2020
05

Working Paper
INSTITUTO DE POLÍTICAS Y BIENES PÚBLICOS (IPP)

Self-serving recall is not a sufficient cause of optimism: An experiment

Adrián Caballero and Raúl López-Pérez

INSTITUTO DE POLÍTICAS Y BIENES PÚBICOS – CSIC

Instituto de Políticas y Bienes Públicos
Consejo Superior de Investigaciones Científicas
C/ Albasanz, 26-28
28037 Madrid (España)

Tel: +34 91 6022300
Fax: +34 91 3045710

http://www.ipp.csic.es

How to quote or cite this document:

Caballero, Adrián and López-Pérez, R. (2020). Self-serving recall is not a sufficient cause of optimism: An experiment. Instituto de Políticas y Bienes Públicos (IPP) CSIC, Working Paper. 2020-05

Available at: digital.csic.es

Self-serving recall is not a sufficient cause of optimism: An experiment*

Adrián Caballero† and Raúl López-Pérez‡

Abstract: A recent experimental literature has documented that people are (sometimes) asymmetric updaters: Good news are over-weighted, relative to bad news. We contribute to this literature with a novel experimental test of a potential mechanism of asymmetric updating, that is, that people recall better the positive than the negative evidence. In our design, this account predicts inflated posteriors regarding some future financial prize. Contrary to that, the average subject (slightly) underestimates the mode of the posterior beliefs about that payoff prospect. Although subjects tend to exhibit self-serving recall (SSR) in that they remember better the positive realizations of a signal in a memory task, this has little effect on their estimates and the extent and direction of the bias. A difficulty to recall accurately, we conclude, is not a sufficient cause for a positivity bias.

Keywords: Belief Updating; Biases; Motivated Beliefs; Optimism; Self-Serving Recall.

JEL Classification: D03, D80, D83, D84**.**

* We are grateful to Alexander Coutts, Eli Spiegelman, Florian Zimmerman, and participants at a seminar in the IPP-CSIC for helpful comments and suggestions. Needless to say, all errors remain our own. We also 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 Fabio Casalegno, Sergio Rubio, and Sara Yuste.
† Department of Economic Analysis, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain. e-mail address: adrian.caballero@uam.es
‡ Institute of Public Goods and Policies (IPP), Spanish National Research Council (CSIC), C/ Albasanz, 26–28, 28037, Madrid, Spain. E-mail address: raul.lopez@csic.es

1. Introduction

A growing literature documents a positivity bias in human beliefs, in that people *sometimes* arrive to excessively optimistic expectations regarding self-relevant events and future material outcomes, relative to the Bayesian benchmark –Bénabou and Tirole (2016), Epley and Gilovich (2016), Kunda (1990), Sharot et al. (2011), Wicklund and Brehm (1976). The specific mechanisms leading to such bias, however, are still not well understood: When is it more likely to appear? What personal characteristics correlate with it? This paper uses experiments to explore the idea that memory, or more precisely a selective recall of memories conditional on their valence (negative/positive), is *one* source of the bias. In other words, we explore whether the (non-Bayesian) optimists tend to be those who better recall the good news, particularly in scenarios where memory is obstructed by the absence of feedback or records.

Numerous researchers have defended the idea that optimism can be caused or *reinforced* by self-serving recall (SSR), or similar ones. Epley and Gilovich, (2016, p. 133) contend that preferences influence "the way evidence is gathered, arguments are processed, and memories of past experience are recalled", while Bénabou and Tirole (2016, p. 149) note that "several complementary and de facto equivalent cognitive mechanisms can sustain motivated updating, but the simplest one is selective recall or accessibility of past signals".[1] Kunda (1990, p. 483) illustrates the phenomenon, noting that "people who want to believe that they will be academically successful may recall more of their past academic successes than of their failures". As we will argue later, however, the existing evidence on the role of SSR is somehow mixed. Further, it is also relatively scarce: in many controlled studies on motivated inference, for instance, recall of signals is always extremely easy because subjects get feedback –e.g., Barron (2020), Coutts (2019), Eil and Rao (2011), Ertac (2011), Gotthard-Real (2017), and Möbius et al. (2011). While these studies are clearly important, they cannot shed much light on our research question, as selective recall seems highly unlikely in these settings (indeed, these studies had different research goals than ours). However, understanding whether SSR affects the formation of optimistic beliefs is relevant because such beliefs can motivate suboptimal individual and collective decisions. If the hypothesis is correct and we want to prevent those decisions, it follows that the appropriate strategy should be focused on the 'selective' individuals, giving them feedback or some kind of reminder on a regular basis.[2]

[1] Bénabou and Tirole (2002, p. 871) cite one of Friedrich Nietzsche's apothegms in *Beyond Good and Evil*: "I have done this, says my memory. I cannot have done that, says my pride, remaining inexorable. Finally —memory yields".

[2] Optimism has also a positive side. For example, a positive view about one's own abilities or morality can boost self-esteem. In fact, optimism has been associated to a good mental and physical health (Rasmussen et al., 2009; Strunk et al., 2006). Also, a positivity bias can motivate individuals to pursue their goals and overtake the obstacles that may arise

To clarify further our assumptions, the paper first provides a parsimonious model of inference with self-serving recall, which can be applied to any scenario in which people update their beliefs about the prevalence of some group, class, or category, or about the frequency of occurrence of some repeatable event, e.g., the infection fatality rate (IFR) within some age group of COVID-19 or some other disease. When people observe some relevant signal, specifically, the model predicts that they estimate frequencies by extrapolation from the signals that they *recall* having observed. Importantly, people have preferences regarding the frequency of the event. Typically, they will prefer low rates if the event is 'bad', e.g., fatality within the person's age group, and high rates if it is 'good', e.g., earning a large financial payoff after investing in some asset. This preference for some rates or states instead of others is the basis of the SSR hypothesis, i.e., the likelihood of recalling some signal increases if it has positive valence, namely, is in line with the preferred rates/states. To illustrate, consider an agent called Adam who has access to four COVID-19 studies; two of them suggest a relatively low IFR in Adam's age group, whereas the other two indicate a larger IFR. Whereas a Bayesian Adam would probably adopt a rather circumspect stance given the evidence available, an optimistic (pessimistic) Adam would tend to 'recall' better the first (last) two studies, i.e., the positive (negative) ones.[3]

In short, the model of optimism just described is based in two ideas: (i) people extrapolate from the signals they recall and (ii) the SSR hypothesis. To test this model, we run a balls-and-urns experiment where each subject faces a box with 100 balls. Each ball has a different boy or girl name; the proportion $\theta \in [0, 1]$ of 'female' balls in the subject's urn is randomly determined for each participant. The subject then observes one by one an indeterminate number of draws with replacement from her urn –30 draws, in fact– and must afterwards provide an estimate $\hat{\theta}$ of θ, with a payoff for accuracy. From a statistical viewpoint, it is a very simple problem which requires extrapolation from the sample observed: If the empirical frequency of female balls in the sample is $f \in [0, 1]$, the best estimate is $\hat{\theta} = f$. Given our research goal, though, we introduce two aspects so as to induce optimism, that is, an 'inflated' estimate of θ ($\hat{\theta} > f$). First, subjects earn 0.50 euros for each female ball in their urn, so that they have a preference for θ to be high –note that if $\theta = 1$, i.e., in the

(Bénabou and Tirole, 2002, 2004). If we wanted instead people to be positively-minded, therefore, the hypothesis would recommend a blurring of prior memories.

[3] This has possibly implications for behavior: Even if he cares about others, an optimistic Adam might wash less his hands and keep less physical distance, particularly if others come from the same age group and are unlikely to interact and thus spread the virus within other groups where the IFR is higher (note that Adam may have Bayesian beliefs on the IFR in other age groups). For evidence that optimism might play a role in health decisions, see for instance Oster et al. (2013), who find significant differences in the behavior of tested and untested individuals at risk for Huntington disease, a hereditary condition. Specifically, individuals at risk who refuse to get tested are optimistic about their health and behave as those who certainly do not have Huntington disease regarding some events of their lives (financial decisions, retirement, marriage, etc.) in which diagnosed individuals behave significantly different.

'best of the worlds', the prize amounts to 50 euros. It follows that female draws are good news and hence more likely to be recalled according to the SSR hypothesis. Second, we give no feedback to subjects, who are moreover not allowed to keep records of the extractions, and are explained the incentivized estimation task only after they have seen the 30 consecutive draws. In addition, recall is greatly hindered, as numerous distracting tasks are placed between the extractions. We expect the SSR hypothesis to be particularly relevant in this setting.

The first test of the model is of an indirect nature, and the evidence is arguably negative. Specifically, we observe that most people do not report 'inflated' estimates, and when they do so, the difference $\hat{\theta}$ - f is rather small. Specifically, around 34 percent of the subjects overestimate θ, and the median deviation amounts to just 4 balls. Interestingly, these subjects tend to face samples with a relatively small f. This squares badly with the idea that optimism is caused by SSR. To clarify, suppose that both A and B recall better the 'good news' and that A sees 10 female extractions and B 20 (out of 30). Given their selective memories, both should report higher estimates than their respective Bayesian estimates, i.e., $f = 1/3$ and $f = 2/3$. Contrary to this, we find that subjects like A are more likely to overestimate. In our setting, therefore, optimism is relatively infrequent, of a rather limited extent, and depends on characteristics of the sample that should be irrelevant by assumption.

