UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

TESIS DOCTORAL

Motivated beliefs, self-serving recall and data omission: economic implications and experimental evidence

Adrián Caballero Castillejo

Supervised by

Ph.D. Carmen Arguedas Tomás

Ph.D. Raúl López Pérez

Madrid, 2021

Agradecimientos/Acknowledgments

En primer lugar, me gustaría agradecer a los organismos que han financiado mi formación e investigación durante estos años. Al Ministerio de Educación y Ciencia y la Universidad Autónoma de Madrid (UAM), a través el contrato predoctoral para Formación de Personal Investigador y las ayudas propias de la UAM para la realización de estancias de investigación en el extranjero. También al Ministerio de Economía, Industria y Competitividad, a través de los proyectos de investigación ECO2014-52372-P y ECO2017-82449-P.

A Carmen Arguedas y Raúl López, mis padres académicos, por su confianza, por guiarme y por enseñarme tanto a lo largo de esta etapa. Gracias por iluminar la ciencia lúgubre y hacerla mucho más humana.

No podría olvidar a ninguna de las personas que han contribuido a mi educación: desde Vicen y Lola, que me enseñaron a leer y escribir, hasta Rocío, con quien aprendí a llevar a cabo análisis econométricos de datos de panel. Por todo ello, mi agradecimiento a todos y cada uno de mis maestros y profesores en el Colegio Público El Parque, el Instituto de Educación Secundaria Las Lagunas, la Universidad Rey Juan Carlos y la Universidad Autónoma de Madrid. Y a la Educación Pública.

Gracias también a mis amigos, por quererme y aguantarme durante tantos años (y lo que te rondaré). Gracias Laura, Esther, Marina, Alberto, Aitana, Rubén P., Rubén R., Diana, Silvia, Habiba, Noelia, Toni, Franco y Nico.

Por último, el agradecimiento más especial no puede ser sino para mi familia. A Felipe, quien lleva acompañándome desde que inicié esta Tesis y ojalá lo siga haciendo por muchos años más. A mis abuelos, y particularmente a mi abuelo Mariano, que siempre quiso ser maestro y lo habría sido si la guerra no lo hubiese obligado a crecer antes de

tiempo. A mi hermana Ana, a quien puedes encontrarte salvando vidas en los hospitales de Madrid al ritmo del último musical de Broadway. Y, por supuesto, a mis queridísimos padres, Mariana y Vicente, sin quienes nunca podría haber llegado hasta aquí. A ellos, que siempre han creído y lo han dado todo por sus hijos.

Acknowledgements

I wish to thank Kai Barron, Alexander Coutts, Hubert Kiss, Guy Mayraz, Juan César Palomino, Jonathan Parker, Eil Spiegelman, and participants at seminars at UAM, CSIC and TSE for helpful comments and suggestions. I am also grateful to the Department of Economics at the Toulouse School of Economics, for their hospitality during my visit in 2018.

Contents

Introduction

It has been generally assumed that economic agents integrate all the available information to form beliefs about an uncertain world. Moreover, they aggregate each piece of evidence in a statistically consistent manner, following Bayes' rule, to get a depiction of the environment as realistic and accurate as possible. As of today, this sense of rationality enjoys such a good health that many economic models assume it even without stating it explicitly. For the last decades, however, an increasing amount of evidence from Psychology and Behavioral Economics has been gathered that shows that what is true for *homo economicus* is not necessarily true for *homo sapiens*. Cognitive limitations like bounded memory or attention can constrain the amount and the type of information that we perceive and store in our minds. For example, more recent or salient information may be better recalled than older, inconspicuous one, resulting in an incomplete -and probably biased- sample. Cognitive limitations can also condition how we process and integrate the available evidence to form beliefs. Thus, we may rely on relatively simple heuristic strategies rather than more cognitively demanding statistical computations (Tversky and Kahneman, 1974). Furthermore, just as people prefer one good over another, they may also have preferences over different beliefs. Some beliefs may provide us joy or comfort even when they are ill-founded -I will live a prosperous and long live, I am a good professional, I will get large returns from my last investment, etc.-. On the contrary, ominous thoughts may threaten our welfare and self-esteem. Consequently, people may be prone to look for or to recall better the information that reinforces their preferred beliefs, while avoiding, forgetting or underweighting the negative evidence (Bénabou and Tirole, 2016).

This Thesis is composed of four essays that provide some economic models and experimental evidence to answer several questions regarding the determinants and the consequences of data omission and biased beliefs due to cognitive limitations and preferences over beliefs. While related, each chapter can be read independently. A summary of each chapter is presented below.

In Chapter 1, we explore the consequences of self-deception in strategic settings where the consumers' behavior may entail uncertain risks for themselves or others. This includes plenty of real-life situations. For example, most consumers are not fully aware of the specific impact of recycling plastic, flying from Madrid to Boston, or donating to a certain NGO. We present a new model in which consumers (i) may experience anticipatory utility (Caplin and Leahy, 2001; Kőszegi, 2010; Loewenstein, 1987) -i.e. their current utility may depend on their expectations about their future utility- and (ii) can choose their beliefs regarding the risks associated to a hazardous product. In contrast to the existing literature on motivated reasoning and self-deception (see, for example, Akerlof and Dickens, 1982; Bénabou, 2013; Brunnermeier and Parker, 2005), we focus on strategic settings and the relations of interdependence both between the consumers and between these and the firm(s). Thus, not only prices determine optimal beliefs and choices, but consumer's preferences over beliefs also alter the firms' incentives. Consumers, for example, may alleviate their anxiety by underestimating the risk associated to the hazardous product. This, however, has consequences for the market equilibrium, since optimistic consumers are more willing to pay for that product than consumers with more realistic beliefs, other things equal. In this sense, our model provides some interesting results: when self-deception is not possible, for example, those individuals that feel more anxious or excited about their future are relatively more prone to take action in the present in order to improve their future situation. In contrast, when individuals can choose their beliefs, anticipatory utility may trigger self-deception rather than action. Also, our model predicts that self-deception is more likely to arise when there is uncertainty about the externalities associated to a product (or action), when individual behavior has a negligible impact on the economy, or when the potential consequences will realize only in a distant future. This seems particularly relevant, for example, regarding environmental problems. Interestingly, however, the existence of individuals that underestimate the risks associated to the hazardous product does not necessarily lead to a larger aggregated consumption of it in the economy. The specific result will depend on the characteristics of the economy, including the intensity of the anticipatory utility among individuals. Overall, we believe that the model presented in this chapter provides some new insights that may prove useful to deal with a large range of phenomena, including belief polarization, social responsibility or climate change denial.

In Chapter 2 we present an experimental design to test the implications of some models of motivated beliefs and optimism. All these models assume that individuals have preferences over the possible states of nature and they derive utility directly from keeping beliefs in line with these preferences (for example, in the form of anticipated utility). On the other hand, forming and keeping unrealistic beliefs may entail some costs. We focus on two families of theories that differ mainly in the factors that prevent individuals' beliefs from departing too much from reality. According to some models (e.g. Akerlof and Dickens, 1982; Brunnermeier and Parker, 2005), optimism is more likely when the material costs associated to biased beliefs are relatively low. On the contrary, individuals may be more reflective and form more accurate beliefs in situations where biases lead to relatively costly or hazardous decisions. Alternatively, models by Rabin (1994) and Bracha and Brown (2012) stress the fact that self-deception is costly also in the sense that the individuals must selectively look for favorable information, while avoiding unfavorable evidence or restraining challenging thoughts. Remarkably, while there is abundant evidence of situations in which people seem optimistic, tests about the specific implications of these models are much less common, and far from conclusive. On account of this, we contribute to the existing literature with a novel experimental design. Succinctly, the experiment consists of an estimation task in which each participant faces a virtual urn with 100 balls, each containing a boy or girl name. After observing a series of random draws, the participant must estimate the actual share of female balls in the urn. Importantly, monetary incentives are designed to induce a preference for that proportion to be as large as possible. Overall, the results of our experiment find scarce support for these models, although models by Rabin (1994) and Bracha and Brown (2012) fit our data relatively better. Still, we do not find systematic optimism among our participants. Moreover, the sign of the bias seems more related to the specific characteristic of the observed sample rather than individual characteristics or material incentives.

The estimation task presented above was followed by a memory task, in which the participants are incentivized to recall as many boy and girl names in their urn as possible. Using data from both tasks, we explore in Chapter 3 the connection between biased recall and optimism. In the last years, it has been suggested that the optimistic bias may be caused by self-serving recall, i.e. people often recall favorable information better than negative evidence (Epley and Gilovich, 2016; Bénabou and Tirole, 2016). Based on this literature, we provide a model of inference with self-serving recall and test experimentally its main predictions. In a nutshell, our model is based in two ideas: (i) people extrapolate from the evidence they recall; and (ii) people are more likely to recall information that is favorable given their preferences, i.e. when it supports their preferred states of nature. Our results provide some support for the latter: in our experiment, the participants are more likely to recall girl names -which are associated to a larger payoff-. In addition, they do not seem aware of this bias: in fact, they expect to recall bad news (i.e. male balls) relatively better than good news. Nevertheless, several results suggest that self-serving recall does not induce optimism in our experimental design. First, our subjects do not systematically overestimate the proportion of female balls. Further, we find no correlation between optimism and biased recall: people who provide inflated estimations are not relatively more likely to recall girl names better. In this sense, the participants' estimations are better fitted by the Bayesian model than by a model assuming that people track the proportion of female balls in the recalled sample. Overall, our results suggest that the link between memory tasks and estimation tasks is not straightforward, and that people may infer from a different sample than the recalled one. We consider this result to be particularly relevant for experimental research, where these tasks are widely used.

Finally, Chapter 4 explores other sources of bias regarding belief formation. Specifically, we focus on data omission in contexts where there is no preference for any state of the world. To that end, we propose both an analytical framework and a lab experiment in which participants face a quite simple problem of inference. Succinctly, each participant observes a series of random draws with replacement from one urn containing red and blue balls. The subject knows that the specific rate of red balls her urn is randomly determined with uniform probability from a set of three rates. Based on the evidence observed, the participant is asked to estimate the true rate in an incentive-compatible manner. In contrast to the experimental design from Chapters 2 and 3, the participant's payoff depends exclusively on the accuracy of her estimate and not on the rate per se. Based on the experimental evidence, we explore some relevant questions regarding data omission. First, we consider heterogeneity among the participants. For example, some individuals may omit more data than others when elaborating their estimates. On this matter we find that, while the estimates of most participants fit the

Bayesian model relatively well, a non-negligible portion of them seem to rely on very small subsamples. We find this quite striking given the simplicity of the problem. Further, our evidence suggests that differences in data omissions are more likely to do with differences in attention than memory constraints. Finally, we find experimental support for the hypothesis that experience and incentives can alleviate data omission in our experimental design. Yet, the extent of the improvement seems modulated by the complexity of the problem.

Introducción (Spanish)

Ha sido generalmente asumido que los agentes integran toda la información disponible para formar creencias sobre un mundo incierto, y que la evidencia es agregada de manera estadísticamente consistente, siguiendo la regla de Bayes, para obtener una representación del entorno lo más realista y precisa posible. Incluso hoy, la racionalidad entendida de este modo goza de buena salud, y son muchos los modelos que la asumen, incluso sin declararlo explícitamente. Durante las últimas décadas, sin embargo, la evidencia recogida desde Psicología y la Economía del Comportamiento ha revelado que, lo que es cierto para el *homo economicus*, no lo es necesariamente para el *homo sapiens*. Las limitaciones cognitivas, como la memoria y la atención limitadas, pueden restringir la cantidad y el tipo de información que podemos percibir y almacenar en nuestra mente. Por ejemplo, la información más reciente o destacada puede ser mejor recordada que aquella más antigua o que no llama tanto la atención, dando como resultado una muestra subjetiva incompleta -y probablemente sesgada-. Las limitaciones cognitivas también pueden condicionar la forma en la que procesamos e integramos la evidencia disponible para formar creencias. Así, en ocasiones recurrimos a estrategias heurísticas relativamente sencillas en lugar de a cálculos estadísticos, cognitivamente mucho más laboriosos (Tversky y Kahneman, 1974). Por otra parte, igual que las personas prefieren un bien sobre otro, también pueden tener preferencias por determinadas creencias. Algunas creencias pueden hacernos sentir más felices o confortables, incluso aunque puedan ser infundadas -voy a tener una vida larga y próspera, soy un buen profesional, mi última inversión va a generar grandes beneficios, etc.-. Por el contrario, los pensamientos más agoreros o amenazantes pueden poner en riesgo nuestro autoestima y bienestar. En consecuencia, es posible que las personas tiendan a buscar o a recordar mejor aquella información que refuerza sus creencias preferidas, evitando, olvidando o infravalorando la evidencia desfavorable (Bénabou y Tirole, 2016).

Esta Tesis doctoral se compone de cuatro ensayos que aportan nuevos modelos económicos y evidencia experimental, con los que se trata de dar respuesta a algunos interrogantes relacionados con las causas y consecuencias de la omisión de datos y creencias sesgadas debido a limitaciones cognitivas y a la preferencia por ciertos pensamientos. Aunque están relacionados, cada capítulo puede leerse de manera independiente. El resumen del contenido de cada capítulo se presenta a continuación.

En el Capítulo 1, exploramos las consecuencias del autoengaño en contextos estratégicos donde el comportamiento de los consumidores puede conllevar riesgos para ellos mismos u otros. Esto incluye un sinfín de situaciones reales. Por ejemplo, la mayoría de los consumidores no son plenamente conscientes del impacto real de reciclar plástico, volar de Madrid a Boston o donar a una determinada ONG. En este sentido, presentamos un nuevo modelo en el que los consumidores (i) pueden experimentar utilidad anticipada (Caplin and Leahy, 2001; Kőszegi, 2010; Loewenstein, 1987) -esto es, la utilidad actual puede depender de las expectativas sobre la utilidad futura- y (ii) pueden elegir sus creencias con respecto a los riesgos asociados a su consumo. A diferencia de la literatura existente en materia de razonamiento motivado y autoengaño (véase, por ejemplo, Akerlof y Dickens, 1982; Bénabou, 2013; Brunnermeier y Parker, 2005), nuestro trabajo se centra en marcos estratégicos y en las relaciones de interdependencia tanto entre los propios consumidores, como entre estos y la(s) empresa(s). En este sentido, no solo los precios determinan las creencias óptimas y las decisiones de consumo, sino que las preferencias de los consumidores en relación a sus creencias también alteran los incentivos de las empresas. Los consumidores, por ejemplo, podrían reducir su ansiedad infravalorando los riesgos asociados a productos potencialmente dañinos. Esto tiene consecuencias con respecto al equilibrio de mercado ya que, en igualdad de condiciones, los individuos más optimistas estarán dispuestos a pagar más por esos productos que aquellos con creencias más realistas. En este sentido, de nuestro modelo se obtienen varios resultados de interés: si, por ejemplo, el autoengaño no es posible, aquellos individuos más propensos a sentirse ansiosos o excitados con respecto al futuro son más propensos a llevar a cabo acciones en el presente para mejorar su situación futura. Por el contrario, si los individuos pueden elegir sus creencias, la existencia de utilidad anticipada puede llevar al autoengaño, en lugar de a actuar. Por otro lado, nuestro modelo también predice que el autoengaño es tanto más probable cuanto mayor sea la incertidumbre con respecto a las externalidades asociadas a un producto o acción, cuando el comportamiento individual tiene un impacto insignificante en la economía, o cuando las consecuencias potenciales se materializan en un futuro lejano. Estas condiciones parecen particularmente relevantes, por ejemplo, en relación con los problemas ambientales. Curiosamente, sin embargo, la existencia de individuos que infravaloran los riesgos asociados a un producto potencialmente dañino no lleva necesariamente a un mayor consumo agregado del mismo. El resultado concreto dependerá de las características específicas de la economía, incluyendo la intensidad de la utilidad anticipada entre los consumidores. En general, consideramos que el modelo presentado en este capítulo aporta conocimientos que pueden ser útiles para enfrentarse a multitud de fenómenos, incluyendo la polarización ideológica, la responsabilidad social o el negacionismo del cambio climático.

