where around 70 % chose A. To test the robustness of these results, Larson and Capra (2007) conduct a slightly different experiment, obtaining similar results (in this case, 53 % of participants in the hidden payoff treatment chose to remain ignorant). ■

Prediction 7 (P7): Predictions P1 to P6 above do not depend on the complexity of the data available or the attention paid to it.

Discussion and evidence: the available evidence on self-serving biases does not seem to support P7, as the extent of the bias seems to depend on the quality of the data sample. For instance, Haisley and Weber (2010) find that the behavior observed in a scenario similar to that in Dana et al. (2007) above is even more pronounced in more ambiguous contexts. Specifically, the experiment includes a “simple risk” and an “ambiguous” treatment. In the former, the receiver's payoff scheme is determined by a random draw from an urn which contains 10 red and 10 blue chips. In the “ambiguous” treatment, the specific content of that urn was unknown, although its composition was determined randomly using a uniform distribution from 0 to 20 (note, however, that the a priori probability of sampling a red chip from the ambiguous urn is 0.5 as in the “simple risk” treatment). Haisley and Weber (2010) report data on beliefs, e.g., deciders systematically overestimate the receivers' expected payoff: the mean estimate bias was of \$0.29 among the participants in the simple risk treatment and of \$0.89 among those in the ambiguous treatment. For another example, Dahl and Ransom (1999) find that Mormons who attend church regularly, serve in volunteer church positions, and have previously served as missionaries tend to report an increased tithable income base, which suggests that knowledge prevents

biases.¹⁶ Further, church members exhibit stronger biases in the tithing scenarios that they view more ambiguously, i.e., when they express more uncertainty, possibly as a result of having weak evidence on the issue. Finally, Babcock et al. (1997) explore several ways to reduce self-serving biases in pre-trial bargaining, using the *same* experiment as Babcock et al. (1995). In this study, as the reader may recall from the discussion of *PI*, subjects play either the role of plaintiff or defendant. Further, they read the same case materials, estimate the judge's decision and what a fair settlement would be like, and bargain over the case. Yet the authors find that having subjects who already know their roles list the weaknesses in their own case significantly decreases differences in their estimates of the judge's award and in the occurrence of impasse. This suggests that attention paid to the non-favorable evidence can reduce the extent of the self-serving biases. ■

Another important issue regards the sources of belief heterogeneity. In the models of NBO, including our toy model, two individuals may keep different beliefs if they have access to a different set of information (different priors or observed signals). This is in line with Bayesian models. However, in the models of NBO, belief heterogeneity can be the result as well of heterogeneous preferences. Suppose that each member of a group of people must rank the beauty of each member of the group, including themselves. If beauty is important for their self-esteem and the cost of making unrealistic estimates is low enough, the NBO models predict that their beliefs will differ. Specifically, everyone will overestimate their own beauty, even if all of them have access to the same information. Finally, in Bénabou and Tirole (2002) and Bénabou (2015), different individuals may keep different beliefs depending on how aware they are that their recalled sample is systematically biased. Some individuals may be fully aware that the evidence they recall is a biased sample that favors their preferred state. In this case, they will adequately incorporate this in their beliefs, alleviating the bias to some extent, although not completely. On the other extreme, fully naïve individuals will form their beliefs as if their recalled sample was unbiased.

Prediction 8 (P8; heterogeneity and individual co-variates): In models of optimism, the sources of belief heterogeneity are the evidence observed, the individual preferences and the individual's degree of awareness about their bias.

Discussion: Some evidence suggests a more complex picture. For instance, some studies point out that individuals with a higher ability to solve complex problems are apparently more likely to act as motivated believers. In Kahan, Peters, Dawson, and Slovic (2017) subjects are presented with a problem that requires drawing valid causal inferences from empirical data. One problem is ideologically neutral, dealing with a new skin-rash treatment. Here subjects who scored highest on numeracy tests did substantially better than less numerate ones. A second problem is more ideologically

¹⁶ Indeed, more active church members are significantly less uncertain about what income sources are tithable.