Our second test of the model is more direct. *After* the estimation of θ, subjects are inadvertently asked to recall for a prize as many names observed in the 30 prior extractions as possible, the *recalled sample* henceforth. According to the SSR idea, positive, i.e., female signals should leave a stronger memory trace and, indeed, this is what we find: subjects in our experiment are more likely to recall female than male extractions. Yet several results indicate that such selective recall does not induce optimism in our experiment. As we have noted, first, people do not systematically provide inflated estimates of θ, even if the average recalled sample overstates the actual prevalence of good signals. Second, this pattern of recall does not correlate with the overestimations, i.e., 'optimistic' subjects are not relatively more likely to recall female extractions. Third, the Bayesian standard, which assumes that people extrapolate from the whole sample, outperforms a model in which people estimate θ by extrapolation from the recalled sample. In this respect, the R-squared of a linear regression where the dependent variable is a subject's estimate and the X-variable is the share of female balls in the whole sample equals 0.551, whereas the R-squared of a regression based in the recalled sample goes down to 0.324. In summary, although people display SSR in our recall task, this is insufficient to generate optimism and in fact offers little insight on subjects' previous estimates of θ.

To account for the absence of optimism in our data, we have checked the possibility that subjects are sophisticated enough to anticipate SSR, and hence correct the recalled sample

accordingly –see Bénabou and Tirole (2002) for a formalization of the idea and some psychological justification. Suppose for instance that a female draw is twice more likely to be recalled than a male draw, perhaps because subjects rejoice such lucky events and hence pay more attention to them. If a subject anticipates this and her recalled sample includes 6 female names and 3 male ones, he might conclude that most likely θ is around 0.5, and not around 2/3, as her (selective) recollection indicates. That is, people might not extrapolate from the recalled sample, but from a corrected one. To explore this idea, subjects responded two non-incentivized questions after the recall task: (I) the percentage of female names that they had recalled correctly in that task, relative to the total number of female names sampled, and (II) the corresponding percentage for the male names. Ratio I/II takes value 1 if a subject expects no recall bias, whereas I/II > 1 denotes an anticipated SSR bias. Sophisticated subjects should accurately evaluate this ratio which, recall, tends to be actually larger than 1 for most subjects. Contrary to this, people tend to underestimate the ratio, as they overestimate the denominator II. That is, they expect to have better memory for the bad news than they actually have. As a result, subjects consistently fail to recognize the selective nature of their recall, and there is not much difference in this respect between those who inflate or deflate θ. On top of that, a subject's (previous) estimate of θ is not influenced by her beliefs about I and II. Subjects verge more on naiveté than sophistication, and do not seem to extrapolate from a corrected sample.

All things considered, we find scarce evidence for the idea that 'fuzzy' environments where accurate recall is difficult lead to optimism via SSR. Our findings, we believe, have credibility for several reasons. First, we arguably control the individuals' priors, since they know that θ is uniformly distributed between 0 and 1, as well as the signals observed. This is not so typical in previous studies, particularly psychological ones or those coming from the field. Second, we do not elicit probabilities in our experiment, but just a proportion, and the estimation task is computationally undemanding. We can still test our main hypotheses, but do not face the ensuing confounds if the task instead required, say, the application of Bayes' rule. Third, memorization is hindered in our design, as subjects do not even know that they have to recall something, they observe relatively large samples, no feedback is given, and distracting tasks are placed between the signals. Fourth, we use a within-subjects design: The same subjects (i) observe signals, (ii) estimate θ, and (iii) have their memories about the signals observed in (i) elicited. Fifth, in line with Bénabou and Tirole (2002), we check for the possibility that individuals are at least partially aware of their memory biases and exhibit some degree of sophistication by correcting their estimations accordingly. Finally, we give incentives in many of our relevant tasks, something that is not at all usual in the psychological literature.

The rest of the paper is organized as follows. The next section reviews some prior evidence on SSR, as well as some literature on how SSR relates to optimism.[4] Section 3 presents and discusses our theory, with the objective of illustrating more formally its main intuitions. Section 4 introduces the experimental design and reports results. Section 5 concludes. Note that this paper is part of a broader research program focused on the test of potential accounts of optimism; in a companion paper, Caballero and López-Pérez (2020), we use the data from this experiment to test some models like Brunnermeier and Parker (2005).

2. Literature review

There is evidence suggesting some form of self-serving recall in memory tasks. Part of this evidence comes from lab games, thus providing some support for the SSR hypothesis in social contexts. In Psychology, Shu et al. (2011) and Kouchaki and Gino (2016) show that people exhibit "unethical amnesia", i.e. people who behave dishonestly are more likely to forget over time the details of their actions than those who act ethically; in these studies, the memory tasks are not incentivized. The experiment conducted by Li (2013), in turn, has two parts. In the first one, participants play a version of the trust game. In the second one, run either (i) immediately, (ii) 7 days, or (iii) 43 days after the first part, depending on the treatment, participants complete an incentivized questionnaire about their choices and their counterparts' in the trust game. The main result is that those first movers who were "victims" of the trustee's selfish choice are more likely to forget than those who were benefited by the co-player. Perhaps this is an indication of SSR, as first movers tend to recall better the more positive interactions.

In Carlson et al. (2020), participants play 5 dictator games and are presented, after completing some distracting tasks, an incentivized memory task in which they guess the mean share of the endowment transferred to the recipient in the 5 games. Participants must also indicate the "maximum acceptable share" for the dictator to keep (before or after learning their role, depending on the treatment). Deciders tend to recall being more generous than they actually were, especially among those who kept a larger share than their stated "maximum acceptable share". A related study is Saucet and Villeval (2019). In their baseline IRA treatment, participants play 12 binary dictator games and perform a distraction task. Afterwards, deciders are (unexpectedly) asked to recall the amounts allocated to the receiver in each of the 12 games, which are randomly presented. Correct

[4] This review has therefore a restricted focus. Caballero and López-Pérez (2020) survey the literature on (i) motivated updating and (ii) the role of some potential predictors of optimism.

recalls are incentivized.[5] Motivated memory implies a better recall rate when the subject made an "altruistic" choice, i.e., one favoring the receiver, instead of a "selfish" one. This is confirmed by the data (32% vs. 23%).[6] In an alternative IRAC treatment where the choices are made by a computer, further, there is no evidence of selective recall (16.3% vs. 16.8%). Not all findings are entirely in line with SSR, however. In the baseline, for instance, most dictators who recall inaccurately after a selfish choice overestimate the receiver's amount; this is not the case when the choice was altruistic (57.4 vs. 31.82%, respectively). This seems a priori consistent with SSR. Yet such pattern is also found in the IRAC treatment, where dictators bear no responsibility in the choice of option. This suggests that the differences in overestimation result more from the payoff constellations associated to each option than from behavioral determinants. In addition, the magnitude of the overestimated recalls is not significantly different between altruistic and selfish choices.

In what regards SSR in individual decision problems, Sedikides and Green (2004) show that individuals recall self-threatening information poorly relative to praising information or information about others, whereas individuals who behave unethically are also more likely to forget the moral rules (Shu and Gino, 2012).[7] In Huang et al. (2020), subjects (N = 1143) answer four questions from an incentivized Raven's IQ test. Some months later they are shown the same four questions, plus two which are new but similar, each accompanied by the correct answer, 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 face a prize/loss if they recall correctly/incorrectly. Subjects' recall patterns show some systematic features. The most relevant one for our purposes is that, in aggregate terms, people are more likely to forget errors than successes, i.e., correct answers.[8] In Zimmermann (2020), subjects solved 10 Raven matrices and were then randomly matched into a group with nine other subjects. Subjects' beliefs about their rank in this group according to their prior performance in the IQ test were elicited both before and after they received (noisy) feedback; the quadratic scoring rule was used. In the Direct (1month) treatment, beliefs were elicited immediately (one month) after feedback. In the first case, subjects updated in the appropriate

[5] The role of incentives in motivating better recall is unclear. The authors run a variation of the baseline where correct recalls are not incentivized, finding an increase (32% vs. 25.3%) only when the decider chose the altruistic option. They conjecture that incentives motivate a higher effort to recall, but only to retrieve the memory of desirable decisions.

[6] Note yet that an alternative explanation is that altruistic people pay more attention or meditate more while deciding, thus recalling better any choice

[7] Because of the employed experimental design, accurate recall was not incentivized in Green and Sedikides (2004). The pattern found by Shu and Gino (2012) holds both with and without monetary incentives in the memory task.

[8] Additionally, people are more likely to err on the positive than the negative side, i.e., they have relatively more wrong memories of correct answers. Further, people fabricate events that did not actually happen, but mostly positive ones. 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. Since these two phenomena are de facto equivalent to SSR in our model and experimental design, we abstract from them in our posterior analysis.

directions, irrespectively of the feedback. Yet beliefs elicited one month later substantially underweight negative feedback.[9] When people were incentivized to pay attention, however, they incorporated the negative feedback in their beliefs. This is the main finding from the Announcement treatment, which was based on 1month, except that subjects were informed at the first lab meeting that one month later they would have their beliefs about performance elicited. Of particular interest, Zimmerman (2020) also conducted a Recall treatment, identical to 1month except that instead of measuring beliefs, he measured subjects' recall of the feedback, with an incentive for accuracy. He finds evidence in line with SSR, so that receiving mostly negative feedback leads to relatively less accuracy one month later. In a RecallHigh treatment identical to Recall, except that the prize for accuracy was significantly higher (50 vs. 2 Euros, respectively), however, those who received negative feedback had a better recall accuracy. Results from Announcement and RecallHigh suggest that incentives can foster greater attention and thus more belief accuracy.

In this line, SSR does not seem a universal and unconditional phenomenon. In Li (2019), participants perform five rounds of a word-entry task and then estimate the number of mistakes as well as their position in some ranking (incentives for accuracy are provided for both estimates). Fully informative feedback is provided at the end of each round, so that participants are aware of whether they overestimated or underestimated their absolute and relative performance. In a second part conducted 40 days later, the same subjects participate in an incentivized memory task in which they recall the number of mistakes and ranking position, as well as whether they overestimated or underestimated those numbers. The results show that SSR is not homogeneous among individuals, e.g. some participants recall too many mistakes and others too few.