En el Capítulo 2, presentamos un diseño experimental con el fin de contrastar las implicaciones de algunos modelos de razonamiento motivado y optimismo. Todos estos modelos asumen que los individuos tienen preferencias sobre los distintos estados de la naturaleza y que obtienen utilidad al mantener unas creencias en línea con esas preferencias (por ejemplo, en forma de utilidad anticipada). Por otro lado, no obstante, la formación y mantenimiento de creencias poco realistas puede entrañar ciertos costes. Nosotros nos centramos fundamentalmente en dos familias de teorías que difieren principalmente en los factores que evitan que las creencias individuales se alejen demasiado de la realidad. De acuerdo con algunos modelos (p.ej. Akerlof y Dickens, 1982; Brunnermeier y Parker, 2005), el optimismo es más factible cuando los costes materiales asociados a las creencias sesgadas son relativamente bajos. Por el contrario, si los sesgos pueden llevar a tomar decisiones muy costosas o peligrosas, es más probable que los individuos sean más reflexivos y formen creencias más precisas. Alternativamente, modelos como los propuestos por Rabin (1994) y Bracha y Brown (2012) hacen hincapié en que el autoengaño es costoso también en el sentido de que los individuos tienen que hacer una búsqueda selectiva de la información más favorable, evitando a la vez la evidencia negativa o reprimiendo aquellos pensamientos que puedan cuestionar sus creencias preferidas. Sorprendentemente, aunque existe evidencia de numerosas situaciones en las que la gente parece optimista, son mucho menos frecuentes los tests sobre las implicaciones específicas de estos modelos, y sus resultados distan de ser concluyentes. Por ello, este capítulo pretende contribuir a esta literatura con un nuevo diseño experimental. De manera sucinta, el experimento consiste en una tarea de estimación en la que cada participante se enfrenta a una urna virtual con 100 bolas, conteniendo cada una de ellas un nombre de chico o chica. Tras observar una serie de extracciones aleatorias, el participante debe estimar el porcentaje real de bolas femeninas en la urna. Es importante señalar que los incentivos monetarios han sido diseñados de tal forma que los individuos prefieran que este porcentaje sea lo más alto posible. En conjunto, los resultados de nuestro experimento encuentran un escaso respaldo de estos modelos, si bien modelos como el de Rabin (1994) o Bracha and Brown (2012) se ajustan relativamente mejor a los datos obtenidos. En todo caso, no encontramos un optimismo sistemático entre nuestros participantes, y el sentido del sesgo parece más relacionado con las características de la muestra observada que con las características individuales o los incentivos materiales.

La tarea de estimación presentada en el párrafo anterior fue seguida de una prueba de memoria, en la que los participantes fueron incentivados a recordar tantos nombres de chico y de chica en su urna como fuera posible. Utilizando los datos de ambas tareas, en el Capítulo 3 exploramos la conexión entre el recuerdo sesgado y el optimismo. En los últimos años, ha sido sugerido que el sesgo optimista podría ser causado por el recuerdo interesado, esto es, que la gente suele recordar la información favorable mejor que la evidencia negativa (Epley y Gilovich, 2016; Bénabou y Tirole, 2016). Basándonos en esta literatura, en este capítulo proponemos un modelo de inferencia con recuerdo interesado y contrastamos experimentalmente sus principales predicciones. De forma resumida, nuestro modelo se fundamenta en dos ideas: (i) las personas extrapolan a partir de la evidencia que recuerdan; y (ii) las personas tienden a recordar mejor la información que les es favorable dadas sus preferencias. Nuestros resultados suponen un cierto respaldo de esta última: en nuestro experimento, los participantes tienden a recordar mejor los nombres de chica -que están asociados a un pago mayor-. Además, no parecen conscientes de este sesgo: al contrario, en general esperan recordar las malas noticias (en nuestro caso, las bolas masculinas) relativamente mejor que las buenas. Con todo, varios resultados sugieren que el recuerdo interesado no se traduce en optimismo en nuestro diseño experimental. En primer lugar, los sujetos no sobreestiman de manera sistemática la proporción de bolas femeninas. Más aun, no hallamos una correlación entre optimismo y recuerdo sesgado: los individuos que dan estimaciones optimistas no tienden a recordar las bolas femeninas relativamente mejor. En este sentido, las estimaciones de los participantes se ajustan mejor al modelo bayesiano que el modelo alternativo en el que los individuos se basan en la muestra recordada. En conjunto, nuestros resultados sugieren que la conexión entre las tareas de memoria y de estimación no es directa, y que las personas podrían basar sus estimaciones en una muestra distinta de la recordada. Consideramos que este resultado es particularmente relevante en el ámbito de la investigación experimental, donde ambos tipos de prueba son ampliamente utilizados.

Finalmente, el Capítulo 4 explora otras fuentes de sesgo en la formación de creencias. Específicamente, nos centramos en la omisión de datos en situaciones en las que no hay preferencias por ningún estado del mundo en particular. Con este fin, proponemos tanto un marco analítico como un experimento de laboratorio en el que los participantes se enfrentan a un problema sencillo de inferencia. Sucintamente, cada participante observa una serie de extracciones aleatorias con reemplazo de una urna que contiene bolas azules y rojas. El sujeto sabe que la proporción de bolas rojas en su urna ha sido determinada aleatoriamente con probabilidad uniforme entre un conjunto de tres posibles tasas. A partir de la evidencia observada y de manera incentivo-compatible, se le pide a cada participante que estime la verdadera tasa. A diferencia del diseño experimental utilizado en los Capítulos 2 y 3, el pago de cada participante depende única y exclusivamente de la precisión de su estimación y no de la tasa en sí misma. A partir de los datos obtenidos, exploramos algunas cuestiones relevantes sobre omisión de datos. En primer lugar, consideramos la heterogeneidad entre los participantes. Por ejemplo, algunos individuos podrían omitir más datos que otros al realizar sus estimaciones. En este sentido encontramos que, si bien la mayoría de los participantes se ajustan relativamente bien al modelo bayesiano, una parte considerable de ellos parece utilizar submuestras muy pequeñas, lo que no deja de ser llamativo dada la sencillez del problema. Además, la evidencia experimental sugiere que las diferencias en la omisión de datos tienen posiblemente más que ver con diferencias en el nivel de atención que en limitaciones de memoria. Por último, nuestros resultados refuerzan la hipótesis de que la experiencia y los incentivos pueden reducir la omisión de datos en nuestro diseño experimental, si bien el alcance de la mejora parece modulado por la complejidad del problema.

References:

Akerlof, G. A., & Dickens, W. T. (1982). The Economic Consequences of Cognitive Dissonance. *The American Economic Review*, *72*(3), 307–319.

Bénabou, R. (2013). Groupthink: Collective Delusions in Organizations and Markets. *The Review of Economic Studies*, *80*(2), 429–462.Bénabou, R., & Tirole, J. (2016). Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives*, *30*(3), 141–164.

Bracha, A., & Brown, D. J. (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, *75*(1), 67–80.

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, *95*(4), 1092–1118.

Caplin, A., & Leahy, J. (2001). Psychological Expected Utility Theory and Anticipatory Feelings. *The Quarterly Journal of Economics*, *116*(1), 55–79.

Epley, N., & Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives*, *30*(3), 133–140.

Kőszegi, B. (2010). Utility from anticipation and personal equilibrium. *Economic Theory*, *44*(3), 415–444.

Loewenstein, G. (1987). Anticipation and the Valuation of Delayed Consumption. *The Economic Journal*, *97*(387), 666.

Rabin, M. (1994). Cognitive dissonance and social change. *Journal of Economic Behavior & Organization*, *23*(2), 177–194.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Chapter 1

Economic decisions with uncertain consequences, self-deception, and strategic pricing

1. Introduction

In this paper, we explore the consequences of motivated reasoning and self-deception in situations where individuals' actions may generate uncertain externalities or when the future utility of public or private goods depends on some contingencies that are unknown in advance. For example, individuals may decide whether to buy ecologic products or to donate to an NGO without being fully aware of the real impact of their actions on society. Also, individuals may decide to consume multivitamins although their effects -if any- can be uncertain and may be observable only in the future.

Classical models depict human behavior as purely rational in the sense that individuals maximize their own expected utility and use all the available information to form realistic beliefs using Bayes' rule (Mas-Colell et al., 1995). In these models, information and expectations play an instrumental role. They do not provide utility by themselves, but accurate information and realistic beliefs allow individuals to make better decisions and therefore to improve their future welfare. This result, however, depends on certain assumptions. For example, it is generally assumed that the utility flow at a certain time does not depend on the expectations about future utility flows.

In this paper, we take an alternative approach by assuming that individuals may experience anticipatory utility (Caplin & Leahy, 2001; Kőszegi, 2010; Loewenstein, 1987) so that their expectations about their future utility have a direct impact on their current utility -e.g. they feel anxious about the perspective of a gloomy future-. In that case, we show that the formation and maintenance of realistic beliefs could be suboptimal. Instead, they may prefer keeping more optimistic and comforting beliefs, even if they are at odds with the available information. While previous works have already acknowledged the consequences of anticipatory utility regarding belief distortion and individual decision making, we focus on how this can alter the incentives, prices and profits in the market.

Up to our knowledge, Akerlof & Dickens (1982) is one of the first studies to deal with belief distortion due to anticipatory utility. They consider the problem of an agent that decides whether to work in a safe industry or in a hazardous industry which pays larger wages but where there is a risk of suffering an accident. Workers in the hazardous industry experience a sort of anticipatory utility or fear which depends on her beliefs about the probability of suffering an accident. Further, these beliefs are assumed to be chosen without constraints by the worker before deciding whether to work in the safe or in the hazardous industry. By keeping optimistic beliefs, i.e. by underestimating the probability of suffering an accident, the worker may alleviate her fears and earn a larger wage in the hazardous industry, but at the cost of facing the potential loss associated to the accident. A more general model of optimal expectations in the presence of anticipatory utility can be found in Brunnermeier and Parker (2005). In both cases, optimal beliefs trade-off the present benefits of being optimistic about the future and the future costs of having distorted beliefs in terms of worse decision making. Other models consider explicitly the possibility that individuals update new information asymmetrically -for example, by avoiding or forgetting bad news or, more generally, by weighting favorable evidence relatively more than challenging information, focusing on the dynamics of information processing (Bénabou, 2013; Benabou and Tirole, 2002; Möbius et al. 2014).

Instead, we focus on the effects of anticipatory utility and the optimality of belief distortion in strategic settings. Specifically, we consider a consumption problem in which consumers are uncertain about the future consequences of their current choices and they experience both material and anticipatory utility (Akerlof and Dickens, 1982; Bénabou, 2013; Brunnermeier and Parker, 2005). Thus, the demand for a specific product will depend simultaneously on prices, preferences over goods and preferences over beliefs. Further, these factors are interrelated and condition each other. It is clear from the mentioned models that prices can alter not only consumers' behavior but also their optimal beliefs. To our knowledge, however, the opposite question -how belief distortion may alter the firms' incentives and pricing decisions- has received no attention so far. In this paper, we analyze the incentives faced by a monopolistic firm and the interaction with a mass of consumers with preferences over beliefs. Further, we consider the possibility that these preferences are not homogeneous among consumers, so that different consumers may hold different beliefs about the consequences of the same action.

In our model, a set of consumers are uncertain about the negative consequences associated to the consumption of a product. If a consumer experiences anticipated utility, we show that she may prefer keeping biased beliefs that underestimate the risk, even if she herself does not consume the hazardous product. Further, we show that if individuals are heterogeneous in their preferences over goods, they may form different beliefs even if they have the same information. This has important implications regarding the market equilibrium since, other things equal, optimistic consumers are more willing to pay for the risky product. At the same time, prices can alter consumers' preferences over beliefs. Under certain circumstances, we show that some actors -a monopoly, in our model- may be able to increase their profits by taking advantage of consumers' preference for biased beliefs.

Following Akerlof and Dickens (1982) and Brunnermeier and Parker (2005), we assume that consumers can choose their preferred beliefs with no restraints or cognitive cost. This is certainly an extreme assumption, and other models soften it by introducing a direct cost of belief distortion that captures the psychic cost of avoiding unfavorable information or repressing challenging memories (see, for example, Benabou and Tirole (2002) or Rabin (1994)). However, its implications are quite limited and do not change qualitatively the main results.

Finally, we note that anticipatory utility is not the only factor that may justify the optimality of keeping unrealistic beliefs. Köszegi (2006) and Möbius et al. (2014) assume that the agent has some *ego utility* which depends on her beliefs about her own abilities. In Rabin (1994), it is beliefs about the morality of one's behavior which provides well-being or discomfort. In all these cases, beliefs (about the future, one's abilities or one's morality) provide utility by themselves. Further, Bénabou and Tirole (2002) show that time-inconsistency can also drive the individual to prefer biased, optimistic expectations.

Altogether, these works aim to provide models of behavior consistent with a large amount of literature from Psychology and Economics that suggests that there is plenty of situations in which individuals form and keep unrealistic beliefs that systematically depart from those that should be expected from a rational agent given the available evidence. When comparing themselves to others, individuals generally overestimate their driving skills (Svenson, 1981), teachers are too optimistic about their teaching abilities (Cross, 1977) and CEOs systematically overestimate their ability to generate returns (Malmendier and Tate, 2005).

Individuals also like to believe that they are moral, and they experience a sort of *ethical dissonance* when their moral values and their own behavior are in conflict (Barkan et al., 2012). Some studies show that this inner tension is sometimes alleviated through self-serving interpretations of the consequences of their behavior, avoiding information that could challenge these beliefs (Dana et al., 2007; Haisley and Weber, 2010). In other cases, they may distort their beliefs regarding other's behavior and morality (Barkan et al., 2012; Di Tella et al., 2015). Another example of ethical dissonance that has attracted some attention in the last years is the so-called "meat paradox" (Loughnan et al., 2010): while most people eat meat, they generally feel discomfort with the idea of animal suffering. One way in which meat consumers solve this tension seems to be precisely the upholding of self-serving beliefs about animals' mental capacities and feelings (Bastian et al., 2012; Hestermann et al., 2019; Loughnan et al., 2010).