charged, having to do with a study on a gun-control ban. The impact of numeracy on performance in this condition was minimal. In fact, higher numeracy improves a subject's performance in detecting covariance only when the correct response is congenial to her subjects' political outlooks. Apparently, numerate subjects use their capacity selectively to conform their interpretation of the data to the result most consistent with their political outlooks. In Möbius et al. (2014), in contrast, high and low performers on the IQ test do not update differently: Both exhibit conservatism and asymmetric updating. To complicate matters further, Coutts (2019) reports that performance in his math and verbal skills quiz is (weakly) related to both conservatism and asymmetry; specifically, subjects with higher ability appear to be less pessimistic and conservative. Grossman and Owens (2012), in turn, observe asymmetry only among those participants with the lowest quiz scores and unrealistically high priors. Similarly, Ertac (2011) reports that self-serving tendencies are more prevalent among those subjects who exhibit most confident priors. Perhaps this signals some kind of confirmation bias, although the evidence is not conclusive, as we have noted earlier.

Gender stands out as another individual variable whose potential correlation with positive thinking deserves to be explored. In this regard, Ertac (2011) finds that men interpret positive feedback in the verbal task less conservatively than women, leading to less pessimistic posteriors. This was not observed in the math task, though, which both genders regarded as much less difficult. Möbius et al. (2014) also report that men update beliefs about performance in the IQ quiz significantly less conservatively than women, although no significant differences in asymmetry are found across genders. Similarly, Coutts (2019) finds no difference in asymmetry between genders. In contrast to Möbius et al. (2014), he reports no significant effect of gender on conservativeness, at least when interacted with ability in the quiz. Finally, Mayraz (2011) finds no difference across genders in the size of the bias.

3. Conclusion

Suppose that you get a nicely written e-mail. It includes a list with five names and promises that if you send one Euro to each of these five guys, put your own name on the list, and forward the new list to as many people as you can, you will receive in 20-60 days around 10,000 Euros in the mail. If you do not send the e-mail, you will not only miss out on the money, but bad fortune will befall to you. Encouragingly, the e-mail states that this is not a scam or a pyramid scheme. Indeed, it includes the testimony of some fellow who claims to have earned a HUGE amount of money and assures you that this program works perfectly every time. How does this sound to you?

Economic theories of optimism predict that people will fall victim of illusions when the cost of holding such illusions is not high. As our review of the literature suggests, there seems to be a puzzle in this respect. On one hand, the abundant evidence reviewed after *Prediction 1* suggests that the phenomenon of motivated

inference is a real one. Indeed, alternative, cognitive explanations of the fact that our preferences frequently interact with our beliefs appear to be overall less convincing. On the other hand, the more controlled tests surveyed after *Prediction 2* tend to find little evidence for asymmetric updating, a point made as well by Benjamin (2019). In general terms, the economic models of NBO do not seem to fully capture the phenomenon.

This means that the precise causes of the optimistic biases are still unclear. We find several open questions. First, under what conditions do people update in an asymmetric manner? Contrary to what the NBO models state, the mere existence of an interest and a small risk of a loss are not enough –e.g., Gotthard-Real, 2017; Barron, 2019. In this regard, second, the conditions required on the signal for the occurrence of a bias are still in doubt. One conjecture is that timing and salience are relevant, because the biases might be the result of biased recall and selective attention –Kunda (1990), Benábou and Tirole (2012). This is a point that has been relatively underexplored: in all the economic studies cited in the discussion of *P2*, for instance, subjects get feedback at all times about previous signals. Some of these studies, e.g., Eil and Rao (2011) and Möbius (2014), interestingly, still find a bias. This could be a signal that NBO does not require that subjects “forget” or “misinterpret” signals. From our point of view, however, these studies present confounds because the subjects’ priors are not exogenously fixed (see footnote 12). It must be noted, third, that the studies differ in the characteristics of the event/state about which subjects’ beliefs are controlled. Eil and Rao (2011), Ertac (2011), and Möbius et al. (2014) elicit beliefs about self-relevant states, e.g., one’s intelligence or attractiveness, whereas other studies focus more on financial domains. Perhaps people are relatively more optimistic about some issues. In some studies, fourth, the priors are objective whereas they are of a more subjective nature in others, particularly in those about ego-relevant events. As Coutts (2019) notes, this could affect updating behavior. Further, the signal generating process is always explicit and objective in these studies. In the field, however, inference scenarios are often characterized by high ambiguity.