In summary, the evidence so far seems favorable to the SSR hypothesis, although with some qualifications. A different issue is whether there exists a link between optimism and SSR. Note that the link is not obvious: even if (some) people exhibit self-serving recall in memory tasks, one cannot take for granted that their *prior* behavior and/or inferences are based on the information elicited in those tasks. In this respect, an additional finding in Li (2019) gives particularly noteworthy within-subjects evidence, i.e. participants who overestimated (underestimated) their ranking in the first part of the study exhibit too 'positive' ('negative') memories in the second part. See also the evidence from the 1month and Recall treatments cited above from Zimmerman (2020). In turn, Thompson and Loewenstein (1992) explore labor negotiations, and find that subjects representing opposite sides later remember, from the same case file (presented before the negotiations), more facts favoring their

[9] In a No Feedback treatment in which subjects received no feedback and their beliefs were elicited again one month after the IQ test, these beliefs did not differ from the priors.

position than going the other way. The more divergent their recalls, importantly, the longer and costlier is the delay to agreement in the bargaining phase –see also Loewenstein et al. (1993).

In contrast, the studies by Garrett et al. (2014), Ma et al. (2016); and Sharot et al. (2011) offer negative evidence of a link, using a common design. Participants are presented an adverse event E, e.g., suffering a car accident, and have a few seconds to estimate their chances of facing E in the future. This is repeated for a total of 80 different events. In a second stage, subjects are briefly shown, one by one, the actual frequency with which each event E happens among individuals living in the same socio-cultural environment as them and must guess their posteriors of encountering E in the future. The three studies report evidence for asymmetric updating in favor of good news, but this cannot be *apparently* explained by SSR. In effect, after the scanning session, participants had to recall the (previously presented) actual frequency of each of the 80 events. The errors thus committed did not depend on whether the actual frequency was better or worse than initially expected by the participants, i.e., whether it was bad or good news.

To finish, note well that our research question is not whether SSR is a necessary condition of optimism, but a sufficient one. Indeed, many different factors can generate a positivity bias. In most economic studies on asymmetric updating, for instance, subjects receive feedback so that biased recall should play no role. In Eil and Rao (2011), participants observe the signal three times and are given always feedback about prior rank guesses and signals. Similarly, subjects in Ertac (2011), Möbius et al. (2011), Gotthard-Real (2017), Barron (2020), and Coutts (2019a) respectively observe 1, 4, 4, 5, and 3 signals, always with proper feedback. These studies, that is, are intentionally designed to minimize forgetfulness about prior signal realizations. Still, some of them, e.g., Eil and Rao (2011) and Möbius et al. (2011), find a positivity bias. It seems therefore that asymmetric updating does not require that subjects "forget" or "misinterpret" signals altogether.

3. Inference with SSR: A model

We start by introducing some general notation, together with the standard Bayesian theory. Afterwards, we formalize the idea of inference with self-serving recall.

3.1 General setup & the Bayesian model

Let time be indexed as $t = 1, 2\ldots$ At period $T \geq 1$, an expected payoff-maximizer called Eve must estimate the frequency/rate $\theta \in [0, 1]$ with which some phenomenon f occurs. Specifically, there is an i.i.d. signal S, taking on value $v \in \{f, m\}$, and such that probability $(S = f) = \theta$ –for expositional purposes, we sometimes refer to f as female, and m as male. Eve does not know the exact value of θ. Let $\Theta \subseteq [0, 1]$ denote the space of potential values of θ –for expositional

convenience, we assume that Θ is finite. Eve has prior beliefs over Θ, quantified by a finitely additive probability measure *p* mapping each event or subset of rates $E \subseteq \Theta$ to a probability p(E). Let p_k denote Eve's priors about rate $\theta_k \in \Theta$. In our experiment, to illustrate, $\Theta = \{0, 0.01, \ldots, 1\}$, whereas the (uniform) prior of any rate θ_k is $p_k = 1/101$.

Eve has observed in each period some realizations of S and hence can use that evidence to update her priors. Let Eve's history of observation of f be represented by a T-dimensional vector $F = (f_1, \ldots, f_T)$ where $f_t \geq 0$ is an integer representing the number of female realizations of S observed at t. In addition, let M denote an analogously defined vector so that m_t indicates the number of male observations at t. The number of female observations up to period T is denoted as $f = \sum f_t$, that of male ones as $m = \sum m_t$, whereas the total number of observations is $\sum f_t + m_t$. Given *data* D = (F, M), Eve's posterior beliefs about any $\theta_k \in \Theta$ are obtained by means of Bayes' rule (the last equality is true only if priors are uniform):

$$p_{k|D} = \frac{p_k \cdot \theta_k^f \cdot (1-\theta_k)^m}{\sum_\Theta p_j \cdot \theta_j^f \cdot (1-\theta_j)^m} = \frac{\theta_k^f \cdot (1-\theta_k)^m}{\sum_\Theta \theta_j^f \cdot (1-\theta_j)^m} \qquad (1)$$

3.2 Inference with self-serving recall

A "limited" agent called Adam infers as Eve, except for a single exception: When he observes any evidence, his beliefs over Θ do not change exactly as in expression (1). The intuition here is that Adam may forget or omit some observations of the signal, either due to inattention, limited recall or any other cognitive factor. In this regard, let I_{tf} and I_{tm} respectively denote the 'memory indicator' of any eventual female and male observation at time t ($I_{tf}, I_{tm} \in \{0, 1\}$ for any t), $\tilde{f} = \sum_{t=1}^{T} f_t \cdot I_{tf}$ denote the recalled number of female observations, and $\tilde{m} = \sum_{t=1}^{T} m_t \cdot I_{tm}$ correspondingly denote the recalled number of male observations. Vector $(\tilde{f}, \tilde{m})$ is the *recalled sample*. To form his posteriors, we posit that Adam applies Bayes' rule, but using the recalled instead of the actual numbers of female and male observations (the last equality assumes uniform priors):

$$\tilde{p}_{k|D} = \frac{p_k \cdot \theta_k^{\tilde{f}} \cdot (1-\theta_k)^{\tilde{m}}}{\sum_\Theta p_j \cdot \theta_j^{\tilde{f}} \cdot (1-\theta_j)^{\tilde{m}}} = \frac{\theta_k^{\tilde{f}} \cdot (1-\theta_k)^{\tilde{m}}}{\sum_\Theta \theta_j^{\tilde{f}} \cdot (1-\theta_j)^{\tilde{m}}} \qquad (2)$$

In other words, Adam uses the same rule as Eve, but infers based on a sub-sample of the objective data, due to his limited attention or memory. Specifically, indicators I_{tx} and I_{ty} need not equal 1 for any t. When an indicator is nil for some t, Adam omits/forgets the corresponding observation, which leaves no trace. To formalize this idea, we posit indicators to be random variables. By varying the determinants of the probability $\pi(I_{tv})$ that indicator I_{tv} takes value 1, $v \in \{f, m\}$, one gets different specifications of the model. The Bayesian model of inference assumes of course that all agents are like Eve, with $\pi(I_{tf}) = \pi(I_{tm}) = 1$ $\forall t$; that is, no data omitted. One potential

deviation from this idea says that people can omit or forget data due to 'cold' memory failures, e.g., old data is ceteris paribus more easily forgotten, but also inattention to some contextual event if we are focused on other stimuli. Given our research goal, we omit cold factors in our analysis.

A second deviation from the Bayesian model, most relevant here, considers omissions due to 'hot', i.e., motivated, memory failures. Note that while both cold and hot factors are likely to affect recall, the implicit assumption in the literature seems to be that hot factors have a stronger effect than cold ones. To model these hot factors, assume that Adam has preferences over set Θ, that is, regarding which rate or state of the world is the actual one. By this we simply mean that Adam prefers the realization of some state(s).[10] Let rate $\theta_P \in \Theta$ denote Adam's favorite or preferred one; we assume for parsimony that Adam's preferences have either a single peak θ_P or no peak at all, i.e., absolute indifference with regard to θ. In our experiment, for instance, $\theta_P = 1$. The following SSR hypothesis states that Adam is most likely to recall the evidence that fits him, e.g., the female extractions in our experiment:

Hypothesis (self-serving recall): If there is no peak θ, $\pi(I_{tf}) = \pi(I_{tm})$. If there is a peak $\theta_P > ½$, then $\pi(I_{tf}) > \pi(I_{tm})$ for any t. If $\theta_P < ½$, in turn, $\pi(I_{tf}) < \pi(I_{tm})$ for any t. If $\theta_P = ½$, finally, $\pi(I_{tf}) = \pi(I_{tm})$ for any t, i.e., any realization is equally likely to be recalled, independently of its value.

To sum up, Adam updates as if he recalled the signals self-servingly, but then processes such recalled sample like Eve, i.e., as a Bayesian. Note that a measure of the strength of SSR at time t is the difference $\pi(I_{tf}) - \pi(I_{tm})$, which the SSR hypothesis implicitly assumes non-negligible and constant through time.[11] Observe also that this basic framework admits many extensions. We finish this section by describing one of them, to be later checked with our experiment. This extension is motivated by Bénabou and Tirole (2002) (henceforth BT), who present a model in which individuals can, within limits and possibly at a cost, affect the probability of remembering a given piece of data. BT embed this problem of inference within a decision problem, but we abstract from the latter in what follows. To nest BT into our framework, let $\lambda \in [0, 1]$ denote the share of male realizations that Adam recalls accurately, i.e., $\pi(I_{tm}) = 1$ for any such observation. Further, all female realizations are

[10] In Akerlof and Dickens (1982), for instance, the worker in a risky job prefers the state in which he suffers no accident and hence no material loss. For another example, if Adam is fair-minded and acts as decider in a Dictator game, the best of the worlds is one where he takes all money and still acts fairly, e.g., because he is (or presumes to be) much needier than the counterpart.