Finally, individuals are generally too optimistic about their future. We underestimate the probability of getting divorced, losing our job or having a heart attack, while we overestimate the probability of having a long life or having a good salary (Weinstein, 1989). The unemployed overestimate how fast they will find a new job (Spinnewijn, 2015). Smokers underestimate their risk of developing lung cancer relative to both non-smokers and to other smokers (Weinstein et al., 2005). Far from anecdotal, the estimates suggest that around 80 percent of people show a certain degree of optimism bias (Sharot, 2011). Although there is evidence that optimism is associated to a good mental and physical health (Rasmussen et al., 2009; Strunk et al., 2006), it entails some risks, since inaccurate beliefs may lead to suboptimal decision making. Among the economic consequences of an excessive optimism we find, for example, the formation of speculative bubbles (Lamont and Thaler, 2003; Minsky, 1974; Shiller, 2000), poor investment decisions by CEOs (Malmendier and Tate, 2005), or an underprovision of insurance against health problems or unemployment (Spinnewijn, 2015).

The rest of the paper is structured as follows. In Section 2 we introduce our benchmark model in which self-deception is not possible and beliefs are determined exogenously. Then, Section 3 extends this model to allow for the possibility that consumers can choose their beliefs. In Section 4, we compare the benchmark and the extended models and analyze the main differences between the two. We conclude in Section 5. All the technical details as well as several extensions of the model can be found in the Appendix.

2. A benchmark model with exogenous beliefs

In this section, we consider an individual that experiences anticipatory utility, and can choose among two varieties of a consumption good manufactured by a monopolistic firm. These varieties may have different private and public effects that are realized only after consumption. Prior to purchase, the consumer has only some beliefs about the probability distribution of these effects. Throughout this section, we assume that these beliefs coincide with the objective probability distribution and that the consumer cannot strategically self-deceive or manipulate her beliefs.

For presentation purposes, we first consider in Subsection 3.1. the behavior of one individual. We then extend our analysis in Subsection 3.2 to a mass of consumers which differ in the intensity of their anticipatory utility. Finally, in Subsection 3.3 we consider the pricing decision faced by a monopolistic firm that produces the two varieties considered in our model, and the resulting market equilibrium.

2.1. Consumer's behavior

In this subsection, we consider an individual i from a mass of consumers in the interval $[0,1]$. There are two periods $t = \{1,2\}$. At time $t = 1$, the consumer chooses whether she buys one unit of a good. We denote by $\emptyset$ the absence of consumption. There are two varieties of that good, namely a regular version (R) and a superior one (S). Both varieties provide the same consumption utility $u_R = u_S = 1$ at $t = 1$ (we assume that $u_\emptyset = 0$) and differ only in that the regular version produces a damage $\theta = \bar{\theta} > 0$ at $t = 2$ with objective probability $x \in (0,1)$ or no damage $\theta = 0$ with complementary probability $1 - x$. The expected unitary damage is therefore $E(\theta) = x\bar{\theta}$.

Thus, at time $t = 2$, consumer's utility from the consumption of $j = \{\emptyset, R, S\}$ is

$$U_{j,2} = \begin{cases} -\beta\theta\Theta_R & if\ j = \{\emptyset, S\} \\ -\alpha\theta - \beta\theta\Theta_R & if\ j = R \end{cases} \tag{1}$$

where $\alpha, \beta \geq 0$ and $\Theta_R \in [0,1]$ is the subset of consumers choosing the regular product. The coefficient α can be interpreted as guilt, i.e. the consumer feels bad from contributing to a public bad. More generally, it is any private negative effect on consumer's utility derived from the consumption of R, if $\alpha > 0$. The coefficient β captures the public nature of the damage θ. If $\beta > 0$, then θ has a public good nature -more precisely, a public bad nature- whose effect on the consumer's utility depends on the specific value of β, θ and on the total amount of the regular product consumed in the economy (Θ_R).

Further, at time $t = 1$, the consumer may experience anxiety or, more generally, *anticipatory utility*, i.e. even before any future effect from consumption is realized, the consumer feels utility or disutility -anxiety, sadness, excitement etc.- from thinking on her future welfare (Akerlof and Dickens, 1982; Bénabou, 2015; Benabou and Tirole, 2002). The temporal structure of the consumer's problem is summarized in Figure 1.

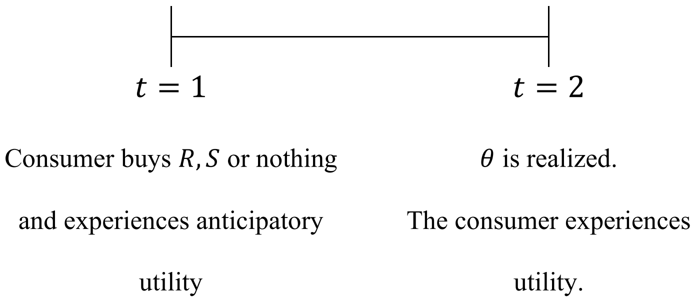


FIGURE 1. Timing with exogenous beliefs

Assuming that anticipatory utility is proportional to the expected utility at $t = 2$, the consumer's expected utility at $t = 1$ is

$$\hat{E}U_{j,1} = \begin{cases} -(a+\delta)\hat{E}[\beta\theta\Theta_R] = -(a+\delta)\beta\Theta_R\hat{x}\bar{\theta} & if\ j = \emptyset \\ 1 - p_S - (a+\delta)\hat{E}[\beta\theta\Theta_R] = 1 - p_S - (a+\delta)\beta\Theta_R\hat{x}\bar{\theta} & if\ j = S \\ 1 - p_R - (a+\delta)\hat{E}[\alpha\theta + \beta\theta\Theta_R] = 1 - p_R - (a+\delta)(\alpha + \beta\Theta_R)\hat{x}\bar{\theta} & if\ j = R \end{cases} \quad (2)$$

where u_j and p_j are the consumption utility and the price of the variety j, respectively; $a \geq 0$ is the weight associated to anticipatory utility, $\delta \in (0,1)$ is the discount factor[1] and $\hat{x} \in [0,1]$ stands for the individual subjective probability assigned to the event $\theta = \bar{\theta}$.[2] Along this Section, we assume that $\hat{x}$ is fixed and equal to the objective probability associated to the event $\theta = \bar{\theta}$, so that $\hat{x} = x$.

The following result shows the consumer's optimal choice under exogenous beliefs (see the proof in the Appendix):

[1] Alternatively, the discount factor δ can also be interpreted in terms of how far in time is period 2. Other things being equal, the further the time at which θ, the smaller the value of δ.

[2] An important remark must be made here. While the anticipatory utility component and the expected utility at $t = 2$ are mathematically equivalent, they represent quite different concepts. At $t = 1$, the individual experiences a certain anticipatory utility equal to $-a(\alpha e_j + \beta\Theta_R)\hat{x}\bar{\theta}$, which directly depends on her beliefs ($\hat{x}$). On the other hand, at $t = 2$ the individual will experience an uncertain utility of 0 or $-(\alpha e_j + \beta\Theta_R)\bar{\theta}$ depending on whether $\theta = 0$ or $\theta = \bar{\theta}$, respectively. Although this difference lacks importance for the results of this section, it will be critical in Section 4.

Proposition 1. Consumer's choice. For given prices p_S and p_R, beliefs $\hat{x}$ and parameters a, α, δ and $\bar{\theta}$, there are four possible scenarios:

I. If $p_R \leq 1-(a+\delta)\alpha\hat{x}\bar{\theta}$ and $p_S \leq 1$, the consumer buys the superior variety S if $\hat{x} \geq \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}}$ and the regular variety R otherwise.

II. If $p_R > 1-(a+\delta)\alpha\hat{x}\bar{\theta}$ and $p_S \leq 1$, the consumer buys S.

III. If $p_R \leq 1-(a+\delta)\alpha\hat{x}\bar{\theta}$ and $p_S > 1$, the consumer buys R.

IV. If $p_R > 1-(a+\delta)\alpha\hat{x}\bar{\theta}$ and $p_S > 1$, the consumer does not buy anything.

Intuitively, if the price of both varieties is low enough, the consumer will buy the superior variety S if she expects that the premium payed for that variety is worth given the reduction of future and anticipatory disutility. Otherwise, she chooses R (Scenario I). On the other hand, if one of the varieties is relatively expensive while the other is affordable, the consumer will buy the latter (Scenarios II and III). Finally, if both varieties are relatively expensive, the consumer will refrain from buying any of the two (Scenario IV). The precise definition of what is an affordable or an expensive price depends on the specific parameters. From Scenarios I-IV we notice that the maximum willingness to pay for the superior variety is smaller or equal to 1. This is just because this variety provides a consumption utility of $u_S = 1$ and it does not entail any potential harm. On the other hand, the consumer's maximum willingness to pay for the regular variety is never larger than $1-(a+\delta)\alpha\hat{x}\bar{\theta}$, since, in addition to a consumption utility $u_R = 1$, the regular variety entails an anticipatory disutility equal to $a\alpha\hat{x}\bar{\theta}$ and an expected future disutility equal to $\delta\alpha\hat{x}\bar{\theta}$. We constrain our analysis to cases where $\delta\alpha\hat{x}\bar{\theta} < 1$, since no consumer would ever buy the regular variety otherwise. Since in our model the individual consumer is too small relative to the whole economy, her sole actions cannot affect the total provision of the public bad, which is given by Θ_R. In consequence, disutility from the provision of the public bad does not play any role in the consumer's choice.

From the above results, we can derive two main conclusions. First, we find that public effects and others' choices are completely irrelevant for the consumer. As long as $\hat{x}$ is exogenous, the consumer's choice at $t = 1$ depends only on her guilt -or, more generally, on the private part of her expected utility-. It is straightforward from Proposition 1 that it does not depend on either β or the other consumers' choices, represented by Θ_R. Prosocial behavior (avoidance of contributing to the provision of the public bad) is triggered solely by private motives (α) including impure altruism[3].

Second, the discount rate (δ) and the anticipatory utility weight (a) affect individual behavior in the same way. Consumers who are more patient -weight the future more- and those who feel more anxious or that experience anticipatory utility from future consequences more intensely are more likely to consume the superior variety. Specifically, in our linear model, it is the sum of both components ($a + \delta$) which determines individual behavior jointly with the rest of parameters.

As we see later in Section 3, these two results change drastically when consumers are free to choose their beliefs.

2.2. Consumers' heterogeneity

Once we have analyzed the behavior of an isolated consumer, we assume for the rest of the paper that there is a mass of consumers that are heterogeneous in the intensity

[3] Pure altruistic individuals are concerned about the others' utility. This implies that an increase in others' welfare (for example, through a larger provision of a public good) makes the individual better off, regardless of whether they are responsible for this improvement or not. Impure altruism (Andreoni, 1989; 1990), on the other hand, considers that the individual also cares directly about her own contribution to others' welfare and gets a private "warm-glow" from giving. In our model, the parameter α can be interpreted -although not exclusively- as the intensity of this "warm-glow" effect.

of their anticipatory utility. Specifically, there are $y \in [0,1]$ consumers in the population with anticipatory utility $a = \bar{a} > 0$, while there are $1 - y$ consumers who experience no anticipatory utility at all ($a = 0$). The following Proposition describes the demand for both varieties in this case:

Proposition 2. Heterogeneity and consumption. Assuming $p_S \leq 1$, the consumption pattern of our economy is the following:

I. If $p_S - p_R \leq \delta\alpha\hat{x}\bar{\theta}$, every consumer in the economy buys S.

II. If $(\bar{a} + \delta)\alpha\hat{x}\bar{\theta} \geq p_S - p_R > \delta\alpha\hat{x}\bar{\theta}$, then the y consumers with $a = \bar{a}$ buy S and the $1 - y$ consumers with $a = 0$ buy R.

III. If $p_S - p_R > (\bar{a} + \delta)\alpha\hat{x}\bar{\theta}$, every consumer in the economy buys R.

The three scenarios are depicted in Figure 2. The intuition of this result is as follows: If the premium for the superior variety is not so large ($p_S - p_R \leq \delta\alpha\hat{x}\bar{\theta}$), every consumer in the economy will buy it (Scenario I). The reason is that, in Scenario I, the expected disutility associated to the regular variety is larger than the premium associated to the superior variety. For moderate values of the premium ($(\bar{a} + \delta)\alpha\hat{x}\bar{\theta} \geq p_S - p_R > \delta\alpha\hat{x}\bar{\theta}$), only those individuals which experience anticipatory disutility will buy the superior variety, while the rest will choose the regular one (Scenario II). Although every consumer considers her expected utility at $t = 2$ when choosing among varieties, a fraction y of the consumers experience also anticipatory disutility in $t = 1$ directly from thinking about the future consequences of their (and other's) behavior. Therefore, they will be willing to pay more for the superior variety than the consumers that do not experience anticipatory disutility at all. Specifically, consumers with anticipatory utility are willing to pay a maximum premium of $(\bar{a} + \delta)\alpha\hat{x}\bar{\theta}$ for the superior variety, while consumers without anticipatory utility are willing to pay a maximum premium of just

$\delta\alpha\hat{x}\bar{\theta}$. Finally, if the premium is too large (larger than $(\bar{a}+\delta)\alpha\hat{x}\bar{\theta}$), every consumer in the economy will buy the regular variety, since the premium associated to the superior variety is larger than the sum of anticipatory and expected disutility associated to the regular one.

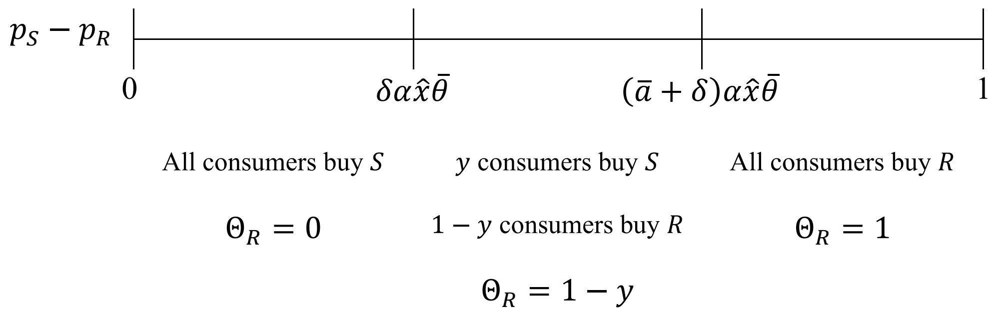


FIGURE 2. Consumption scenarios (assuming $p_S \leq 1$).

2.3. Firm's pricing

A monopolistic firm manufactures varieties R and S at a unitary cost of c_R and c_S. We abstract from fixed costs or capacity constraints. Without loss of generality, we assume that $1 > c_S > c_R$. Given the consumers' preferences, the firm must choose whether manufacturing one or both varieties, as well as their respective prices. For simplicity, we focus only on the price choice, considering that, by setting a sufficiently large price for a given variety, the firm can reduce its demand to zero. The firm's problem is hence:

$$\max_{p_R,p_S} \pi(p_R,p_S) = \Theta_R(p_R,p_S)(p_R - c_R) + [1 - \Theta_R(p_R,p_S)](p_S - c_S) \tag{3}$$

Given the maximization problem and the specific values of each parameter, the following result shows the firm's pricing strategy.