A fifth question is heterogeneity. What (if any) personal characteristics reinforce asymmetric updating? As we discussed after *P8*, the NBO models are relatively agnostic in this regard, but there are several variables that might be worth considering. Gender stands as a very prominent candidate. While it seems well-established that men are more self-confident than women, even when actual performance is controlled for (e.g. Niederle & Vesterlund, 2007; Barber & Odean, 2001), it is still unclear whether women and men differ in the way they process information. Sixth, a potential confound in the existing studies regards the belief elicitation mode. In the vast majority of the studies cited in *P2*, specifically, subjects must estimate posterior probabilities using incentive compatible methods like the quadratic scoring rule or the lottery method. A concern therefore is that subjects are required to exert substantial mental effort, as conditional probabilities are difficult to compute, and the elicitation methods are complex –indeed, they often require intuitive additional explanations,

e.g., Coutts (2019), Barron (2019).¹⁷ Could people infer in a very reflexive manner in these scenarios, perhaps diminishing optimistic biases? Although the evidence so far seems negative (see *P4*), seventh, the role of material incentives on motivated inference possibly warrants further research. Does the risk of a (significant) loss lead to more Bayesian updating, as the NBO models predict?

Another question, eighth, refers to the persistence of the biases, particularly when people interact and exchange ideas. NBO models predict that optimistic biases should be relatively resistant, at least if people do not change their perceptions about their present and future preferences.¹⁸ This means a reluctance to admit mistakes and change opinion when others (including experts) think otherwise; although this resistance would be moderated if people are sophisticated and anticipate their biases, Bénabou and Tirole (2002). Arguably, a policy designed to make citizens' beliefs more accurate would be useless, since beliefs will hardly change, and harming, because it would make people unhappier (if it works). Further research should be devoted to explore the relationships between optimism and socialization: Are (optimistic) people immune to others' opinions?¹⁹ When do people expect to be able to convince others, if they are thought to be wishful thinkers? Last, but not least, there is the question of functionality and welfare. That is, economists have long considered Bayes' rule as a normative benchmark for belief updating, based on an instrumental argument. If beliefs yield intrinsic utility, however, it might be optimal in some circumstances to hold distorted, non-Bayesian beliefs. What are these circumstances, that is, when does the positive side of distortion dominate the negative one?

References

- Akerlof, G. A., & Dickens, W. T. (1982). The economic consequences of cognitive dissonance. *The American Economic Review*, 72(3), 307-319.
- Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. *The self in social judgment*, 1 (pp. 85-106).
- Babcock, L., & Loewenstein, G. (1997). Explaining Bargaining Impasse: The Role of Self-Serving Biases. *Journal of Economic Perspectives*, 11(1), 109-126.
- Babcock, L., Loewenstein, G., & Issacharoff, S. (1997). Creating convergence: Debiasing biased litigants. *Law and Social Inquiry*, 22(4), 913-925.

¹⁷ Ertac (2011) runs some non-incentivized sessions and concludes that her results are not an artifact of the belief elicitation method. Still, subjects must report numerical likelihoods.

¹⁸ Conversely, NBO predicts that beliefs should *rapidly* change if preferences over them change –e.g., if the monetary prizes associated to the states change their relative positions. We do not know of any evidence in this respect.

¹⁹ In Dahl and Ransom (1999), to illustrate some issues here, slightly over half of the survey respondents discuss what items to tithe with their spouse, while around 40% have sought outside advice about tithing. Further, individuals who seek advice from spouses are significantly more generous in their tithing donations.

- Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. (1995). Biased judgments of fairness in bargaining. *American Economic Review*, 85(5), 1337-1343.
- Babcock, L., Wang, X., & Loewenstein, G. (1996). Choosing the wrong pond: social comparisons in negotiations that reflect a self-serving bias. *Quarterly Journal of Economics*, 111(1), 1-19.
- Barber, B. M., & Odean, T. (2001). Boys will be boys: gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 116(1), 261-292.
- Bar-Hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking and Reasoning*, 1(1), 71-103.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141(4), 757-773.
- Barron, K. (2019). *Belief updating: Does the 'good-news, bad-news' asymmetry extend to purely financial domains?* Unpublished manuscript.
- Baumeister, Roy F. (1998). The Self. In D. Gilbert, S. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology*. McGraw-Hill.
- Bénabou, R. (2015). The economics of motivated beliefs. *Revue d'économie politique*, 125(5), 665-685.
- Bénabou, R. & Tirole, J. (2002). Self-Confidence and Personal Motivation. *Quarterly Journal of Economics* 117(3), 871-915.
- Bénabou, R. & Tirole, J. (2016). Mindful economics: the production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3), 141-164.
- Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 2, pp. 69-186). North-Holland.
- Benoit, J. & Dubra, J. (2011). Apparent Overconfidence. *Econometrica*, 79(5), 1591-1625.
- Brenner, L. A. (2003). A random support model of the calibration of subjective probabilities. *Organizational Behavior and Human Decision Processes*, 90(1), 87-110.
- Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes*, 65(3), 212-219.
- Brock, T. C., & Balloun, J. L. (1967). Behavioral receptivity to dissonant information. *Journal of Personality and Social Psychology*, 6(4), 413-428.
- Brunnermeier, M. K. & J. A. Parker (2005). Optimal Expectations. *American Economic Review*, 95(4), 1092-1118.
- Budescu, D. V., Wallsten, T. S., & Au, W. T. (1997). On the importance of random error in the study of probability judgment. Part II: Applying the stochastic judgment model to detect systematic trends. *Journal of Behavioral Decision Making*, 10(3), 173-188.
- Buser, T., Gerhards, L., & Van Der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, 56(2), 165-192.
- Caponecchia, C. (2010). It won't happen to me: An investigation of optimism bias in occupational health and safety. *Journal of Applied Social Psychology*, 40(3), 601-617.
- Case, K. E., & Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity* (2) (pp. 299-362).
- Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior*, 104 (pp. 1-23).

- Chew, S. H., Huang, W., & Zhao, X. (2019). Motivated False Memory. <https://ssrn.com/abstract=2127795> or <http://dx.doi.org/10.2139/ssrn.2127795>
- Clark, J., & Friesen, L. (2009). Overconfidence in forecasts of own performance: an experimental study. *Economic Journal*, 119(534), 229-251.
- Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2), 369-395.
- Cross, K. P. (1977). Not can, but will college teaching be improved? *New Directions for Higher Education* (17) (pp. 1-15).
- Dahl, G. D., & M. R. Ransom (1999). Does Where You Stand Depend on Where You Sit? Tithing Donations and Self-Serving Bias. *American Economic Review*, 89(4), 703-727.
- Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1), 67-80.
- Di Tella, R., Perez-Truglia, R., Babino, A., & Sigman, M. (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *American Economic Review*, 105(11), 3416-42.
- Eil, D., & Rao, J. M. (2011). The Good News–Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics*, 3(2), 114-38.
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133-40.
- Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior and Organization*, 80(3), 532-545.
- Festinger, L. (1957). A theory of cognitive dissonance. Stanford University Press.
- Forsythe, R., Rietz T., & Ross, T. (1999). Wishes, Expectations, and Actions: Price Formation in Election Stock Markets. *Journal of Economic Behavior and Organization*, 39 (pp. 83-110).
- Ganguly, A., & Tasoff, J. (2016). Fantasy and dread: The demand for information and the consumption utility of the future. *Management Science*, 63(12), 4037-4060.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European review of social psychology*, 2(1), 83-115.
- Ging-Jehli, N. R., Schneider F. H., & Weber, R. A. (2019). *On self-serving strategic beliefs* (Working Papers No 315). University of Zurich.
- Gino, F., Norton, M. I., & Weber, R. A. (2016). Motivated Bayesians: feeling moral while acting egoistically. *Journal of Economic Perspectives*, 30(3), 189-212.
- Gotthard-Real, A. (2017). Desirability and information processing: An experimental study. *Economics Letters*, 152 (pp. 96-99).
- Grossman, Z., & Owens, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior and Organization*, 84(2), 510-524.
- Haisley, E. C., & Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, 68(2), 614-625.
- Heger, S. A., & Papageorge, N. W. (2018). We should totally open a restaurant: How optimism and overconfidence affect beliefs. *Journal of Economic Psychology*, 67 (pp. 177-190).
- Kahan, D. M., Peters E., Dawson E. C., & Slovic P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), 54-86.
- Kahneman, D., & Tversky A. (Eds.) (2000). *Choices, Values, and Frames*. Cambridge University Press and the Russell Sage Foundation.

- Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, 38(2), 95-115.
- Kellerman, S. E., Lehman, J. S., Lansky, A., Stevens, M. R., Hecht, F. M., Bindman, A. B., & Wortley, P. M. (2002). HIV testing within at-risk populations in the United States and the reasons for seeking or avoiding HIV testing. *Journal of Acquired Immune Deficiency Syndromes*, 31(2), 202-210.
- Kindleberger, C. P., & Aliber, R. Z. (1978). *Manias, Panics, and Crashes: A History of Financial Crises*. 4th ed. Wiley, 2000.
- Kruger, J. (1999). Lake Wobegon be gone! The 'below-average effect' and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77(2), 221-232.
- Kuhnen, C. M. (2015). Asymmetric learning from financial information. *The Journal of Finance*, 70(5), 2029-2062.
- Kunda, Z. (1987). Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories. *Journal of Personality and Social Psychology* 53(4), 636-47.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.
- Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes*, 102(1), 76-94.
- Larson, T., & Capra, C. M. (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making*, 4(6), 467-474.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, 20(2), 159-183.
- Mayraz, G. (2011). *Wishful Thinking* (Working Paper).
- Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat T. (2014). *Managing Self-Confidence* (Working Paper).
- Moore, D. A., & Cain, D. M. (2007). Overconfidence and underconfidence: when and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, 103(2), 197-213.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3), 1067-1101.
- Oster, E., Shoulson, I., & Dorsey E. R. (2013). Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease. *American Economic Review*, 103(2), 804-830.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: divergent perceptions of bias in self versus others. *Psychological Review*, 111(3), 781.
- Rabin, M. (1994). Cognitive Dissonance and Social Change. *Journal of Economic Behavior and Organization*, 23(2), 177-94.
- Rabin, M., & Schrag, J. (1999) First Impressions Matter: A Model of Confirmatory Bias. *The Quarterly Journal of Economics*, 114(1), 37-82.
- Samuelson, L. (2004). Modeling Knowledge in Economic Analysis. *Journal of Economic Literature*, 42(2), 367-403.
- Schwardmann, P., & van der Weele, J. (2019). Deception and self-deception. *Nature Human Behavior*, 3(10), 1055-1061.

- Sharot, T., Korn, C. W., & Raymond Dolan (2011). How Unrealistic Optimism is Maintained in the Face of Reality. *Nature Neuroscience* 14(11), 1475-79.
- Sharot, T., & Garrett, N. (2016). Forming Beliefs: Why Valence Matters. *Trends in Cognitive Science*, 20(1), 25-33.
- Shiller (2000). *Irrational Exuberance*. Princeton University Press..
- Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63(2), 129-138.
- Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association*, 13(1), 130-167.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2), 143-148.
- Thompson, L., & Loewenstein, G. (1992). Egocentric Interpretations of Fairness and Interpersonal Conflict. *Organizational Behavior and Human Decision Processes*, 51(2), 176-197.
- Tversky, A., & Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5(2), 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131.
- von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34(1), 1-16.
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and social psychology*, 39(5), 806-820.
- Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine*, 10(5), 481-500.
- Weinstein, N. D., Marcus, S. E., & Moser, R. P. (2005). Smokers' unrealistic optimism about their risk. *Tobacco Control*, 14(1), 55-59.
- Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on cognitive dissonance*. Lawrence Erlbaum.
- Wiswall, M., & Zafar, B. (2015). How do college students respond to public information about earnings? *Journal of Human Capital*, 9(2), 117-169.