[11] Note also that this difference is not very sensitive to the specific value of θ_P (leaving aside the SSR hypothesis). This means that the mode of Adam's posterior beliefs might not coincide with θ_P. The model could be changed to prevent this issue. If $\theta_P = 1$, for instance, Adam might forget any male observation, at the same time considering any female one –i.e., $\pi(I_{tf}) = 1$ and $\pi(I_{tm}) = 0$ for any t. If $\theta_P = 0.6$, in contrast, Adam might have a more balanced pattern of recall, generating a sample $(\tilde{f}, \tilde{m})$ such that 0.6 is exactly the mode of the posterior distribution determined by (2). Still, this variation of the SSR hypothesis is hardly consistent with the evidence we later report, as people rarely report a mode equal to 1,

correctly recalled. Adam can decrease λ with respect to its "natural" value $\lambda_N \leq 1$, but "choosing" a smaller recall probability involves a "memory cost" M(λ), with M (λ_N) = 0, and $\frac{dM}{d\lambda} \leq 0$ for $\lambda < \lambda_N$.[12]

An interesting insight in BT is that, although the individual can manipulate λ, he is aware that there are incentives that result in selective memory. If Adam is sophisticated, that is, he anticipates some motivated omissions and hence a biased recalled sample. More precisely, he thinks that a share $\lambda^* > 0$ of the male realizations are recalled accurately. If the number of male observations that he recalls is $\tilde{m}$, in other words, he concludes that the actual number of realizations is $\tilde{m}^* = \tilde{m}/\lambda^*$. This can be introduced into expression (2) instead of $\tilde{m}$ so as to derive Adam's beliefs. If $\lambda^* = 1$, Adam is naïve and unaware of any self-serving recall (provided that $\lambda < 1$). If $\lambda^* = \lambda$, on the other hand, Adam accurately anticipates the degree to which he engages into self-serving recollection.

4. Experimental design & data analysis

4.1 Experimental design and procedures

In our experiment, any subject faces her own virtual urn, with 100 balls inside. Each ball in the urn has either a boy or a girl Spanish name, and the 100 names in the urn are different. Balls with a girl/boy name are called henceforth female/male balls –we did not use these terms in the subjects' instructions; see Appendix I. The precise rate θ of female balls in a subject's urn is a multiple of 0.01 selected by the computer with uniform probability over the interval [0, 1] at the start of the session; the rate of male balls is hence $1 - \theta$. Although the subject does not know θ, the method to determine it is common information.[13] Priors are hence arguably fixed.

Each subject then observes the realization, i.e., name, of an a priori undetermined number (in fact, 30) of consecutive random draws with replacement from her/his box.[14] Subjects did not observe others' samples. After the first 15, 22 and 30 extractions, further, the subject is asked to provide a point *estimation* of θ –therefore, she gives estimates in 3 rounds, each one with a progressively enlarged dataset. In the analysis below, however, we will focus on the third round unless otherwise noted –Caballero and López-Pérez (2020) offer data on the other rounds. Subjects were explained

[12] BT also allow for the possibility that λ is increased over its natural value, again at a cost.

[13] To determine the specific names in each urn, we used two lists with the most popular, non-compound female and male names in Spain, respectively. These lists, elaborated by the Spanish National Statistics Institute, order the names according to frequency; see https://www.ine.es/en/welcome.shtml. We excluded foreign names from the lists, e.g., Mohamed, as some subjects might find them relatively unfamiliar. We are hence rather sure that our subjects were able to discern whether a name was female or male, and also to spell it, something very relevant for the recall task (see below). Once θ had been randomly determined for a subject, we randomly selected $100 \cdot \theta$ different girl names and $100 \cdot (1-\theta)$ boy names in the corresponding lists to 'fill' the urn. Subjects were just told that the selected names were used with a relatively high frequency in Spain.

[14] To ensure that all subjects really 'observe' the extractions, each name is displayed in the screen besides a button that the subject must click to proceed to the next extraction; the position of the button in the screen changes in each extraction.

each estimation task only immediately after observing the corresponding extractions, and did not receive any feedback about prior extractions.

Subjects get either a 'state prize' that depends on the rate/state θ or an 'estimation prize' depending on the accuracy of the participant's last estimation of θ. The prize that a subject finally gets is randomly determined with probability 0.5 at the end of the experiment, to prevent hedging problems (Blanco et al., 2010). As a 'state prize', specifically, the subject gets 0.50 Euros for each female ball in the urn, e.g., a maximum of 50 Euros if θ = 1. For the 'estimation prize', in turn, let $\hat{\theta} \in [0, 1]$ denote a subject's last, i.e., third, estimation. The subject earns 10 euros if the corresponding error $|\theta - \widehat{\theta}|$ is smaller or equal to 0.02, and 0 euros otherwise; we opted for eliciting the mode of the subject's beliefs because it is a rather straightforward statistical problem which does not require de facto the computation of the probability of any rate, a substantially more demanding problem. The elicitation of the first two estimations of θ, in turn, was not incentivized. While the participants are informed about the nature of the 'state prize' before they observe any extractions, the structure of the 'estimation prize' is only revealed just before the last estimation task, i.e., after the 30 extractions. Indeed, the initial instructions only stated that with probability 0.5 the subject will get either the 'state prize' or an undefined prize whose nature will be specified later (otherwise, we were afraid that subjects would count the female and male draws from the outset).

Additional tasks and questions are inserted *between* some extractions, so as to hinder memorization of the names drawn. After the first 7 extractions, specifically, we included a brief questionnaire where we gathered information on personal and socio-demographic characteristics (age, gender, major, religiosity, and political ideology). A risk aversion index was elicited after the first 19 extractions.[15] Also, subjects completed an expanded cognitive reflection test or CRT (Frederick, 2005), including the three classical questions and two additional ones, after the first 26 extractions. Furthermore, after the third estimation task, i.e., the incentivized one, subjects had to report the shortest 95% confidence interval they could figure out –Caballero and López-Pérez (2020) analyze this data.

After this interval estimation, additionally, we included a 'recall task': Subjects had 90 seconds to introduce as many extracted (female and male) names as possible and were paid 0.40 euros for each 'correct' name, i.e., actual extraction. From this payoff, we deducted 0.20 euros for each 'incorrect', non-extracted name, so as to prevent subjects from introducing well-known,

[15] This variable does not predict biases in our experiment; see Caballero and López-Pérez (2020) for a fuller discussion.

common names that had not been extracted.[16] Note yet that the minimum payoff from this recall task was zero, i.e., subjects could not lose money here. The goal of this recall task was to elicit a subject's recalled sample, namely, the dataset from which she theoretically extrapolates and estimates θ. Some readers may argue though that the SSR hypothesis makes sense only for signals that are 'relevant' to the inference problem at hand. In other words, since the specific names observed are inconsequential for the estimation task, the hypothesis cannot be properly tested using our recalled samples. Alternatively, one could have asked subjects the number of female and male draws that they recalled having observed. We pondered this issue, and finally opted for our design choice for three reasons. First of all, subjects are not informed in advance that the names extracted are inconsequential; hence the argument does not seem to apply in our context. Regardless, second, the idea that the only stimuli that leave a memory trace are the consequential ones seems to fit badly with introspection and memory research. For instance, in the first experiment conducted by Shu and Gino (2012), participants are presented two essays (an academic honor code and a text about eligibility for a Massachusetts license) before they perform an incentivized problem-solving task. Both texts were irrelevant for this task and incentives to recall them were not provided, yet participants generally recalled some of the content in them. In general, if an episode like an extraction constitutes good news and good news are better recalled than bad news, we find natural that people keep a more accurate memory of the episode, including of aspects that are ex post neutral (like the names). Since the recall task came after the estimation task, finally, we were afraid that the estimate of θ would act as an anchor in a question like: how many female balls have you observed? The correlation between both answers would then be very high, but possibly highly artificial. The evidence coming from a different memory task, performed five months later, is in fact consistent with this presumption, as we will detail below.[17]

In summary, there are three prizes. **1**: State prize equal to $50 \cdot \theta$. **2**: A prize equal to 10 Euros, if the last point estimation was accurate enough. **3**: Letting C and W denote the number of correct and wrong names in the recall task, the subject got a prize for recall equal to $0.4 \cdot C - 0.2 \cdot \min\{2 \cdot C, W\}$ Euros. As explained above, each subject gets **3** for sure and either **1** or **2**, randomly determined. After the recall task, subjects responded two questions so as to check whether they expected to recall better female than male extractions, i.e., good than bad news; we describe them in detail later. In

[16] Since we wanted to elicit the recalled sample, subjects were allowed to introduce the same name several times, which could be relevant if the name was actually drawn several, i.e., $m > 1$, times. If they introduced that name n times, they earned $0.4 \cdot n$ Euros if $n \leq m$, and $0.4 \cdot m - 0.2(n-m)$ otherwise. That is, incorrect entries were penalized.

[17] Relatedly, our initial plan was to run an additional treatment where the male names pay, i.e., not the female ones as in our control. This would prevent confounds in case we had found evidence in favor of the SSR hypothesis, i.e., to make sure that female names are not just intrinsically easier to remember. Since the evidence in favor of the hypothesis is so scarce, however, we have abstained from running that treatment.

addition, they answered two questions on statistical knowledge, the LOT-R test on optimism (Scheier et al., 1994; Scheier and Carver, 1985), and a test on disappointment, in this order, thus ending the experiment.

The study consisted of six computerized sessions at Universidad Autónoma de Madrid, with a total of 68 participants. The software used was z-Tree (Fischbacher, 2007). Participants were not students of the experimenters. After being seated at a visually isolated computer terminal, each participant received written instructions that described the decision problem. Subjects could read the instructions at their own pace and we answered their questions in private. Understanding of the rules was checked with a computerized control questionnaire that all subjects had to answer correctly before they could start making choices –see Caballero and López-Pérez (2020) for details. At the end of the experiment, subjects were informed of their final payoff and paid in private. Each session lasted approximately 60 minutes, including paying subjects individually, and on average subjects earned 20.50 euros, including a show-up fee of 3 euros.