Proposition 3. Firm's choice. For given costs c_S and c_R, beliefs $\hat{x} = x$ and parameters $\bar{a}$, α, δ and $\bar{\theta}$, there are three possible scenarios (proof in the Appendix):

I. Scenario S. If $x > \frac{c_S - c_R}{\delta \alpha \bar{\theta}}$, the firm sells only the variety S at a price $p_S = 1$ and all consumers buy S. Hence, $\Theta_R = 0$ and $\pi_S = 1 - c_S$.

II. Scenario RS. If $\frac{c_S - c_R}{\delta \alpha \bar{\theta}} > x > \frac{(c_S - c_R)y}{(\bar{a} + \delta y)\alpha \bar{\theta}}$, the firm produces both varieties and set prices $p_S = 1$ and $p_R = 1 - \delta \alpha \hat{x} \bar{\theta}$. In this scenario, $\Theta_R = 1 - y$ and $\pi_{RS} = (1 - y)(1 - \delta \alpha x \bar{\theta} - c_R) + y(1 - c_S)$.

III. Scenario R. If $x < \frac{(c_S - c_R)y}{(\bar{a} + \delta y)\alpha \bar{\theta}}$, the firm sells only the variety R at a price $p_R = 1 - (\bar{a} + \delta)\alpha x \bar{\theta}$ and all consumers buy R. Hence, $\Theta_R = 1$ and $\pi_R = 1 - (\bar{a} + \delta)\alpha x \bar{\theta} - c_R$.

The intuition of this result is as follows. The firm will sell only the superior variety S (Scenario S) if its cost is not much larger than the cost of producing R or if the expected private effects of R ($\delta \alpha \hat{x} \bar{\theta}$) are large enough. The firm can charge the maximum price to the superior variety, compensating its costs. If the additional cost of producing S rather than R is relatively high, the other two scenarios arise. The key feature of the problem is that individuals of type $a = \bar{a}$, i.e. those who experience anticipatory (dis)utility, are less willing to pay for R than those of type $a = 0$, due to the perspective of future negative consequences associated to the regular variety. If the proportion of individuals of type $a = \bar{a}$ (y) is large enough, the firm may find it more profitable to set a low price for R in order to sell it to all the consumers (Scenario R). On the other hand, if there are relatively few individuals who experience anticipatory utility, the firm prefers selling the superior variety to them (at p_S) while at the same time selling the regular variety to the rest of the consumers at a relatively high price (Scenario RS).

Related to this, Figure 3 shows the different scenarios as a function of the probability that the consumption of the standard variety has negative effects (x) and the proportion of consumers that experience anticipatory utility (y). For relatively large values of x, the firm finds it optimal to produce only the superior variety, irrespective of y (Scenario S). The reason is simple: the expected disutility associated to the regular variety is increasing in x, and therefore the maximum willingness to pay for R decreases for all the consumers in the economy, reducing the firm's profitability of selling this variety.[4]

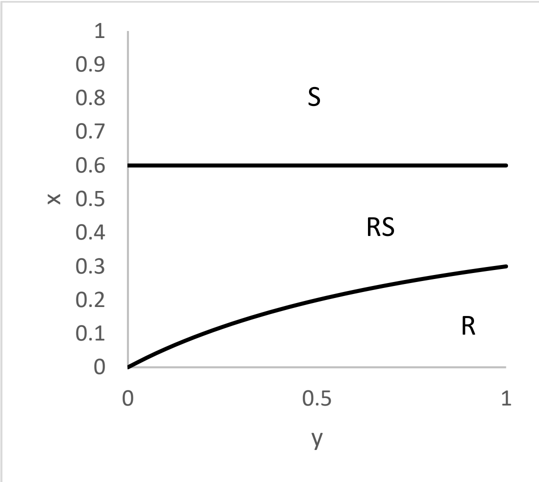


Figure 3. Scenarios with exogenous beliefs.

The most interesting cases take place when x is small enough. In those cases, either Scenario RS or Scenario R and the precise composition of the demand -i.e. the proportion of consumers of each type- plays a crucial role. In general, Scenario R is more likely to happen the smaller the value of x and the larger the proportion y of consumers of type $a = \bar{a}$. The rationale is as follows: consumers of type $a = 0$ are more willing to pay for R than consumers of type $a = \bar{a}$. This implies that the price charged to R in the Scenario RS is larger than in the Scenario R. Further, since $c_S - c_R > \delta\alpha\hat{x}\bar{\theta}$, the unitary profit from selling R to the consumers of type $a = 0$ is larger than the unitary profit from selling S to the consumers of type $a = \bar{a}$ in the Scenario RS. This implies, however, that the firm's profit is strictly decreasing in y, potentially making Scenario R more profitable.

[4] For illustration purposes, the values considered for the parameters are $\alpha = 0.2$, $\delta = 0.5$, $\bar{a} = 0.5$, $c_S - c_R = 0.3$ and $\bar{\theta} = 5$. Different values do not alter substantially the shape of the graph.

On the other hand, profits in both Scenarios RS and R are decreasing in x. However, they decrease faster in the Scenario R, eventually making Scenario RS more appealing.

3. Extended model with endogenous beliefs

In this Section, we present a modified version of the benchmark model that accounts for the possibility that consumers strategically choose their beliefs. In Subsection 3.1., we redefine the consumer's problem considering that she can form and hold beliefs that depart from the objective probability x. While distorted beliefs may alleviate anticipatory disutility, they also increase the risk of making wrong decisions. In Subsection 3.2. we revisit the firm's choice of which varieties to produce and at which prices when consumers' beliefs are endogenous. As we will see, the assumptions about the timing and the firm's conjectures about the consumers' beliefs affect critically the pricing decision.

3.1. Consumer's beliefs

So far, we have assumed that the subjective probability $\hat{x}$ that the consumers assign to the negative event $\theta = \bar{\theta}$ was equal to the objective probability x of that event. In this section, we relax this assumption to allow for the possibility that consumers may manipulate and hold biased beliefs about the consequences of their (and others') acts. Specifically, we assume that there exists a period $t = 0$ before the consumption period at which the consumer freely chooses any belief $\hat{x} \in [0,1]$. This is assumed to be an unconscious cognitive process so that, once $\hat{x}$ has been chosen, the consumer at $t = 1$ truly believes or behaves as if $\hat{x}$ were the objective probability of the state $\theta = \bar{\theta}$. The time structure of the model is depicted in Figure 4.

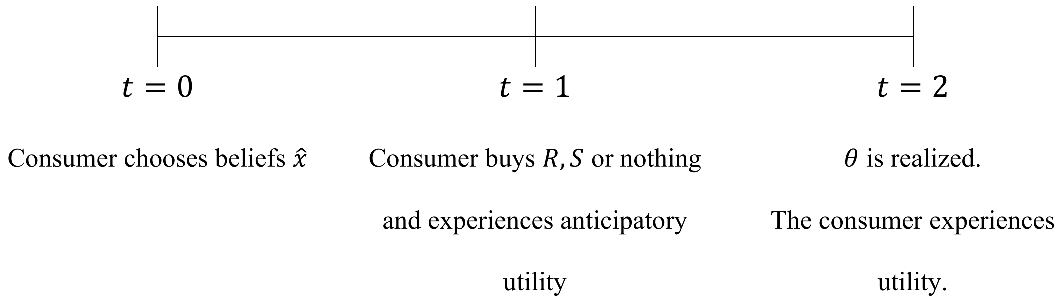


FIGURE 4. Timing in the extended model

Our approach to modelling endogenous belief is similar to Akerlof and Dickens (1982) and Brunnermeier and Parker (2005) in the sense that individuals can choose any belief $\hat{x} \in [0,1]$ regardless of the value of the objective probability x, and that they do not incur in any cognitive cost of manipulation or self-deception. Although this approach seems quite extreme and unrealistic if taken literally, we find it useful for several reasons. First, it still captures the trade-off between the benefits from keeping optimistic beliefs and the costs of making decisions based on unrealistic beliefs. As Brunnermeier and Parker (2005) note, these models are consistent with psychological evidence that finds that individuals tend to report optimistic beliefs particularly when the potential losses from having distorted beliefs are small. Second, some of the results apply qualitatively well, yet in quantitatively softer terms, to more restrictive models that include explicit costs of self-deception. Finally, this approach can be regarded as the opposite extreme to the one presented in Section 3. We expect that reality lies somewhere between these two extreme situations.

At $t = 0$, we assume that consumer's prior beliefs coincide with the objective probability x. The consumer's problem at $t = 0$ is therefore

$$\max_{\hat{x}} EU_{j,0} = \begin{cases} -a\hat{x}\bar{\theta}\beta\Theta_R - \delta x\bar{\theta}\beta\Theta_R & j = \emptyset \\ 1 - p_S - a\hat{x}\bar{\theta}\beta\Theta_R - \delta x\bar{\theta}\beta\Theta_R & j = S \\ 1 - p_R - a\hat{x}\bar{\theta}(\alpha + \beta\Theta_R) - \delta x\bar{\theta}(\alpha + \beta\Theta_R) & j = R \end{cases} \tag{4}$$

$$s.t.\ \hat{x} \in [0,1]$$

Note that, since the true effects from R are not realized until period $t = 2$, the anticipatory utility flow at $t = 1$ depends on the subjective probability $\hat{x}$, that is, on how likely the consumer at $t = 1$ thinks that $\theta = \bar{\theta}$. Conversely, the last term in each sub-function shows the expected utility flow at $t = 2$, when the actual damage is realized. At $t = 2$, the consumer's utility does not depend on her beliefs about θ but on the actual value of θ.

When choosing the optimal belief $\hat{x}^*$ at $t = 0$, the individual must take into consideration two different effects. First, from the maximization problem in (4) it is straightforward that, if $a > 0$, the individual can increase is anticipatory utility just by choosing a small subjective probability $\hat{x}$. By judging that the future negative effects from R will be most likely innocuous, the consumer may reduce her anxiety at $t = 1$. Second, note that the consumer at $t = 1$ still behaves accordingly to Proposition 1. In this sense, consuming the regular product R becomes more attractive the lower is the subjective probability $\hat{x}$. Underestimating x may lead the consumer to buy R instead of S, possibly experiencing a larger disutility at $t = 2$. In other words, when forming her beliefs, the consumer must balance the trade-off that arises between being more comfortable at $t = 1$ by holding an optimistic perspective of the future and the future damages derived from mistakes due to her distorted, too optimistic beliefs.

Hence, at $t = 0$ each consumer chooses her beliefs $\hat{x}$ taking into account that at $t = 1$ she will behave as stated in Proposition 1. For example, if at $t = 0$ she chooses $\hat{x} \geq$

$\frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}}$, at $t = 1$ she will buy the variety S as long as $p_S \leq 1$. Thus, assuming that $p_S \leq 1$, the optimal belief $\hat{x}^*$ can be defined as follows:

$$\hat{x}^* = \arg\max_{\hat{x}} EU_0 \begin{cases} 1 - p_R - a\hat{x}\bar{\theta}(\alpha + \beta\Theta_R) - \delta x\bar{\theta}(\alpha + \beta\Theta_R) & if\ \hat{x} < \underline{x} \\ 1 - p_S - a\hat{x}\bar{\theta}\beta\Theta_R - \delta x\bar{\theta}\beta\Theta_R & if\ \hat{x} \geq \underline{x} \end{cases} \quad (5)$$

$$s.t.\ \hat{x} \in [0,1]$$

where $\underline{x} = \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}}$ is the minimum value of $\hat{x}$ for which the consumer buys the superior variety. The following proposition shows the consumers' optimal choice of beliefs.

Proposition 4. Assuming that $p_S \leq 1$ and that $p_S \geq p_R$, the solution to the maximization problem (5) is

I. If $x \geq \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}} \cdot \frac{\alpha\delta + a(\alpha+\beta\Theta_R)}{\alpha\delta}$, the optimal belief is $\hat{x}^* = \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}}$ and the consumer will buy the superior variety at $t = 1$. In the particular case in which $\beta\Theta_R = 0$, then any $\hat{x} \in \left[\frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}}, 1\right]$ is optimal. Further, if $a = 0$, any $\hat{x} \in \left[\frac{p_S - p_R}{\delta\alpha\bar{\theta}}, 1\right]$ is optimal.

II. If $x < \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}} \cdot \frac{\alpha\delta + a(\alpha+\beta\Theta_R)}{\alpha\delta}$, there is only one solution given by $\hat{x}^* = 0$ and the consumption of the regular variety at $t = 1$. If $a = 0$, any $\hat{x} \in \left[0, \frac{p_S - p_R}{\delta\alpha\bar{\theta}}\right)$ is optimal.

Note that for consumers that do not experience anticipatory utility ($a = 0$), the strategy $\hat{x} = x$ is always optimal. In other words, they do not have incentives to distort their beliefs. This is not surprising, considering that the only advantage from distorting beliefs in our model is to alleviate anticipatory disutility.

Now we take a closer look to some implications that we can extract from Proposition 4.

Self-deception and optimistic bias. In the presence of anticipatory utility, it is generally true that $\hat{x}^* < x$ (only if $\beta\Theta_R = 0$ there could be additional solutions in which $\hat{x}^* \geq x$). Even those consumers choosing the superior variety will be too optimistic about the negative consequences associated to the consumption of the regular product. The reason is that, even if they can avoid part of these consequences by consuming the superior variety, they cannot affect the total provision of the public bad given by Θ_R. Therefore, keeping a limited optimism about the future can alleviate their anxiety at $t = 1$ while at the same time guaranteeing their consumption of S. Belief manipulation is more aggressive for those consumers who prefer the regular variety. These consumers will categorically deny any negative effect from the consumption of R ($\hat{x}^* = 0$).[5]

Guilt and private effects. In our model, the decisions regarding belief formation or consumption depend critically on the assumption that consumers can feel guilty or that there are private harmful effects derived from the consumption of R ($\alpha > 0$). Otherwise, consumers would always be over optimistic and hold denial beliefs (defined as $\hat{x}^* = 0$ and $\hat{E}(\theta) = 0$), choosing R as long as it is the less expensive variety.

Public bad, denial and consumers' choices. The total provision of the public bad does not trigger the reduction of the provision of the variety that produces it. On the contrary, in the presence of anticipatory utility, it may in fact encourage denial beliefs ($\hat{x} = 0$) even among altruistic consumers. Hence, the more consumers buy the regular product R, the more likely other consumers will do the same. Specifically, consumers are more likely to keep denial beliefs the larger $\beta\Theta_R$. This is the opposite to what we found

[5] This extreme behavior is due to the design of our model (in particular, to our assumption that the individuals can freely choose any belief and that there are not cognitive costs from forming and keeping beliefs that are far from the reality) and the inclusion of more restrictive assumptions would attenuate this result. Still, the main result -that consumers, and especially those who will buy R, keep systematically optimistic beliefs- would not be compromised.

in the benchmark model presented in Section 3, where other's action and the provision of the public good were irrelevant in the individual consumption choice. The reason is that, since the individuals are negatively affected by the total provision of R (Θ_R) but their individual action cannot affect Θ_R, some of them may find optimal to deny the negative effects associated to R even if they are altruistic, since the feeling of being a responsible consumer by reducing e is overwhelmed by the anxiety produced by the perspective of a gloomy future.