4. 2 Research hypothesis and data analysis

Consider first the Bayesian model presented in Section 3.1. If Eve were a subject in our experiment, she would face a rather simple problem of inference. Let $f \in [0, 1]$ denote the (rounded) frequency of female balls in the sample observed by Eve, i.e., $f = \frac{f}{m+f}$. Since priors are uniform in our experiment, it follows from a standard Bayesian argument that Eve's posterior beliefs have a unique mode at $\theta_k = f$ and a concave shape. Given the structure of the estimation prize in the third round, therefore, Eve reports there an estimate $\hat{\theta} = f$ except when the sample observed is 'extreme', i.e., contains 0, 1, 29, or 30 female balls; in these cases, she reports an estimation slightly different than *f*, a point that we take into account in our analysis below –e.g., by distinguishing between *f* and the Bayesian prediction given *f*, denoted by $\hat{\theta}_B(f)$ in what follows.[18] Our first research hypothesis is hence direct.

Hypothesis I (Bayesian): A subject who observes a sample where the (rounded) share of female balls equals $f \in [0, 1]$ chooses *f* as an estimate of θ. The only exception appears if the sample observed contains extremely few or extremely many female balls.

Evidence: On average, the subjects' urns have around 56.7 female balls, i.e., the mean θ equals 0.567. After the 30 extractions, furthermore, the mean $\hat{\theta}_B(f)$ is 0.578, while the subjects' average estimate of θ equals 0.530. Hence we observe a systematic (although small) underestimation

[18] For an example, suppose that Eve observes 30 (0) female balls, so that there are most likely 100 (0) female balls in the urn. Since the estimation prize allows for a maximum error of 2 balls, however, she maximizes her chances to get that prize if her estimate is of 98 (2) balls instead –see Caballero and López-Pérez (2020) for a more detailed discussion.

of the number of female balls For a deeper analysis, not focused on average figures, we define a subject's deviation in a round as the difference between her actual estimate of θ and the predicted Bayesian estimate (given the sample so far observed by the subject). Figure 1 shows the distribution of deviations in the last estimation; the intervals are of size 0.1. As we can see, subjects rarely deviate in more than 10 balls from the Bayesian estimate, and they tend to err on the negative side.

Note that the difference between the mean Bayesian and subjects' estimates is marginally significant (paired t-test, p-value = 0.068). This difference persists when we exclude the 10 subjects who observed extreme samples (paired t-test, p-value = 0.015). Hence, the under-estimation observed does not seem an artifact of the estimation prize. Overall, the evidence observed is rather favorable to Hypothesis I. ■

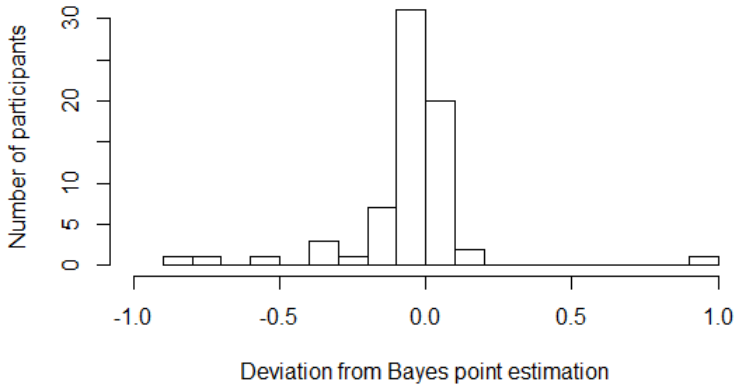

Figure 1: Distribution of subjects' deviations in the last, incentivized estimation round

We move now to the model of biased recall presented in Section 3.2 above. The next research hypothesis is again straightforward and parallels Hypothesis I above.

Hypothesis II (extrapolation): Leaving aside extreme cases, Adam reports $\tilde{f} \in [0, 1]$ as an estimate of θ, where $\tilde{f} = \frac{\tilde{f}}{\tilde{m}+\tilde{f}}$ is the (rounded) share of female balls in the sample that he recalls, specifically in the recalled sample obtained in the memory task.

Let $\hat{\theta}(\tilde{f})$ denote Adam's estimate given $\tilde{f}$. Hypothesis II says that $\hat{\theta}(\tilde{f}) = \tilde{f}$ in general, except for instance if $\tilde{f} = 0$, in which case $\hat{\theta}(\tilde{f}) = 0.02$. If we posit that the recall probabilities $\pi(I_{tv})$ follow the SSR conjecture, so that Adam displays self-serving recall, we get in addition the following corollary:

Hypothesis III (estimation with SSR): In average, $\tilde{f} > f$ so that the average Adam over-estimates θ, i.e., reports $\hat{\theta}$ larger than $\hat{\theta}_B(f)$.

Evidence: We focus for the moment on Hypothesis III. As we have seen, the average subject does not inflate, i.e., overestimate θ. Therefore, the hypothesis is rejected. Still, the average can mask some heterogeneity so that we can consider a more nuanced version of Hypothesis III, according to which only some agents infer as our model predicts. In this respect, it must be noted that 33.82 percent of the participants overestimate θ in the third round, i.e., report $\hat{\theta} > \hat{\theta}_B(f)$. Among these 'optimistic' subjects the median deviation $\hat{\theta} - \hat{\theta}_B(f)$ equals 0.040 and the average one 0.086. For the sake of comparison, the median and average deviation were -0.055 and -0.145 among the 52.94 percent of subjects who underestimated θ in the same round, i.e., $\hat{\theta} < \hat{\theta}_B(f)$. Thus the extent and strength of the optimistic bias is arguably limited; if any, it is the pessimistic bias that stands out. For further illustration, the share of subjects who overestimate by more than 10 (20) balls is 4.41 (1.47) percent, while 17.65 (10.29) percent of subjects underestimate to the same extent.

Note also that the subjects who inflate (deflate) θ tend to face samples with a small (large) f or more precisely leading to a small (large) estimation $\hat{\theta}_B(f)$: the average $\hat{\theta}_B(f)$ equals 0.47 for the over-estimators and 0.63 for the under-estimators, a significant difference (t-test p-value = 0.0394). This is something that the SSR model cannot explain because the rate of subjects who inflate should be independent of the sample distribution, and seems perhaps more consistent with the joint hypothesis that people are Bayesian but may commit some random errors: If a subject observes, say, $f = 0.2$, he is more likely to deviate above 0.2, as there are more rates between 0.2 and 1 than between 0 and 0.2. Interestingly, this story can explain as well the prevalence of underestimation in our data: most (58.2%) subjects observed a sample with $f > 0.5$.[19] In summary, the evidence in favor of Hypothesis III is very limited, even allowing for heterogeneity. ■

The following result summarizes our key findings so far.

Result 1: The average and median subjects slightly underestimate θ. The share of subjects who overestimate is relatively small and these subjects deviate little from the Bayesian benchmark, or at least less than the under-estimators. Further, overestimation is more likely when the observed rate of female balls is relatively small, which cannot be explained by the SSR hypothesis.

Subjects do not inflate θ, contrary to what the model predicts. Yet, do they exhibit biased recall? The results from the recall task, which subjects completed after the third estimation round, allow us to check the following hypothesis, which is a straightforward implication of SSR.

[19] As an alternative explanation, we have received the comment that individuals might have underestimated the number of female names because of inattention, coupled with the fact that they were unaware of the total number of draws, i.e., 30. As we note in footnote 14, however, subjects were somehow 'forced' to see the extractions. Note also that subjects possibly estimate θ by extrapolation. If this is the case, the explanation amounts to say that subjects are relatively more inattentive on the female than the male extractions. This is possible, but somehow odd and not confirmed by the data cited below in footnote 21. In addition, this theory cannot explain the correlation just cited between inflation and a low f.

Hypothesis IV: Subjects are significantly more likely to accurately recall positive, i.e., female, extractions. This is particularly the case among those subjects who overestimate θ.

Evidence: As a starter, Figure 2 below depicts the distribution of subjects' accurate recollections, net of errors. As can be inferred from the graph, subjects forgot a large share of the 30 extractions actually observed by them: In average, only 23.53% of the extractions were accurately recalled. Further, there were also wrong recollections. On average, 18.43% of each subject's recalled names were wrong, either because a non-observed name was introduced or because a name was written more times than it had been sampled.

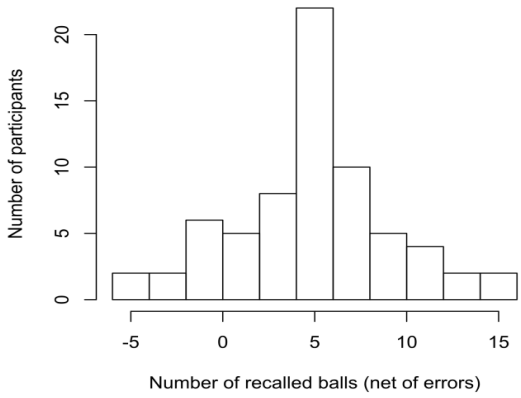

Figure 2: Distribution of subjects' correct insertions in the recall task, net of errors

In what directly regards Hypothesis IV, the likelihood of accurately recalling a female name equals 0.3023, while the corresponding figure for the male names equals 0.2126. In other words, subjects correctly recall around 30% of the female extractions, and 21% of the male extractions; this difference is significant (paired t-test, $p = 0.0064$). Also, the mean proportion of female names in the recalled sample[20] is 0.6794, which is significantly larger than the share of female names in the observed sample (paired t-test, $p < 0.001$). For a visual illustration, each dot in Figure 3 represents a subject, with coordinates (share of female extractions in the actual sample, share of female names in the set of recalled names). As we can see, most subjects are above the diagonal, a signal that they are more likely to accurately recall a female name. In summary, the evidence is favorable for the first half of Hypothesis IV.

As an aside, one might wonder why subjects tend to recall better the female names. Is this perhaps driven by the fact that they are objectively easier to recall, or due to their higher desirability

[20] Unless otherwise indicated, the recalled sample includes both accurate and wrong recollections, that is, names that were not actually observed by the participant –or even not included in our lists (see Footnote 13)–, as well as misspelled names. Our results do not change substantially if the analysis centers exclusively on the accurate recollections.

in our context, i.e., the SSR conjecture? Some *preliminary* evidence goes against the first interpretation. Conjecturally, that is, the ease-of-recall effect (if it exists) should be more pronounced among our female subjects. However, a regression analysis indicates that, keeping the actual share of female extractions constant, our female subjects do *not* introduce in the recalled sample a significantly higher share of female names (p = 0.756).[21]

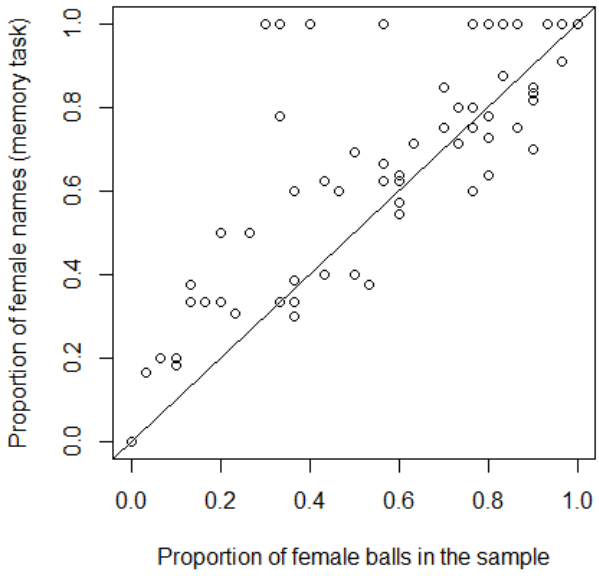

Figure 3: share of recalls that are female vs. objective sample

In what concerns the second half of Hypothesis IV, the data is less reassuring than for the first half. Specifically, we check the probability of recalling a female name conditional on whether the subject under or over-estimates. If inflation is due to self-serving recall, that is, the subjects who inflate θ should have a different recall pattern. We can hence compare the memory bias, defined as the difference between the share of female names introduced in the recalled sample and f. The mean value is 0.0909 for the 'pessimistic' subjects, i.e., those who deflate θ in the last round, and 0.1072 for the 'optimistic' subjects; these two values are not significantly different (p-value = 0.7593). This is hardly consistent with the idea that overestimation is triggered by a memory 'hot' bias. ■

Result 2: People forget most extractions, but the rates of posterior recall are significantly higher for the 'positive', female extractions. This pattern, however, is displayed by both inflators and deflators.

[21] An interesting finding is that, during the extractions, subjects dedicate relatively more time in average to a screen where a female name is drawn, although the difference is neither extremely large (2.75 and 2.40 seconds for each female and male name, respectively) nor very statistically significant (p-value = 0.0484). We are not totally sure how to interpret this result, but at least it suggests that people pay more attention to the female draws, which might partly explain the selective recall.

For a more detailed analysis of the potential mechanisms explaining behavior in our experiment, recall Hypothesis II above. It says that Adam's estimation depends on the sample he recalls. Importantly, it admits different specifications conditional on the assumptions about recall. One possibility is that subjects extrapolate from the recalled sample; this idea amounts to say that such sample accurately reflects the properties of the sample that subjects actually used during the estimation.[22] As we will show now, however, this story has less empirical support than the Bayesian model (Hypothesis I), which is incidentally also a special case of Hypothesis II. For an aggregate analysis, first, we define a participant's subjective deviation as the difference, *in absolute terms*, between her estimate (in the third round) and the share of female names in her recalled sample. The average and median subjective deviation in the third round is 0.2062 and 0.1100, respectively. This can be compared with the subjects' average and median absolute deviation from the Bayesian estimate, equal to 0.1060 and 0.0350 in that round, respectively. In other words, the average subject tracks more closely the actual frequency than the frequency in the recalled sample.

For more detailed econometric evidence, we first run a simple linear regression where the dependent variable is the subject's estimate of θ in the last round and the independent variable is (i) the observed empirical frequency f –results are basically identical if we use instead the Bayesian estimate $\hat{\theta}_B(f)$. This regression, therefore, considers the fit of Hypothesis I. In this model, the R-squared and the coefficient of variable (i) equal 0.551 and 0.745 ($p < 0.001$), respectively. For comparison, if the regression model includes (ii) the share of female names in the recalled sample instead of variable (i), the R-squared of this new model goes down to 0.324, while the coefficient of variable (ii) is highly significant ($p < 0.001$) and equals 0.614. The idea that people extrapolate from the recalled sample, therefore, fits worse the data than the Bayesian theory. Alternatively, if we regress the subject's estimate of θ on variables (i) and (ii) above, the fit of this regression is rather high, as measured by an R-squared equal to 0.673. In turn, the coefficient of variable (i) happens to be very close to one, more precisely 0.949, and very significant. Variable (ii), in turn, is marginally significant ($p = 0.085$) but its estimated coefficient equals -0.167, i.e., it has *negative* sign.

To check some potential heterogeneity, finally, we extend the prior model by adding (iii) a dummy taking value 1 when the subject overestimates θ, interacted with variable (ii). That is, perhaps the optimistic types focus on the recalled sample, while the remaining subjects are basically Bayesians. In this extended model, the R-squared increases up to 0.729. But the most interesting finding concerns the estimated coefficients, which are (i) 1.004, (ii) -0.229, and (iii) 0.223, all of

[22] As we have noted above, the recalled sample includes wrong recollections. Although the model in Section 3.2 excludes the possibility of wrong memories, we find more natural to assume that inference is based on the whole set of recollections, and not on the actually accurate ones. In any case, our findings are robust and do not change when the recalled sample is defined as the set of accurate recollections.

them significant at the 1% level. Note well that figures (ii) and (iii) have roughly the same absolute value, but different sign: this means that the over-estimators track better frequency *f*, that is, they estimate θ in a more Bayesian fashion! The reverse finding is that the under-estimators deviate more, which is perhaps not so surprising if we recall our discussion above on Hypothesis III. Overall, therefore, we find little evidence, if any, that people estimate based on the recalled sample.

Although the prior analysis is relatively favorable to the Bayesian theory, note well that other specifications of our model outperform that theory, and do not assume full recall of the sample. While we have conceived a few different specifications of this idea, and although a full analysis of this point is out of the scope of this paper, we propose for expositional purposes to distinguish two groups of subjects. The first group are those whose estimate deviates in less than 10 balls from the Bayesian one; they comprise 77.94% of the sample and fit almost perfectly the Bayesian model –a regression of these subjects' estimates of θ on variable (i) above gives an R-squared of 0.9788. The second group exhibit larger deviations and, as shown in Figure 1, the sheer majority of them under-estimate θ. While the first group got on average larger scores than the second group in the CRT (2.72 and 2 out of 5, t-test p = 0.1587) and in the recall task (5.68 and 4.53 accurately recalled names, p = 0.4676), these differences are not significant. Interestingly, however, individuals in the first group took on average considerably less time to successfully complete the control task (108.1 seconds rather than 155.2 seconds for the second group, p = 0.0376). This suggests that it might have been harder for individuals in the second group to understand the instructions of the experiment. While this may be one of the reasons under the observed heterogeneity, it still does not explain why individuals in the second group consistently underestimate θ. Specifically, the mean deviation in the second group is -0.208 and 80 percent of its members underestimate θ. ■

Result 3: A Bayesian model fits betters the subjects' estimations than a model assuming that people track the empirical frequency of female balls in the recalled sample. This is particularly true for the subjects who over-estimate θ. Most deviations from Bayes are underestimations.

We move now to a slightly different issue. That is, a potential reason for the scarce evidence on optimism in our experiment is that people are sophisticated, as Bénabou and Tirole (2002) suggest. That is, subjects might anticipate that they recall things in a biased, self-serving manner. As a result, they may not extrapolate from the sample obtained in the recall task, but from a corrected one; see Section 3.2. To explore this hypothesis, subjects responded two questions after the recall task so as to check whether they expected to recall better female than male extractions. More precisely, participants were asked (I) the percentage of female names that they had recalled correctly in the memory task, relative to the total number of female names sampled, as well as (II) the corresponding percentage for the male names; this is λ* in Section 3.2. For clarification, subjects

were noted that (I) and (II) should be the same if they believed that gender had not influenced the likelihood of recalling each name. In this respect, ratio I/II measures a subject's anticipated recall bias, taking value 1 if the subjects expects no bias, and a value larger than unity if female extractions are expected to be recalled more easily, consistent with the self-serving bias. The ratio can be compared with the actual figure, derived from the subject's recalled sample. In this line, Figure 4 represents, for each subject, her ratio I/II (Y-axis) and the actual rate of recall bias (X-axis).[23]