Anticipatory utility and temporal discounting. It is straightforward from Proposition 4 that the maintenance of positive beliefs $\hat{x} = \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}}$ and the subsequent consumption of the superior variety at $t = 1$ is more likely the larger the value of δ, i.e. if consumers are patient or if period 2 is close enough. Prosocial behavior (in the sense of not contributing to the provision of the public bad) is more likely to be associated to those individuals who are more patient and put a larger weight on future periods. This is in line with the results of the benchmark model. However, we find now that denial beliefs and consumption of the regular variety associated to the public bad are more likely the larger the value of a, which measures the intensity of anticipatory utility (see proof in the Appendix). In other words, those individuals who tend to feel more anxious about the future are more likely to form optimistic beliefs about the future in order to alleviate bad feelings in the previous periods. This is in contrast with the benchmark model, where we found that a and δ worked in the same direction, encouraging prosocial behavior. The reason of this difference is pretty clear, however. At $t = 1$, the discomfort experienced from anticipatory utility is increasing in both a and $\hat{x}$. If the individual cannot choose her beliefs, the only way of reducing anticipatory disutility is by consuming S rather than R. This, however, alleviates discomfort only partially, since in our large economy it does not affect the total consumption of R. However, if consumers can choose their beliefs, anticipatory discomfort can be fully eliminated just by choosing $\hat{x} = 0$, which is associated to the consumption of R as long as $p_S > p_R$.

Given these radically opposite results between the two models, an interesting question arises. In the presence of uncertainty, would those individuals who are more apprehensive about the future be more or less likely to take actions in order to improve future payoffs? The answer seems to be that it depends on how easily they can deceive themselves about that future. If, for example, the existing evidence is too hard to be ignored, it seems unlikely that individuals can easily hold unrealistic beliefs, and more apprehensive individuals are more likely to take corrective actions. On the other hand, if the individual can -at least partially- choose her beliefs, apprehension may trigger self-deception rather than action.

3.2. Firm's pricing

We keep all the assumption from Section 3.3 regarding the monopolistic firm and the cost structure. In the presence of endogenous beliefs, a critical question arises: whether the firm does or does not consider that its pricing decision may influence the process of belief formation depicted previously. If the firm is aware that the prices will determine consumers' beliefs (and therefore their consumption choices), then pricing gains even more strength as a strategic tool for the firm. Therefore, in this Section we will focus on that case, although the interested reader can find a description of the alternative simultaneous game in the Appendix.

If the firm knows that the consumers will form their beliefs as depicted in Proposition 4 and the consumers believe that their actions will have no impact on prices, then we can analyze the problem as a dynamic game in which the firm first choses the prices of each variety and then the consumers form their beliefs and chose the variety they prefer. The result is presented in the following proposition:

Proposition 5. Subgame Perfect Equilibria. Depending on the specific values of the parameters, several Subgame Perfect Equilibria arise:

I) Scenario S. If $x > \frac{c_S - c_R}{\delta\alpha\bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta y}{\alpha(\bar{a}+\delta)}$, the firm will sell only the superior variety to all the consumers at a price $p_S = 1$, and any belief $\hat{x}_0$ and $\hat{x}_{\bar{a}}$ is optimal. The total provision of public bad in the economy is $\Theta_R = 0$.

II) Scenario RS. If $\frac{c_S - c_R}{\delta\alpha\bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta y}{\alpha(\bar{a}+\delta)} > x > \frac{c_S - c_R}{\delta\alpha\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$, the firm will sell both varieties at prices $p_S = 1$ and $p_R = 1 - \alpha\delta\bar{\theta}x\frac{\alpha(\bar{a}+\delta)}{\alpha(\bar{a}+\delta)+\bar{a}\beta y}$. Consumers of type $a = 0$ will choose the superior variety and those of type $a = \bar{a}$ will buy the regular one. The total provision of public bad in the economy is $\Theta_R = y$.[6]

III) Scenario R. Finally. if $x < \frac{c_S - c_R}{\delta\alpha\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$, the firm will sell only the regular variety to all the consumers at a price $p_R = 1 - \delta\alpha x\bar{\theta}$, and consumers' optimal beliefs are defined by $\hat{x}_0 \leq x$ and $\hat{x}_{\bar{a}} = 0$. The total provision of public bad in the economy is $\Theta_R = 1$.

While the three scenarios seem similar to those found in the benchmark model (see Proposition 3), there are substantial differences, particularly in Scenario RS.

In the first place, the consumers of each variety in the Scenario RS are different in the benchmark model and in the model with endogenous beliefs. With exogenous beliefs, those individuals with anticipatory utility ($a = \bar{a}$) were the ones consuming the

[6] In a strict sense, p_R is infinitesimally smaller than $1 - \alpha\delta\bar{\theta}x\frac{\alpha(\bar{a}+\delta)}{\alpha(\bar{a}+\delta)+\bar{a}\beta y}$.

superior variety, while consumers with $a = 0$ chose the regular one. In the current situation with endogenous beliefs it is precisely the opposite.

Remember that, in the benchmark model, Scenario RS is more likely to take place when the proportion y of consumers with anticipatory utility is relatively small (see Figure 3). The reason is that they are less willing to pay for the regular variety. Unless they represent a large proportion of the consumers, the firm usually will find it more profitable to sell them the superior variety, while selling the regular one at a large price to the rest of consumers. Again, now the opposite is true. Figure 4 shows the different scenarios in the model with endogenous beliefs as a function of the probability x and the proportion of consumers that experience anticipatory utility. With endogenous beliefs, the consumers with anticipatory utility are more willing to pay for the regular variety than those of type $a = 0$. Thus, the larger the proportion of consumers with $a = \bar{a}$, the more likely the firm will prefer selling the regular variety only to them but at a larger price, while selling the superior variety to the rest of consumers.[7]

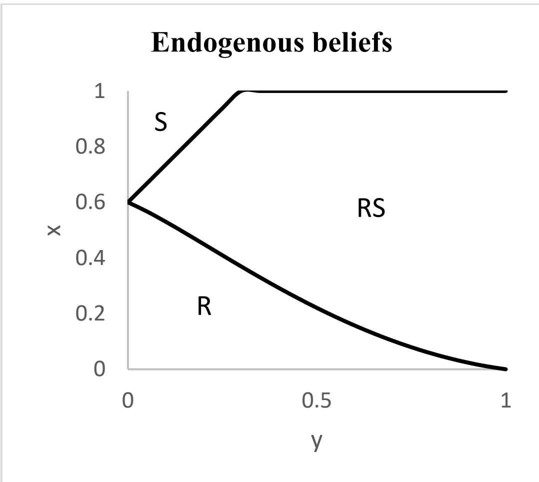


Figure 4. Scenarios with endogenous beliefs.

[7] For illustration purposes, the values considered for the parameters are $\alpha = 0.2$, $\delta = 0.5$, $\bar{a} = 0.5$, $c_S - c_R = 0.3$ and $\bar{\theta} = 5$. Different values do not alter substantially the shape of the graph.

Finally, the price charged to the regular product in Scenario RS is larger now than it was in the benchmark model, as long as $a_0, \beta, y > 0$ and the difference is increasing in all these parameters. Remember that, in the benchmark model, the consumers of variety R in Scenario RS where those of type $a = 0$. While this type of consumer does not experience anticipatory utility, it still takes into account her future expected utility when making her consumption choice. Now, in contrast, consumers of type $a = \bar{a}$ are the ones buying R in Scenario RS. While they experience anticipatory utility, they completely distort their beliefs so that they truly believe that variety R entails no future harm at all. Thus, they are more willing to pay for R than consumers of type $a = 0$ in the benchmark model.

4. Endogenous vs. exogenous beliefs

We now compare the main results obtained in the benchmark model (exogenous beliefs) and in the extended model with endogenous beliefs. Some differences have been already pointed out in Section 4. For example, while in the benchmark model the discount factor δ and the intensity of anticipatory utility a affect the consumer's behavior in the same way, this is no longer the case in the extended model. Also, public effects matter only when beliefs are endogenous. Hence, in this Section we focus on the differences regarding the firm's profits and the demand for the regular variety in the economy (Θ_R), which is also the total provision of the hazardous component e in the economy.

The following proposition illustrates the relationship between the firm's profits and the precise nature of beliefs (see the proof in the Appendix):

Proposition 6. In our model, the firm's profits are always equal or larger when beliefs are endogenous than when these are exogenous, regardless of the specific values of the parameters.

In our model, there is a hazardous component e associated to the regular product whose consumption (by oneself or by others) may reduce the consumer's future utility. Therefore, as we discussed in Proposition 4, consumers that experience anticipatory disutility will generally favor beliefs that underestimate the probability of e having bad consequences in the future. This increases their willingness to pay for the regular variety, while the willingness to pay for the superior variety remains constant. Since the supply side of the economy is formed by a monopolistic firm that produces the two varieties, it seems clear that its profits will never be smaller in the model with endogenous beliefs than in the benchmark model with exogenous beliefs. It must be noticed, however, that this result depends critically on our specific setting[8].

If there are public effects ($\beta > 0$), another relevant question is whether endogenous beliefs lead to larger levels of the provision of the public bad in the economy, which is given by the demand of the regular product Θ_R, due to an excess of optimism regarding its future negative consequences. As the following Proposition shows, however, there is no definite answer. Under some circumstances, in fact, it could be the case that the existence of too optimistic consumers who underestimate the negative consequences associated to the regular variety leads to a lower demand for this product and, therefore, a smaller provision of the public bad in the economy.

Proposition 7. The total provision of the public bad, Θ_R, is smaller when consumers' beliefs are endogenous than in the benchmark model only when any of the following two conditions hold:

[8] In an alternative model of perfect competition, profits are zero with endogenous and exogenous beliefs. Also, if we assume that a single firm produces the superior variety while the regular product is manufactured by a competitive industry, then the firm's profits are larger when the consumers' beliefs are exogenous -opposite to our result-. See Appendix II for more details.

(i) $\frac{(c_S-c_R)y}{(\bar{a}+y\delta)\alpha\bar{\theta}} > x > \frac{c_S-c_R}{\alpha\delta\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$

(ii) $\frac{c_S-c_R}{\alpha\delta\bar{\theta}} > x > \frac{c_S-c_R}{\alpha\delta\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$ and $y < 0.5$

Note that the first condition may be not even feasible, since nothing guarantees that the first term in the inequality is larger than the last one. In general, for that scenario to be feasible, it is necessary that the proportion y of individuals that experience anticipatory utility is large enough.

The intuition of this result is as follows: in some circumstances, if the number of individuals who experience anticipatory utility is large enough, the firm may find more profitable selling the regular variety to everyone in the economy at a price $p_R = 1 - (\bar{a} + \delta)\alpha\hat{x}\bar{\theta}$ when beliefs are exogenous but only to those individuals with experience anticipatory disutility when they can self-deceive. The rationale is that, in that case, it can charge them a larger price $p_R = 1 - \alpha\delta\bar{\theta}x\frac{\alpha(\bar{a}+\delta)}{\alpha(\bar{a}+\delta)+\bar{a}\beta y}$, while selling the superior variety at a price $p_S = 1$ to the rest of consumers. Thus, endogenous beliefs would lead to a lower provision of the public bad in the economy, with those consuming the regular variety paying a larger price for it than in the benchmark model.

Figure 5 shows the demand of the regular variety in the benchmark model and in our model with endogenous beliefs as a function of the probability x. To better illustrate this, we include several scenarios using different values for the parameters.

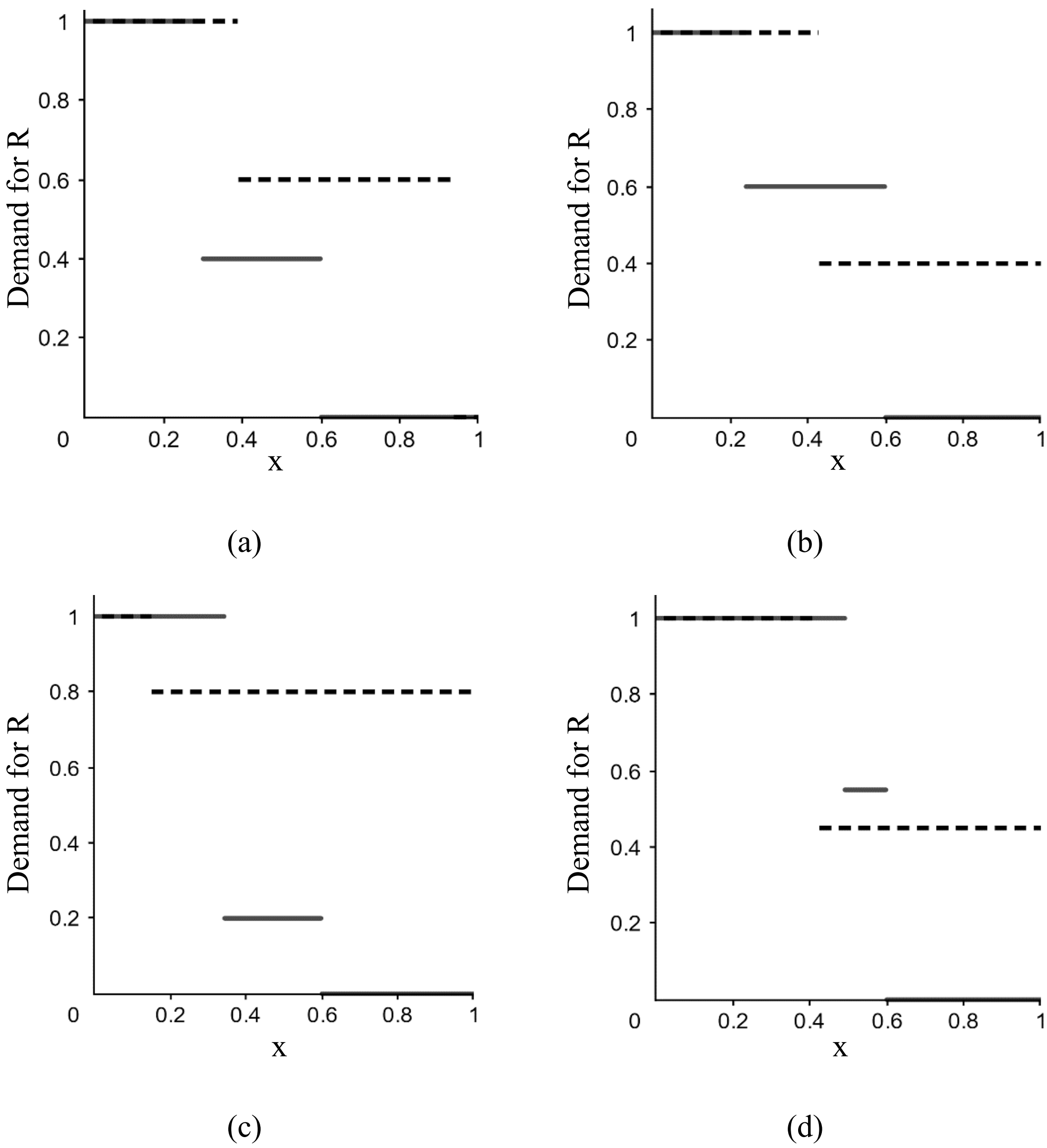


Figure 5. Demand for the regular product in the benchmark model (continuous line) and in the model with endogenous beliefs (dashed line)[9].

Figure 5(a) shows a typical case in which self-deception leads to an equal or larger demand for the regular variety irrespective of the probability that this product will have negative consequences in the future. Conversely, figure 5(d) presents a case in which the model with self-deception predicts an equal or smaller demand for R, irrespective of x. Finally, figures 5(b) and 5(c) show situations in which this comparison depends on the

[9] The specific values of the parameters used in all cases are $\alpha = 0.2, \delta = 0.5, \bar{a} = 0.3,\ \bar{\theta} = 5, c_R = 0.2, c_S = 0.5$. The four cases differ only in the values of β and y. Specifically, (a) $\beta = 0.5$ and $y = 0.6$; (b) $\beta = 2$ and $y = 0.4$; (c) $\beta = 2$ and $y = 0.8$; and (d) $\beta = 5$ and $y = 0.45$.

specific values of x. In 5(b), the demand of R is equal or larger in the model of self-deception for relatively low values of x, while it is equal or smaller for larger values. The opposite is true in 5(c).

6. Conclusions

In this paper, we have modelled an economy in which consumers may experience anticipatory utility and can choose their beliefs. In contrast with classical models, individuals may find optimal to keep unrealistic beliefs in the presence of anticipatory utility. Here we take a step further and show how this can alter the economic agents' incentives and, consequently, the market outcome.

Specifically, we have assumed that consumers are uncertain about the risk associated to the regular variety of a product, whose consumption could be damaging for themselves and for others. Because of the existence of anticipatory utility, consumers have preferences not only over actions but also over their own beliefs about the true state of the world. Specifically, they prefer to think that the risk associated to the regular variety is lower than it is in fact. This contrasts with classical models, in which agents do always prefer to hold beliefs as much accurate and realistic as possible. Underestimating the risks associated to the standard variety alleviates the anticipatory disutility experienced by the consumers in the present but it has consequences for the market equilibrium. Other things equal, optimistic consumers are more willing to pay for the standard product than consumers with realistic beliefs. In turn, this may alter the incentives and behavior of the firms participating in the market. Under some circumstances, as in our monopolistic model, a firm could take advantage from consumers' delusion and increase its profits.

From the comparison between our model with endogenous beliefs and the benchmark model in which self-deception is not an option, three main conclusions can be extracted. First, under the assumption of no self-deception, those consumers who experience larger levels of anticipatory utility -e.g. those who experience more excitement or anxiety in the present by thinking about their future- are more prone to take action in the present in order to increase their future utility. This is no longer the case when individuals can hold their preferred beliefs. In that case, larger levels of anticipatory utility may trigger self-deception rather than action. This has important implications for policymaking. Think, for example, on climate change. Should the government or NGOs appeal to people's sentiments in order to promote individual action? If different beliefs are easy to maintain without challenge -for example, because there are a multitude of opinions in the media-, this approach may backfire and foster denial of climate change.

Second, self-deception is more likely to arise when there is uncertainty about the effects of a public good or about externalities and individual actions have little or no effect on the economy. Think again on climate change. If the total emissions of greenhouse gases are too large and each individual action is negligible, it could be optimal for those that experience anticipatory utility to underestimate the risks associated to greenhouse gases. Since they cannot amend the level of total emissions, at least they could alleviate their distress by keeping unrealistic and optimistic beliefs about the future.

Third, we have seen that a firm can benefit from self-deception. In our model, this is because optimistic consumers have a larger willingness to pay for the risky product. This, however, does not mean that self-deception is necessarily associated to a larger level of consumption of this product. As we discussed in Section 5, it could be the case that the firm find more profitable to sell the risky product to a less portion of the consumers - relative to the model without self-deception- but at a larger price. Eventually, this will depend on the specific characteristics of the economy, including the intensity of anticipatory utility among individuals.

There are several extensions of our model. We focus on the case of an industry formed by a monopolistic firm that produces several varieties of the same product. In the Appendix, we also consider the cases of perfect competition and a monopoly that produces only one of the varieties. While there are some changes -particularly regarding the firm's profits-, most results do not change drastically. Alternative approaches could consider some types of imperfect competition, like oligopoly or monopolistic competition.

Following other works (Akerlof and Dickens (1982); Brunnermeier and Parker (2005)), our model approaches the process of belief choice in the simplest fashion: consumers exhibit certain preferences over a set of beliefs and they just choose the ones that maximize their expected utility. Alternative models (Bénabou (2013), Bénabou and Tirole (2002), Köszegi (2006), Möbius et al. (2014)) depict this process in a more sophisticated way that puts emphasis on the role of information. In this case, consumers do not directly choose their preferred beliefs, but they process and internalize new evidence in an asymmetrically, convenient way. Our future research will incorporate this approach, which we think that could provide a deeper and interesting insight about the strategic use of information by the firms or third parties like the government, NGOs, etc.

Also, a call for caution must be made. While there are plenty of neuroscientific and field studies that seem to confirm the existence of the optimistic bias, experimental evidence is much more limited. So far, experimental works have provided mixed results and their procedures are not always easily comparable (see Benjamin (2019) for a review on experimental research on optimism bias). This supports the idea that the appearance and importance of the optimism bias is context dependent. The determination of the factors and circumstances that may favor or prevent this bias is one of the most important challenges for future research on this subject.

Finally, our agenda also includes further research on the design of public policies under the presence of anticipatory utility and endogenous beliefs.

References

Akerlof, G. A., & Dickens, W. T. (1982). The Economic Consequences of Cognitive Dissonance. *The American Economic Review*, *72*(3), 307–319.

Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, *141*(4), 757–773.

Bastian, B., Loughnan, S., Haslam, N., & Radke, H. R. M. (2012). Don't Mind Meat? The Denial of Mind to Animals Used for Human Consumption. *Personality and Social Psychology Bulletin*, *38*(2), 247–256.

Bénabou, R. (2013). Groupthink: Collective Delusions in Organizations and Markets. *The Review of Economic Studies*, *80*(2), 429–462.

Bénabou, R. (2015). The Economics of Motivated Beliefs. *Revue d'économie politique*, *125*(5), 665.

Benabou, R., & Tirole, J. (2002). Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, *117*(3), 871–915.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, *2*, 69–186.

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, *95*(4), 1092–1118.

Caplin, A., & Leahy, J. (2001). Psychological Expected Utility Theory and Anticipatory Feelings. *The Quarterly Journal of Economics*, *116*(1), 55–79.

Cross, K. P. (1977). Not can, but will college teaching be improved? *New Directions for Higher Education*, *1977*(17), 1–15.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Di Tella, R., Perez-Truglia, R., Babino, A., & Sigman, M. (2015). Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism. *American Economic Review*, *105*(11), 3416–3442.

Haisley, E. C., & Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, *68*(2), 614–625.

Hestermann, N., Le Yaouanq, Y., & Treich, N. (2019). *An economic model of the meat paradox*. (No. 164). Discussion Paper.

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

Kőszegi, B. (2010). Utility from anticipation and personal equilibrium. *Economic Theory*, *44*(3), 415–444.

Lamont, O. A., & Thaler, R. H. (2003). Can the Market Add and Subtract? Mispricing in Tech Stock Carve-outs. *Journal of Political Economy*, *111*(2), 227–268.

Loewenstein, G. (1987). Anticipation and the Valuation of Delayed Consumption. *The Economic Journal*, *97*(387), 666.

Loughnan, S., Haslam, N., & Bastian, B. (2010). The role of meat consumption in the denial of moral status and mind to meat animals. *Appetite*, *55*(1), 156–159.

Malmendier, U., & Tate, G. (2005). Does Overconfidence Affect Corporate Investment? CEO Overconfidence Measures Revisited. *European Financial Management*, *11*(5), 649–659.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

Minsky, H. P. (1974). The Modelling of Financial Instability: An Introduction. *Hyman P. Minsky Archive*, *467*.

Möbius, M., Niederle, M., Niehaus, P., & Rosenblat, T. (2014). *Managing Self-Confidence*. *working paper*.

Rabin, M. (1994). Cognitive dissonance and social change. *Journal of Economic Behavior & Organization*, *23*(2), 177–194.

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.

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945.

Shiller, R. J. (2000). Measuring Bubble Expectations and Investor Confidence. *Journal of Psychology and Financial Markets*, *1*(1), 49–60.

Spinnewijn, J. (2015). Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs. *Journal of the European Economic Association*, *13*(1), 130–167.

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.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*(2), 143–148.

Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science*, *246*(4935), 1232-1233.

Weinstein, N. D., Marcus, S. E., & Moser, R. P. (2005). Smokers' unrealistic optimism about their risk. *Tobacco Control*, *14*(1), 55–59.

Appendix I: Proofs

Proof of Proposition 1

At $t = 1$, the consumer's subjective expected utility from buying S, R or nothing is, respectively:

$$\hat{E}U_{S,1} = 1 - p_S - (a+\delta)\beta\Theta_R\hat{x}\bar{\theta}$$

$$\hat{E}U_{R,1} = 1 - p_R - (a+\delta)(\alpha + \beta\Theta_R)\hat{x}\bar{\theta}$$

$$\hat{E}U_{\emptyset,1} = -(a+\delta)\beta\Theta_R\hat{x}\bar{\theta}$$

The consumer chooses S if $\hat{E}U_{S,1} \geq \hat{E}U_{R,1}$ and $\hat{E}U_{S,1} \geq \hat{E}U_{\emptyset,1}$. That is, if

$$\hat{x} \geq \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}} \quad \text{and} \quad p_S \leq 1$$

Analogously, the consumer chooses R if $\hat{E}U_{R,1} > \hat{E}U_{S,1}$ and $\hat{E}U_{R,1} \geq \hat{E}U_{\emptyset,1}$. That is, if

$$\hat{x} < \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}} \quad \text{and} \quad p_R \leq 1 - (a+\delta)\alpha\hat{x}\bar{\theta}$$

Finally, the consumer does not buy nothing at all if $\hat{E}U_{\emptyset,1} > \hat{E}U_{S,1}$ and $\hat{E}U_{\emptyset,1} > \hat{E}U_{R,1}$. That is, if $p_S > 1$ and $p_R > 1 - (a+\delta)\alpha\hat{x}\bar{\theta}$.

Proof of Proposition 3. The firm must decide whether to supply the whole market or not, as well as whether to produce one or both varieties. Since we assume that $\mathbf{1} > \boldsymbol{c_S}$ and consumers' maximum willingness to pay for $\boldsymbol{S}$ is 1, it is clear that the firm will always find profitable to supply the whole market. Hence, the question is whether to supply one or both varieties and at which price.

If the firm wants to sell S to every consumer, from Proposition 1 the maximum price it can charge is $p_S = 1$. Its profit would be then $\pi_S = 1 - c_S$.

Alternatively, if the firm wants to sell R to every consumer, from Lemma 1 the maximum price it can charge is $p_R = 1 - (a+\delta)\alpha\hat{x}\bar{\theta}$. Since we have heterogeneous consumer that differ in the value of a, the maximum price that it can charge if it wants that every consumer buys R is $p_R = 1 - (\bar{a}+\delta)\alpha\hat{x}\bar{\theta}$. Its profit would be then $\pi_R = 1 - (\bar{a}+\delta)\alpha\hat{x}\bar{\theta} - c_R$.

Finally, the firm can sell both varieties. Specifically, it could sell R only to consumers of type $a = 0$ at a larger price $p_R = 1 - \delta\alpha\hat{x}\bar{\theta}$, while selling S at price $p_S = 1$ to the rest of consumers. Its profits would be then $\pi_{RS} = y(1 - c_S) + (1 - y)(1 - \delta\alpha\hat{x}\bar{\theta} - c_R)$.

Firm's choice will be therefore the one that maximizes its profits. For example, it will sell only S if $\pi_S > \pi_R$ and $\pi_S > \pi_{RS}$, i.e. if $x > \frac{c_S - c_R}{(\bar{a}+\delta)\alpha\bar{\theta}}$ and $x > \frac{c_S - c_R}{\delta\alpha\bar{\theta}}$. Since the latter condition is more restrictive, it is the necessary and sufficient condition for the scenario in which the firm sells S to all consumers. The other scenarios in Proposition 3 come straightforward from these comparisons.

Proof of the role of anticipatory utility and temporal discount in the determination of optimal beliefs.

From Proposition 4, the condition for forming beliefs $\hat{x}^* = \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}}$ and consuming the superior variety is $x \geq \underline{x}$, where

$$\underline{x} = \frac{p_S - p_R}{(a+\delta)\alpha\bar{\theta}} \cdot \frac{\alpha\delta + a(\alpha + \beta\Theta_R)}{\alpha\delta}$$

Taking partial derivatives on δ

$$\frac{\partial \underline{x}}{\partial \delta} = \frac{(p_S - p_R)(a+\delta)\alpha^3\delta\bar{\theta} - (p_S - p_R)[\alpha\delta + a(\alpha + \beta\Theta_R)]\alpha^2\delta\bar{\theta}}{[(a+\delta)\alpha^2\delta\bar{\theta}]^2}$$

$\frac{\partial \underline{x}}{\partial \delta} < 0$ if

$$\alpha\delta + a(\alpha + \beta\Theta_R) > (a + \delta)\alpha$$

$$\beta\Theta_R > 0$$

As long as this condition is satisfied and other things being equal, $\underline{x}$ will be lower the larger δ is, meaning that those consumers who weight the future more will keep positive beliefs ($\hat{x} > 0$) and consume the superior variety even for smaller values of x.

Taking partial derivatives on a_0

$$\frac{\partial \underline{x}}{\partial a} = \frac{(p_S - p_R)(\alpha + \beta\Theta_R)(a + \delta)\alpha^2\delta\bar{\theta} - (p_S - p_R)[\alpha\delta + a(\alpha + \beta\Theta_R)]\alpha^2\delta\bar{\theta}}{[(a + \delta)\alpha^2\delta\bar{\theta}]^2}$$

$\frac{\partial \underline{x}}{\partial a} > 0$ if

$$(\alpha + \beta\Theta_R)(a + \delta) > \alpha\delta + a(\alpha + \beta\Theta_R)$$

$$\alpha a + \alpha\delta + \beta\Theta_R a + \beta\Theta_R\delta > \alpha\delta + \alpha a + \beta\Theta_R a$$

$$\beta\Theta_R\delta > 0$$

Therefore, other things equal, as long as $\beta\Theta_R\delta > 0$, $\underline{x}$ will be larger for those consumers with more intense anticipatory utility (larger a), meaning that they require a larger objective probability x in order to form positive beliefs ($\hat{x} > 0$) and consume the superior variety.

Proof of Proposition 5. The different equilibria and their conditions can be determined by backward induction. Hence, we first consider the consumers' choice given prices and then we move to the firm's decision.

Scenario S requires that, given prices, every consumer in the economy chooses the variety S. For this Scenario to constitute a Subgame Perfect Equilibria, it is necessary in first place that, given p_S, p_R and $\Theta_R = 0$, no consumer has incentives to deviate and consume R (or nothing) instead. From Propositions 1 and 4, this requires that $x \geq \frac{p_S - p_R}{\alpha\delta\bar{\theta}}$ and $p_S \leq 1$. Thus, the maximum profit that the firm could obtain in this case is $\pi_S = 1 - c_S$, for $p_S = 1$.