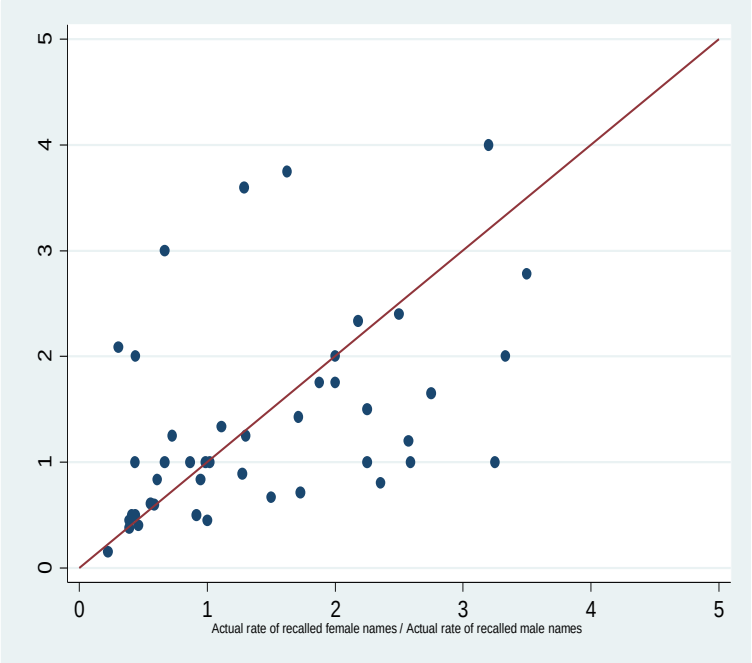

Figure 4: Subjects' actual and perceived recall bias (favoring 'good news')

On average, participants estimated that they had recalled 34.09 and 29.50 percent of the female and male names sampled, respectively, although this difference is not significant (paired t test, p-value = 0.5222). In contrast, they actually recalled 30.23 and 21.26 percent of the female and male names sampled. While the difference between the estimation and the actual rate of female names recalled is not significant (paired t test, p-value = 0.1489), it seems that participants overestimated their percentage of male names recalled (p-value = 0.0056). To sum up, the average subject does not anticipate, at least in statistically significant terms, that positive signals are recalled relatively better. Moreover, she expects to recall the negative, male signals better than she actually does. All this goes contrary to the sophistication idea and is more in line with the naiveté hypothesis. These aggregate findings are possibly apparent in Figure 4. On one hand, there are a few more subjects below the diagonal than above it. Further, many subjects did not expect recall to adopt a

[23] Note that the anticipated recall bias I/II cannot be computed if (II) equals zero; we face a similar problem with the actual recall bias if the subject recalled no male names, e.g., if her sample consisted only of female names. In total, 23 observations are not present in Figure 4 for these reasons. To facilitate visual analysis, further, 2 more observations were excluded for which the ratio I/II took values larger than 15.

self-serving pattern, even failing to anticipate the direction of their own memory biases, i.e., implicitly reporting an incorrect ratio of less than 1. Since subjects are heterogeneous, we have also explored whether those who expect more self-serving recall tend to inflate less, as they should according to Sophistication. The aggregate picture is not encouraging in this respect: the average (median) value of ratio I/II is 2.262 (1.464) among the subjects who *overestimate* in the third round, and 1.114 (1.000) among those who underestimate, although this difference is not significant (t-test, p-value = 0.1351). Further, a logit regression where the dependent variable is a dummy such that 1 = subject overestimated in round 3, and the independent variable is another dummy taking value 1 if subject's ratio I/II is larger than 1 also shows that the expectation of self-serving recall does not predict less overestimation (p-value = 0.804).

So far, therefore, we have found little evidence (if any) in favor of the sophistication idea. To further check this hypothesis, however, we follow the logic of the sophistication hypothesis in 3.2 and compute each subject's *corrected recalled sample*, based on the estimated rates of recall and the recalled sample obtained in the memory task.[24] For example, assuming that the individual thinks that all the names in the recalled sample are correct, her estimate of the number of female names in the corrected sample is computed as

$$Number\ of\ female\ balls = \frac{Number\ of\ female\ names\ in\ recalled\ sample}{Estimated\ rate\ of\ recall\ of\ female\ names}$$

The computation of the number of male balls is analogous. From these numbers, computing the share of female names in the corrected sample is straightforward; we denote this share as $\tilde{f}^*$. Then, we can study whether the estimated θ tracks $\tilde{f}^*$. For this, we compute a regression in which the dependent variable is the subject's actual estimation of θ and the independent variable is $\tilde{f}^*$. While the coefficient associated to $\tilde{f}^*$ is positive and significantly different from zero (0.523, p-value < 0.001), the estimated model fits the data worse (R-squared = 0.313) than the model that considers instead the actual frequency *f*, which has an R-squared of 0.551, as we noted above.[25]

Result 5: Many subjects underestimate the extent of their self-serving recall or even the existence of such type of bias. Inflation is not predicted by unawareness of a self-serving bias. The hypothesis that subjects infer based on a corrected sample has less explanatory power than the Bayesian model.

We finish with a final test of the SSR hypothesis presented in Section 3.2. We have found before that people typically display biased recall (although this does not explain inflated estimates or

[24] Note, however, that the size of the subject's corrected recalled sample is not necessarily coherent with the actual size of the sample, i.e., 30 balls.

[25] Note that $\tilde{f}^*$ can be calculated only if the individual estimated a non-nil rate of recall of both female and male names. For this reason, a total of 10 observations were excluded in the estimation of this model.

optimism; recall Result 2). In our experiment, however, the recall task was presented almost immediately after the last estimation, when the state prize was still uncertain. But what happens afterwards, particularly in the medium/long term? To explore this issue, we (unexpectedly) contacted our subjects by electronic mail around 5 months after the last session of the experiment was run.[26] The message text, available upon request, consisted of a brief reminder of the experimental design. In particular, we reminded subjects of the state prize and informed them that the total number of random extractions was 30. Further, we made two questions (Q1 and Q2), i.e., Q1: Each of the 30 draws you observed had a written name, how many had a female name? (the answer was requested to be an integer from 0 to 30), and Q2: In the recall task, how many female names did you remember correctly? And male names? (we noted that these two numbers could not add up to more than 30). Both questions were incentivized. In Q1, a subject earned 10 Euros only if the error was not above two balls. Further, she got 10 additional Euros if both numbers in Q2 were correct (she earned nothing for that question otherwise). Subjects had to answer both questions to be eligible for any of the prizes. We note that subjects did not know the answers to these two questions, that is, they were never informed about the correct figures when they participated in the experiment five months before – still, at the end of the corresponding session in November 2019, each subject was informed about the actual value of θ, i.e., the rate of female balls in her urn, and about the *aggregate* number of correct name insertions in the recall task.

Observe that, five months after the experiment, subjects should have no preference over θ, i.e., no peak θ_P, as there is no state prize coming. Hence, the SSR hypothesis predicts that they should be equally likely to forget female and male extractions, and hence not overestimate the answer to Q1. Alternatively, it could be the case that 'good news remain good news', irrespective of whether they are instrumental, and hence are recalled better. This account predicts overestimation.

Out of the 68 participants, 40 of them (58.82%) responded. Regarding Q1, the mean deviation from the real value was equal to 0.65 balls. The deviation was strictly negative, i.e. they underestimated, for 16 of the subjects, strictly positive for 13 of the subjects, and nil for the rest. In cumulative and absolute terms, 21 responders deviated at most in 2 balls, 34 in at most 4 balls, and 38 in at most 8 balls. Further, one subject deviated in 16 balls and another one in 20, both in an upwards direction. Leaving aside these two outliers, the average deviation is -0.26 balls. In summary, we find very little evidence of over- or under-estimation in what regards Q1. This is evidence in line with the SSR hypothesis although, for granted, it must be taken with some care given potential

[26] Subjects were contacted in April 2020, that is, during the strict lockdown that Spain endured due to the Covid-19 pandemic. The lockdown measures effectively banned people from leaving their homes except to go to work, buy essential supplies, or walk the dog. We do not know if these circumstances have affected our results, which should be taken therefore with some caution.

selection issues, i.e., the subjects who replied to our call might happen to be those who are best described by the hypothesis.

The answers to Q2 allow an additional test of two aspects regarding self-serving recall. First of all, we can check further whether people are sophisticated and hence anticipate some form of biased recall, e.g., Bénabou and Tirole (2002). For this, we compare (a) the share of female names in the sample, as indicated by the subject's response to Q1, with (b) the share of female names in the recalled sample, taking into account the responses to Q2. If (b) is larger than (a), subjects think that recall *at the time of the experiment* exhibited self-serving recall. In this respect, the mean value of the difference (b) - (a) was equal to 0.05%, which means that the average responder expected a negligible degree of biased recall –that is, if x% of the balls in the actual sample were female, then the average responder thinks that x +0.05 % of the balls in the recalled sample were female.[27] If we recall Result 2, this again suggests naiveté rather than sophistication.

Second, if individuals like to think that they have a good memory, the SSR hypothesis says that they should overestimate in April the aggregate amount of correct recollections in November, *even though*, as we said above, they were informed about the actual figure –i.e., aggregating female and male names– at the end of their session in November. This is indeed what we find. Among our 40 responders, specifically, the mean estimation of the total number of correct recollections was 8.55, significantly larger than the mean number of accurately recalled names among the same participants five months ago, which was 6.85 ($p < 0.001$). If we explore the overestimation of the number of recalled female and male names separately, results are somehow similar, as they are overestimated in around 23.16 and 28.57%, respectively.

5. Conclusion

Optimism is based on a learning pattern called asymmetric updating in the literature: Signal observations are over-weighted or under-weighted depending on the decider's goals, i.e., the target, optimal beliefs. In our experiment, the female extractions are 'good news', as they are evidence in favor of the most desirable state, i.e., $\theta = 1$. Hence, they should be over-weighted: Subjects recollect evidence so as to reinforce their rosy beliefs about the world. In this regard, researchers have suggested that asymmetric updating operates via biased recall. That is, people recall relatively better the positive signals. Since the recalled sample is biased, the estimates based on that subjective sample are biased as well in the positivity direction.

[27] The calculation of this mean value omits one responder (out of a total of 40) who answered 0 to both questions in Q2, i.e., who (accurately, in fact) responded that he/she remembered no name correctly in the recall task in November.