Analogously, Scenario R requires that, given p_S, p_R and $\Theta_R = 1$, consumers do not have incentives to deviate and buy S (or nothing) instead. Consumers of type $a = 0$ and $a = \bar{a}$ have no incentives to deviate if $x < \frac{p_S - p_R}{\alpha\delta\bar{\theta}}$ and $x < \frac{p_S - p_R}{\alpha\delta\bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta}{\alpha(\bar{a}+\delta)}$, respectively -note that the first condition is more restrictive-, and $p_R \leq 1 - \alpha\delta\bar{\theta}x$. Thus, the maximum profit that the firm could obtain in this case is $\pi_S = 1 - \alpha\delta\bar{\theta}x - c_R$, for $p_R = 1 - \alpha\delta\bar{\theta}x$.

Finally, Scenario RS requires that, given p_S, p_R and $\Theta_R = y$, the y consumers of type $a = \bar{a}$ and the $1 - y$ consumers of type $a = 0$ have no incentives to deviate from buying R and S, respectively. This requires that $\frac{p_S - p_R}{\alpha\delta\bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta y}{\alpha(\bar{a}+\delta)} > x \geq \frac{p_S - p_R}{\alpha\delta\bar{\theta}}$ and $p_S \leq 1$. The maximum profit that the firm can make in this scenario is $\pi_{RS} = y\left(1 - \alpha\delta\bar{\theta}x\frac{\alpha(\bar{a}+\delta)}{\alpha(\bar{a}+\delta)+\bar{a}\beta y} - c_R\right) + (1-y)(1-c_S)$ with prices $p_S = 1$ and $p_R = 1 - \alpha\delta\bar{\theta}x\frac{\alpha(\bar{a}+\delta)}{\alpha(\bar{a}+\delta)+\bar{a}\beta y}$.

Taking this into account, the firm will set the prices that lead to the scenario that maximizes its profits. Thus, for Scenario S to constitute a Subgame Perfect Equilibrium, in addition to the previous conditions it is necessary that $\pi_S \geq \pi_R$ and $\pi_S \geq \pi_{RS}$, i.e. if $x \geq \frac{c_S - c_R}{\alpha\delta\bar{\theta}}$ and $x \geq \frac{c_S - c_R}{\alpha\delta\bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta y}{\alpha(\bar{a}+\delta)}$. Note that the latter condition is more restrictive and therefore it is the necessary and sufficient condition for Scenario S.

Similarly, for Scenario R to constitute a Subgame Perfect Equilibrium it is necessary that $\pi_R \geq \pi_S$ and $\pi_R \geq \pi_{RS}$ or, substituting and reordering, $x \leq \frac{c_S - c_R}{\alpha\delta\bar{\theta}}$ and $x \leq \frac{c_S - c_R}{\delta\alpha\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$. The latter condition is more restrictive and it is therefore the necessary and sufficient condition.

Finally, Scenario RS is a Subgame Perfect Equilibrium if $\pi_{RS} \geq \pi_S$ and $\pi_{RS} \geq \pi_R$, i.e. if $\frac{c_S - c_R}{\alpha\delta\bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta y}{\alpha(\bar{a}+\delta)} \geq x \geq \frac{c_S - c_R}{\delta\alpha\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$.

Proof of Proposition 6. Let denote π_S^{exo}, π_{RS}^{exo} and π_R^{exo} the firm's profits in the corresponding Subgame Perfect Equilibria given in Proposition 3 for the model with exogenous beliefs. Similarly, let denote π_S^{endo}, π_{RS}^{endo} and π_R^{endo} the firm's profits in the different Subgame Perfect Equilibria given in Proposition 5.

Note the following properties:

1. $\pi_S^{exo} = \pi_S^{endo}$
2. $\pi_R^{endo} > \pi_R^{exo}$ (as long as $\bar{a}\alpha x\bar{\theta} > 0$)
3. $\pi_{RS}^{exo} = (1-y)\pi_R^{endo} + y\pi_S^{endo}$. Thus, if $\pi_{RS}^{exo} > \pi_R^{endo}$, then $\pi_{RS}^{exo} < \pi_S^{endo}$ and vice versa.

In the model with endogenous beliefs, the firm chooses Scenario R if $\pi_R^{endo} > \pi_S^{endo}$ and $\pi_R^{endo} > \pi_{RS}^{endo}$. The previous properties guarantee that, in this case, π_R^{endo} is necessarily larger than π_S^{exo} (because of property 1), π_R^{exo} (property 2) and π_{RS}^{exo} (property 3). To better illustrate the last case, let suppose that $\pi_R^{endo} < \pi_{RS}^{exo}$. Then, because of property 3, it must be the case that $\pi_S^{endo} > \pi_{RS}^{exo}$. However, we have established that $\pi_R^{endo} > \pi_S^{endo}$ so it cannot be the case that $\pi_R^{endo} < \pi_{RS}^{exo}$.

Similarly, the firm will choose Scenario S if $\pi_S^{endo} > \pi_R^{endo}$ and $\pi_S^{endo} > \pi_{RS}^{endo}$. Using these conditions and the previous properties, this implies that π_S^{endo} is equal to π_S^{exo} and larger than π_R^{exo} and π_{RS}^{exo}.

Finally, the firm chooses Scenario RS if $\pi_{RS}^{endo} > \pi_R^{endo}$ and $\pi_{RS}^{endo} > \pi_S^{endo}$. The previous properties guarantee that, under these conditions, π_{RS}^{endo} is larger that π_S^{exo}, π_R^{exo} and π_{RS}^{exo}.

Proof of Proposition 7. For Θ_R to be smaller in the model with endogenous beliefs than in the benchmark model it is necessary that the parameters take values so that the corresponding Subgame Perfect Equilibria are:

i. Scenario RS in the model with endogenous beliefs ($\Theta_R = y$) and Scenario R in the model with exogenous beliefs ($\Theta_R = 1$). From the conditions given in Propositions 3 and 5, this requires that $\frac{(c_S - c_R)y}{(\bar{a}+y\delta)\alpha\bar{\theta}} > x > \frac{c_S - c_R}{\alpha\delta\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$.

ii. Scenario RS in the model with endogenous beliefs ($\Theta_R = y$) and Scenario RS or R ($\Theta_R = 1 - y$ or $\Theta_R = 1$) in the model with exogenous beliefs, as long as $y < 0.5$. From the conditions given in Propositions 3 and 5, this requires that $\frac{c_S - c_R}{\alpha\delta\bar{\theta}} > x > \frac{c_S - c_R}{\alpha\delta\bar{\theta}} \cdot \frac{(1-y)\alpha(\bar{a}+\delta)+(1-y)\bar{a}\beta y}{(1-y)\alpha(\bar{a}+\delta)+\bar{a}\beta y}$.

Note that, from Propositions 3 and 5, there is no case in which parameters take values such that Scenario S constitutes a Subgame Perfect Equilibria in the model with endogenous beliefs but not in the benchmark model with exogenous beliefs.

Appendix II: Extensions

1. Simultaneous game with endogenous beliefs.

We briefly present here an alternative version of the game analyzed in Subsection 4.2. In this alternative setting, we will assume that both the firm and the consumers think that their respective choices of prices and beliefs will not affect other's decisions. In other words, we assume now that beliefs and prices are set simultaneously. In this case, there exist only two types of pure strategy Nash equilibria in which every consumer buys the same variety.

For example, for $x > \frac{c_S - c_R}{\alpha\delta\bar{\theta}}$, $p_S = 1$, $p_R = 1 - \alpha\delta x\bar{\theta}$, $\hat{x}_0 = x$ and $\hat{x}_{\bar{a}} = \frac{\delta}{\bar{a}+\delta}x$, there exists a pure strategy Nash equilibrium in which every consumer buys the superior variety ($\Theta_R = 0$). From Proposition 4, both types of consumer have no incentives to deviate from these beliefs given prices. At the same type, the firm does not have incentives to change the prices given the consumers' beliefs as long as $x > \frac{c_S - c_R}{\alpha\delta\bar{\theta}}$, or, in other words, as long as the cost of producing the superior variety relative to the regular one is small enough.

Also, for $x < \frac{c_S - c_R}{\alpha\delta\bar{\theta}}$ there exists a pure strategy Nash equilibrium characterized by $p_S > 1$, $p_R = 1 - \alpha\delta x\bar{\theta}$, $\hat{x}_0 = x$, $\hat{x}_{\bar{a}} = 0$ in which all consumers buy the regular variety ($\Theta_R = 1$).

However, in the simultaneous game there is no pure strategy Nash equilibrium in which consumers of different types choose different varieties. To see why, let assume that the firm set prices $p_S = 1$ and $p_R < p_S$ such that consumers of type $a = 0$ choose the superior variety while consumer of type $a = \bar{a}$ choose the regular variety. Following proposition 4, it is optimal for the latter keeping beliefs $\hat{x}_{\bar{a}} = 0$. Given these beliefs, however, they will buy the standard product at any price lower than 1. Thus, the firm has incentives to deviate and increase p_S. From Proposition 4, however, as $p_R \rightarrow p_S$ every consumer will find optimal to consume the superior variety.

2. Competitive markets for both varieties.

We now consider the case in which both varieties are manufactured by competitive industries so that prices equal the marginal costs ($p_S = c_S, p_R = c_R$) and their profits are equal to zero. Thus, the game is reduced to a simultaneous game in which consumers decide their consumption taking p_S, p_R and Θ_R as given. Assuming that $1 > c_S > c_R$ and following Propositions 1 and 2, if beliefs are exogenous, there are three possible Nash equilibria:

I. Scenario S. If $x > \frac{c_S - c_R}{\delta \alpha \bar{\theta}}$, every consumer in the economy buy S.

II. Scenario RS. If $\frac{c_S - c_R}{\delta \alpha \bar{\theta}} > x > \frac{c_S - c_R}{(\bar{a}+\delta)\alpha \bar{\theta}}$, then the y consumers with $a = \bar{a}$ buy S and the $1 - y$ consumers with $a = 0$ buy R.

III. Scenario R. If $x < \frac{c_S - c_R}{(\bar{a}+\delta)\alpha \bar{\theta}}$, every consumer in the economy buy R.

If beliefs are endogenous, however, the possible Nash equilibria are:

I. Scenario S. If $x > \frac{c_S - c_R}{\delta \alpha \bar{\theta}}$, then there is a Nash equilibrium in which every consumer in the economy buy S.

II. Scenario RS. If $\frac{c_S - c_R}{\delta \alpha \bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta y}{\alpha(\bar{a}+\delta)} > x > \frac{c_S - c_R}{\delta \alpha \bar{\theta}}$, then there exists a Nash equilibrium in which the y consumers with $a = \bar{a}$ buy R and the $1 - y$ consumers with $a = 0$ buy S.

III. Scenario R. Finally, if $x < \frac{c_S - c_R}{\delta \alpha \bar{\theta}}$ there is an equilibrium in which every consumer buy R.

Note that for $\frac{c_S-c_R}{\delta\alpha\bar{\theta}} \cdot \frac{\alpha(\bar{a}+\delta)+\bar{a}\beta y}{\alpha(\bar{a}+\delta)} > x > \frac{c_S-c_R}{\delta\alpha\bar{\theta}}$ both Scenarios S and RS are feasible, depending on whether consumers belief that the rest of consumers of type $a = \bar{a}$ are choosing S (so that $\Theta_R = 0$) or R ($\Theta_R = y$).

3. Competitive market for the standard variety only

Finally, we consider the case in which there is a monopolistic firm manufacturing que superior variety S, while the price of the regular variety is determined in a competitive market so that $p_R = c_R$. There are several possible Subgame Perfect Equilibria:

I. Scenario S. If $c_S < c_R + \delta\alpha x\bar{\theta} - \frac{y}{1-y}\bar{a}\alpha x\bar{\theta}$, the monopolistic firm set a price $p_S = c_R + \delta\alpha x\bar{\theta}$ and all consumers buy S. The monopolistic firm's profit is $\pi_S = c_R + \delta\alpha x\bar{\theta} - c_S$.

II. Scenario RS. If $c_R + (\bar{a}+\delta)\alpha x\bar{\theta} > c_S > c_R + \delta\alpha x\bar{\theta} - \frac{y}{1-y}\bar{a}\alpha x\bar{\theta}$, the monopolistic firm sells S only to consumers of type $a = \bar{a}$ at a price $p_S = c_R + (\bar{a}+\delta)\alpha x\bar{\theta}$ and makes a profit $\pi_{RS} = y[c_R + (\bar{a}+\delta)\alpha x\bar{\theta} - c_S]$

III. Scenario R. If $c_S > c_R + (\bar{a}+\delta)\alpha x\bar{\theta}$, the monopolistic firm does not produce S, all consumers buy R and firm's profits are $\pi_R = 0$.

On the other hand, if beliefs are endogenous, the possible Subgame Perfect Equilibria are:

I. Scenario S or RS. If $x > \frac{c_S-c_R}{\delta\alpha\bar{\theta}}$, the firm sets $p_S = c_R + x\delta\alpha\bar{\theta}$. All the consumers of type $a = 0$ buy S. There are two possible equilibria, depending on whether all consumers of type $a = \bar{a}$ buy S ($\Theta_R = 0$) or R ($\Theta_R = y$). The firm's profits are $\pi_S = c_R + \delta\alpha x\bar{\theta} - c_S$ and $\pi_{RS} = (1-y)(c_R + \delta\alpha x\bar{\theta} - c_S)$, respectively.

II. Scenario R. If $x > \frac{c_S-c_R}{\delta\alpha\bar{\theta}}$, the firm does not sell S and makes no profit.

Two important remarks can be extracted from these results: (i) relative to the case in which the firm monopolizes the industry of both varieties, its profits are now reduced. This is hardly surprising, since both products are substitutes and now one of them is manufactured by a competitive industry; (ii) the firm's profits are never larger if beliefs are endogenous. This contrasts with our previous results (see Proposition 6). The reason is pretty simple, however. In the models considered here, optimal beliefs make more likely that individuals that experience anticipated utility underestimate the expected damage associated to the consumption of R. In that case, their willingness to pay for R will be larger than expected according to the models with exogenous beliefs $\hat{x} = x$. A firm may take advantage of this to the extent that it has some market power in the market of R. This is precisely the case in the model presented in Section 3. In this alternative version, however, the firm cannot influence p_R and therefore it cannot take advantage from consumers' self-deception. On the contrary, the consumer's optimism regarding the consequences of consuming R may limit the maximum price it can charge for S. Finally (iii) similar to our previous results, it remains unclear whether self-deception leads necessarily to a larger consumption of R in the economy. If $c_R + \delta\alpha\bar{\theta}x > c_S > c_R + \delta\alpha x\bar{\theta} - \frac{y}{1-y}\bar{a}\alpha x\bar{\theta}$, then the Scenario S ($\Theta_R = 0$) is a Subgame Perfect Equilibrium if beliefs are endogenous but not if they are exogenous. This is because, in that case, the firm has incentives to sell S to individuals that experience anticipated utility ($a = \bar{a}$), who are more willing to pay for S under the assumption of exogenous beliefs.