According to the SSR hypothesis, positive, i.e., female signals in our experiment should leave a stronger memory trace and, indeed, this is what we find: subjects are more likely to recall female than male extractions. Nonetheless, we do not find at the aggregate level that people overestimate θ –if any, they are more likely to underestimate it. Further, the *sign* of the estimation bias seems more related to the characteristics of the sample, i.e., individuals who observed a small/large proportion of female extractions are more likely to over/underestimate θ, rather than to subject-related variables, whereas the *size* of the bias may be explained at least partially by the participants' understanding of the experiment. When we confront the theoretical models to our data, further, a Bayesian model (based on the whole sample) fits them better than a model based on the recalled sample. Following Bénabou and Tirole (2002), in turn, we check the possibility that participants are to some extent aware of their asymmetric recall, therefore correcting their recalled sample when estimating θ, although no significant evidence is found in this line. These findings are reinforced by the results of an incentivized memory task conducted five months after our experiment.

Overall, therefore, our results indicate that inference in an environment where accurate recall is hindered need not lead to optimism. In our within-subjects design, people cannot recall all signals and yet they rarely overestimate θ significantly. While we do not know if our findings are the exception that proves the rule, at least they show that the absence of accurate memories is not sufficient for a positivity bias (due to SSR). Our results also suggest that the connection between memory tasks and estimation or inference tasks must be made with care, as recalled samples may have different properties than the samples actually used by each subject to elaborate her estimates. In other words, people might extrapolate from a different sample than the one obtained with incentives in a memory task, in circumstances that might be considered artificial. We think that this is an interesting methodological point in itself, given that recall tasks like ours are frequently used in research. In this sense, one should not take for granted that self-serving recall in memory tasks necessarily leads to optimistic beliefs. More research is warranted to discover the conditions most conducive to optimism.

Bibliography

Akerlof, G. A., & Dickens, W. T. (1982). The Economic Consequences of Cognitive Dissonance. *The American Economic Review*, *72*(3), 307–319. Retrieved from http://www.jstor.org/stable/1831534

Barron, K. (2020). Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics*. https://doi.org/10.1007/s10683-020-09653-z

Bénabou, R. (2015). The Economics of Motivated Beliefs. *Revue d'économie politique*, *125*(5), 665. https://doi.org/10.3917/redp.255.0665

Bénabou, R., & Tirole, J. (2002). Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, *117*(3), 871–915. https://doi.org/10.1162/003355302760193913

Bénabou, R., & Tirole, J. (2004). Willpower and Personal Rules. *Journal of Political Economy*, *112*(4), 848–886. https://doi.org/10.1086/421167

Bénabou, R., & Tirole, J. (2016). Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives*, *30*(3), 141–164. https://doi.org/10.1257/jep.30.3.141

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics*, *13*(4), 412–438. https://doi.org/10.1007/s10683-010-9249-1

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, *95*(4), 1092–1118. Retrieved from http://www.aeaweb.org/articles?id=10.1257/0002828054825493

Caballero, A., & López-Pérez, R. (2020). *An experimental test of some economic theories of optimism*.

Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E., & Crockett, M. J. (2020). Motivated misremembering of selfish decisions. *Nature Communications*, *11*(1), 2100. https://doi.org/10.1038/s41467-020-15602-4

Coutts, A. (2019). Good news and bad news are still news: experimental evidence on belief updating. *Experimental Economics*, *22*(2), 369–395. https://doi.org/10.1007/s10683-018-9572-5

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), 114–138. https://doi.org/10.1257/mic.3.2.114

Epley, N., & Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives*, *30*(3), 133–140. https://doi.org/10.1257/jep.30.3.133

Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior &*

Organization, *80*(3), 532–545. https://doi.org/https://doi.org/10.1016/j.jebo.2011.05.012

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Garrett, N., Sharot, T., Faulkner, P., Korn, C. W., Roiser, J. P., & Dolan, R. J. (2014). Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression . *Frontiers in Human Neuroscience* . Retrieved from https://www.frontiersin.org/article/10.3389/fnhum.2014.00639

Gotthard-Real, A. (2017). Desirability and information processing: An experimental study. *Economics Letters*, *152*, 96–99. https://doi.org/https://doi.org/10.1016/j.econlet.2017.01.012

Huang, W., Chew, S. H., & Zhao, X. (2020). Motivated False Memory. *Journal of Political Economy*. https://doi.org/10.1086/709971

Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(22), 6166–6171. https://doi.org/10.1073/pnas.1523586113

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Li, K. K. (2013). Asymmetric memory recall of positive and negative events in social interactions. *Experimental Economics*, *16*(3), 248–262. https://doi.org/10.1007/s10683-012-9325-9

Li, K. K. (2019). False Memory Preference. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3312183

Loewenstein, G., Issacharoff, S., Camerer, C., & Babcock, L. (1993). Self-Serving Assessments of Fairness and Pretrial Bargaining. *The Journal of Legal Studies*, *22*(1), 135–159. https://doi.org/10.1086/468160

Ma, Y., Li, S., Wang, C., Liu, Y., Li, W., Yan, X., … Han, S. (2016). Distinct oxytocin effects on belief updating in response to desirable and undesirable feedback. *Proceedings of the National Academy of Sciences*, *113*(33), 9256 LP – 9261. https://doi.org/10.1073/pnas.1604285113

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). Managing Self-Confidence: Theory and Experimental Evidence. *National Bureau of Economic Research Working Paper Series*, *No. 17014*. https://doi.org/10.3386/w17014

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. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.103.2.804

Rasmussen, H. N., Scheier, M. F., & Greenhouse, J. B. (2009). Optimism and Physical Health: A

Meta-analytic Review. *Annals of Behavioral Medicine*, *37*(3), 239–256. https://doi.org/10.1007/s12160-009-9111-x

Saucet, C., & Villeval, M. C. (2019). Motivated memory in dictator games. *Games and Economic Behavior*, *117*, 250–275. https://doi.org/https://doi.org/10.1016/j.geb.2019.05.011

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*. US: Lawrence Erlbaum Associates. https://doi.org/10.1037/0278-6133.4.3.219

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078. https://doi.org/10.1037//0022-3514.67.6.1063

Sedikides, C., & Green, J. D. (2004). What I Don't Recall Can't Hurt Me: Information Negativity Versus Information Inconsistency As Determinants of Memorial Self-defense. *Social Cognition*, *22*(1), 4–29. https://doi.org/10.1521/soco.22.1.4.30987

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*(11), 1475–1479. https://doi.org/10.1038/nn.2949

Shu, L. L., & Gino, F. (2012). Sweeping dishonesty under the rug: How unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology*. Shu, Lisa L.: Wyss House, Harvard Business School, Boston, MA, US, 02163, lisa.shu@post.harvard.edu: American Psychological Association. https://doi.org/10.1037/a0028381

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest Deed, Clear Conscience: When Cheating Leads to Moral Disengagement and Motivated Forgetting. *Personality and Social Psychology Bulletin*, *37*(3), 330–349. https://doi.org/10.1177/0146167211398138

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy*, *44*(6), 861–882. https://doi.org/https://doi.org/10.1016/j.brat.2005.07.001

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, *51*(2), 176–197. https://doi.org/https://doi.org/10.1016/0749-5978(92)90010-5

Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on cognitive dissonance.*. Oxford, England: Lawrence Erlbaum.

Zimmermann, F. (2020). The Dynamics of Motivated Beliefs. *American Economic Review*, *110*(2), 337–361. https://doi.org/10.1257/aer.20180728

Appendix I: Instructions for the control treatment

Thank you for participating in this experiment on Behavioral and Experimental Economics. You will be paid some money at its end; the precise amount will depend on chance and your decisions. All your decisions will be confidential, that is, the other participants will not get any information about your decisions, nor do you get any information about the others' decisions. In addition, your decisions will be anonymous: during the experiment, you will not have to enter your name at any time.

Decisions are made via the keyboard of your computer terminal. Read the on-screen instructions carefully before making any decision; there is no hurry to decide. These instructions meet the basic standards in Experimental Economics; in particular, all the information that appears in them is true and therefore there is no deception.

Please, do not talk to any other participant. If you do not follow this rule, we will have to exclude you from the experiment without payment. If you have questions, raise your hand and we will assist you. The use of calculators and writing tools is not permitted. Please, switch off your cell phone.

Description of the experiment

There is a 'virtual urn' with 100 balls. Each ball has written a different name of girl or boy; any of these names is used with a relatively high frequency in Spain. Let us call F the actual number of balls with a female name in your urn. You do not know either F or the number of balls in your urn with boy name (that is, 100 – F). You only know that the value of F has been randomly selected by the computer from among all integers between 0 and 100, both included (this means a total of 101 numbers, as 0 is included as well). Therefore, the probability that one of these potential values of F has been chosen is a priori of 1/101, that is, slightly less than 1%. Important: The value of F will not change throughout the experiment; the urn has always the same content.

During the experiment, the computer will perform several extractions from the urn, randomly and with replacement –in other words: each draw is reintroduced into the urn and can therefore be drawn in the next extraction. Each of the 100 balls has the same chance in each extraction. The computer will show you the name written in each extraction, one by one. Between some of the extractions, you will receive instructions to complete some questionnaire of perform some task.

Once you have completed all questionnaires and tasks, you will be paid in private and in cash. In this regard, you will receive 3 Euros for participating in the experiment, plus an additional payment that will depend of three 'prizes'. **Prize 1**: you receive 0.50 Euros for each ball in your urn with a girl

name. In other words, if there are F balls with female name in the urn, this prize equals 0.5 x F. **Prize 2** will be explained later, but will depend on one of the tasks to be performed. The same can be said about **prize 3**. Important: You can only win either prize 1 or prize 2. You do not know now which of them you will win; this will be determined randomly at the end of the experiment, choosing then one of the two prizes with a 50% probability. On the contrary, winning prize 3 is compatible with winning either prize 1 or 2. Observe finally that the prizes are always independent of each other. For example, what you win with prize 3 will not depend on how you have performed in the task corresponding to prize 2, and vice versa.

If you have any questions, please raise your hand and we will attend you.