Chapter 2

An experimental test of some economic theories of optimism

1. Introduction

Numerous studies show that people *sometimes* have 'too' optimistic beliefs about self-relevant events and future material outcomes, as if their beliefs were influenced and aligned with their desires or preferences –for reviews, see Bénabou and Tirole (2016); Epley and Gilovich (2016); Kunda (1990); Wicklund and Brehm (1976). *When such a positivity bias is observed*, further, there is strong evidence that it is caused by asymmetric updating, that is, the under-weighting of undesirable information relative to the desirable one –e.g., Eil and Rao (2011); Möbius et al. (2011); Sharot et al. (2011). Wiswall and Zafar (2015), for instance, report that undergrads update their beliefs about their own future earnings asymmetrically: After learning the actual average earnings for each major, their original beliefs about self-earnings are slightly revised downwards when their prior estimation of average earnings was too positive, whereas the corresponding upward revision is more substantial if they underestimated the actual mean value. On the other hand, it is yet far from clear which environmental and individual factors make people (more) skewed: People fail into positivity biases, but it is not well-known *when* they do so, and *who* is more likely to do so.[10]

Understanding which factors or conditions are more propitious for the formation of optimistic beliefs is important for at least two reasons. *First*, these beliefs can motivate

[10] A similar point has been stressed by Benjamin (2019) in his comprehensive review about biases in beliefs.

suboptimal decisions which sometimes may lead to undesirable collective outcomes, as the next examples illustrate. **I**: If people want to believe that they are often right in their presumptions and hence commit few mistakes, a confirmation bias follows. That is, agents underweight disconfirming evidence and overweight confirming evidence, e.g., Eil and Rao (2011). As a result, voters may develop partisan biases and polarized beliefs on issues such as climate change (Sunstein et al., 2016), as well as a higher credulity for fake news in line with their prejudices. **II**: Optimism may lead to overinvestment during economic booms and boost the formation of financial bubbles (Aliber and Kindleberger, 2015; Shiller, 2000). As an illustration, a few years before the subprime mortgage crisis, a survey conducted by Case and Shiller (2003) to new homeowners in different cities in the US showed that 90 percent of them believed that housing prices in their cities would keep increasing for the next 10 years at an average estimated rate between 9 and 15%, depending on the city. **III**: Oster et al. (2013) study testing among individuals at risk for Huntington disease –a degenerative neurological disorder associated to a genetic alteration. One of their main findings is that those who reject testing generally underestimate their actual risk and behave as if they do not have the disease –diagnosed individuals tend to behave differently: they are more likely to retire, make major financial changes or change their recreation habits. More speculatively, but also in the public health realm, many government's initial reaction (or lack of it) to the Coronavirus threat in January-February 2020 suggest some optimism –e.g., the virus will not arrive, and hence there is no need to buy in advance enough personal protective equipment for the healthcare workers in the public hospitals in case it comes.

Second, there is also a positive side of optimism. Beliefs about the future, the morality of our acts, or about oneself can trigger different emotions like hope, pride, anxiety, guilt, or shame. In this sense, individuals may prefer information and beliefs that contribute to emotional well-being. In fact, optimism has been associated to a good mental and physical health (Rasmussen et al., 2009; Strunk et al., 2006). Also, optimistic beliefs about one's abilities can motivate the individual to undertake difficult tasks and to overcome the obstacles that can arise (Bénabou and Tirole, 2002, 2004). Further, self-confidence can also help to convince others (von Hippel and Trivers, 2011). To sum up, optimism has arguably cons and pros and, depending on the context, we may wish to deter or promote such bias. That requires, however, a precise understanding of the environmental, individual, and institutional conditions leading to more optimism (or more realism).

In this endeavor to find answers for the *when* and *who* questions cited above, models are invaluable tools, as they offer insights, allow precise policy design, and organize the analysis. Several economic theories of optimism have been proposed, and they give different answers to those questions. Our goal here is to test the predictions of two families of theories by experimental means. A common hypothesis in all these models is that individuals have implicit preferences over the possible states of nature and derive utility (disutility) from thinking that their preferred state is (not) true. If people experience anticipatory feelings like excitement, joy, fear or anxiety from thinking on some future uncertain events, for instance, they would rather believe that the future will be favorable, so as to trigger the relatively more positive emotions. Another idea common to these models is that human inference operates *as if* people chose their beliefs, although most likely through a subconscious mechanism. This means that belief acquisition responds to incentives and constraints, as in a usual economic decision like, e.g., buying some good. The families differ mostly in the specific factors that restrain optimism and prevent individuals' beliefs from departing too much from reality. In models like Akerlof and Dickens (1982) and Brunnermeier and Parker (2005) –AD and BP henceforth, respectively–, the ‘demand’ for a biased belief depends on its material price: If biases lead to sufficiently costly or risky decisions, individuals may be more reflective, cooling down their expectations. In Rabin (1994) and Bracha and Brown (2012), alternatively, self-deception requires the individual to selectively avoid or look for information, to rationalize it or to restraint certain thoughts. To formalize this idea, these models explicitly incorporate cognitive costs associated to belief distortion, which increase the further away from Bayes’ rule the beliefs are.

While there is plenty of evidence of situations in which individuals seem to be optimistic, the specific implications of these models have received much less attention. As Coutts (2019b, p. 549) notes, “rigorous tests of existing theory and direct evidence about optimism are scarce”, and the evidence collected so far is not conclusive (see Section 2 below for a review). This paper hence contributes to this literature with a lab experiment. On the adequacy of this methodology, we note that, while field studies offer extremely suggestive evidence from a naturalistic setting, they can rarely assure full control of the agents’ relevant priors and evidence observed. As a result, it is difficult to establish if, say, an optimistic prediction is the result of some bias in the updating procedure or caused by the very positive evidence that the person has received or her skewed priors. More generally, field studies cannot be used to test many of the predictions of the existing theories of motivated inference. Lab studies, in contrast, permit such fine-grained tests.

In a nutshell, participants in our experiment have to estimate the frequency $\theta \in [0, 1]$ of some event after observing a series of i.i.d. signals. Importantly, subjects have an incentive to believe that θ is as high as possible, as they can get a *state prize* that increases with θ (this prize does not depend on the subject’s choices, as θ is randomly determined). Each subject makes three estimations of θ, each one with a progressively enlarged dataset.

These estimations are unexpected by the subject and separated in between with some distraction tasks. While the first two estimations are not incentivized, the subject can get an *estimation prize* of 10 euros if her last estimation is sufficiently accurate –to prevent hedging, in fact, subjects could get either the state or the estimation prize, randomly determined with probability 0.5 at the end of the experiment. After round 3, furthermore, subjects must provide the shortest credible interval for θ that contains 95% of the probability mass (this was not incentivized).

The models by AD and BP predict that (some) individuals will give higher estimates of θ than the Bayesian estimate, so as to correspondingly 'inflate' the state prize. This is particularly true when there are no potential losses associated to inaccurate beliefs, as in the first two estimation rounds, where individuals get no prize for accuracy. Since subjects in our experiment can only get either the state or the estimation prize, further, AD and BP predict a correlation between risk aversion and Bayesianism: relatively more risk averse types would rather have realistic beliefs, so as to maximize the likelihood that they get something in case the estimation prize is selected for payment.

Overall, however, we find scarce evidence supporting the models considered here, and particularly for models like AD and BP. Specifically for these two models, *first*, we do not observe systematic optimism, i.e., overestimation, in any of the three rounds. Indeed, the average and median subject slightly underestimates θ in every round. *Second*, the size of the estimation bias is not reduced by the introduction of incentives for accuracy in the third round. While there are subjects who never underestimate θ, *third*, they account to just 26.47 percent of the sample (N = 68) and inflate θ to a rather limited extent. *Fourth*, optimism is unrelated to any of the individual characteristics that we record, except a relatively higher CRT score. That is, the 'optimistic' subjects look scarcely different from the others, particularly in terms of risk aversion. However, *fifth*, we find that overestimation in our experiment is extremely correlated with the sample observed: Optimists (pessimists) tend to be subjects who observe relatively few (many) female extractions. This is again hardly consistent with models like AD and BP. In what regards the confidence intervals, *sixth*, the model by BP says that optimistic subjects will report 'positively skewed' intervals, i.e., the subject's estimate of θ is the lower limit of the interval. No evidence supports this prediction, though. Note that models with cognitive costs like Rabin (1994) are also inconsistent with the first, fourth, and fifth findings just cited. They are in line though with the second finding and *perhaps*, conditional on the cognitive cost function assumed, with the limited extent of the biases observed in our experiment, i.e., our third finding (we have not analyzed what these models predict for the confidence intervals). Overall, therefore, this second family appears to fit better with our results, at least in relative terms.[11]

The rest of the paper is organized as follows. The next section discusses some related experimental literature and our contributions to it. Section 3 introduces the experimental design. Section 4 starts by presenting the predictions of the Bayesian model, as well as applying AD and BP to our setting. This section also reports experimental results afterwards and discusses the models with cognitive costs. Section 5 concludes by mentioning potential future venues of research.

2. Literature review

To organize this survey, we mention in what follows several predictions of the theories and how they compare with the existing experimental evidence so far.[12] The first

[11] The model in Mayraz (2013) assumes no costs to belief distortion and is hence consistent as well with our second finding, but not with our other findings (leaving aside the sixth one, which we have not checked).

[12] This review cannot make justice to the whole literature in this respect, but consult Caballero and López-Pérez (2020a) for a fuller review of the literature on motivated inference and optimism. See also

prediction refers to the very phenomenon of optimism and hence is common to all of the models considered here. Note yet that the models predict this phenomenon under different conditions, to be specified later. In addition, some models like Möbius et al. (2014) explicitly analyze how optimists infer, explaining belief inflation as the result of asymmetric updating. As explained in the introduction, the idea is that signal observations are over-weighted or under-weighted depending on whether they support or contradict, respectively, the decider's desired beliefs.

Prediction 1: Deciders inflate their beliefs. Specifically, the difference between the subjective and the objective probabilities is correlated with the utility payoff that deciders get from having the desired beliefs about the state space. Inflation occurs because individuals process 'good' and 'bad' news asymmetrically, thus reinforcing their favorite beliefs.

Evidence: Several studies report inflated beliefs and find evidence of asymmetric updating as a potential cause, but not all studies do so. In line with our discussion in the introduction, our interpretation is that inflation requires propitious conditions, still not well understood. Two groups of experimental studies can be perhaps distinguished for the sake of the exposition. In a first one, the beliefs analyzed are arguably relevant for self-esteem. When eliciting a subject's posterior distribution about her rank in a group according to some ego-relevant trait (specifically, physical attractiveness and/or IQ score), for instance, Eil and Rao (2011) and Möbius et al. (2014) find evidence of positively skewed updating (see also Heger and Papageorge, 2018). Specifically, people seem to update beliefs according to Bayes' Rule when the signal is good or desirable, and under-update when the signal is negative. In contrast, Ertac (2011) reports negatively

Caballero and López-Pérez (2020b) for a review of the existing literature on the relation between selective recall and optimism.

skewed updating, while Buser et al. (2018) find no evidence at the aggregate level for asymmetric updating about relative performance. See also Zimmermann (2020), who finds evidence of underweighting of the negative signals when the posteriors are elicited one month after feedback, but not when they are elicited immediately after. Grossman and Owens (2012), in turn, explore learning about *absolute* performance and find no evidence of asymmetric updating.

In a second group of economic experiments, closer to our study, subjects have a financial stake in some specific event E and must report the posterior that E occurs after observing some relevant evidence –Gotthard-Real (2017), the Baseline condition in Barron (2020), Coutts (2019a); and Heger and Papageorge (2018).[13] Little evidence of a positivity bias has been found in the studies just cited, in spite of the fact that they display several differences in what regards the priors on E, the prize if E occurs, or the randomization mechanism, e.g., mechanical or using a computer program. In addition, Coutts (2019a) studies inference in a "value relevant" treatment, i.e., when subjects have a preference for some event e to be true, and a "neutral" one, and finds no differences in belief updating across treatments. This occurs when e is financially relevant, i.e., the E above, but also when it is ego-relevant, i.e., the ranking in a math or verbal quiz.

To our knowledge, there are just two lab studies in Economics reporting a motivated bias when subjects have a financial interest for some state. Both involve between-subjects designs, so the "bias" comes from the fact that one role is more positive than another one, given similar information but different preferences. In Mayraz (2013), subjects are shown a chart of historical wheat prices and have to predict afterwards the price at some future time point, getting a bonus for accuracy. In addition, subjects get a

[13] Some recent studies do not fit exactly within any of the two groups considered. Engelmann et al. (2019), for instance, report that subjects under-estimate the probability of receiving an electric shock that is outside of their control

payoff that increases (decreases) with that price if her randomly-selected role is "Farmer" ("Baker"). In average, farmers make significantly higher predictions than bakers, consistent with a positivity bias. In the Strategic condition of Charness and Dave (2017), in turn, subjects play a 2x2 game. The payoff matrix is a priori uncertain, as there are two possible payoff constellations. While subjects in the Odd role get the same equilibrium payoff in both matrices, those in the Even role have a preference for one of them. Prior to playing the game, participants observe a sequence of six signals and their incentivized posteriors of each state/matrix reveal that Even players underweight more strongly the negative signals, i.e., those confirming the 'worst' matrix.

Neuroscientists and psychologists have also gathered some supportive evidence for inflation and asymmetric updating in beliefs about future outcomes. In Sharot et al. (2011), participants are sequentially presented a total of 80 adverse events, such as being diagnosed with Alzheimer's disease or suffering a car accident, and have 6 seconds to estimate their chances of facing any such event in the future (without incentives). In a second stage, subjects are shown for 2 seconds the actual frequency with which any such event happens among individuals living in the same socio-cultural environment as them and must guess their posteriors of encountering that event. Sharot et al. (2011) report evidence for asymmetric updating in favor of good news. Using a similar design, Ma et al. (2016) find that intra-nasally administered oxytocin promotes optimism and asymmetric updating. This is particularly true in individuals with high depression or anxiety traits, who under-weight undesirable feedback more pronouncedly than similar individuals in a placebo treatment –on how depressed individuals update beliefs, possibly in a relatively more balanced manner, see also Alloy and Abramson (1979) and Garrett et al. (2014). ■

The next prediction follows from models like BP. The idea is that people will be *less* biased when there is more risk. Considering that having inaccurate beliefs often leads to suboptimal choices, that is, the models propose that belief inflation will be attenuated when acting as an optimist leads to large expected losses (relative to a Bayesian). Intuitively, people think more when there is a lot at stake, and hence their beliefs are dominated relatively less by their "animal spirits".

Prediction 2: The correlation between beliefs and preferences will get weaker as the expected material loss for holding inaccurate beliefs increases.

Evidence: several of the papers cited above have systematically analyzed whether the size of the expected loss reduces the degree of inflation. Coutts (2019a,b) runs sessions with different accuracy payments, i.e., low ($3), medium ($10), or high ($20). In addition, participants can either get a nil or high prize ($80) if some target event E occurs. According to Prediction 2, subjects have no incentive to distort their beliefs about the probability of E in case the prize is $0, provided that they get no other utility from the occurrence of E, e.g., if E is not ego-relevant. In contrast, distortion should be maximal if the prize is high and the accuracy payment low, i.e., $3. Despite this, the author concludes that neither prizes nor accuracy payments alter updating. Mayraz (2013) varies the size of the accuracy bonus from £1 to £5 and reports that the magnitude of the bias does not depend on the scale of the bonus (neither on the size of the prize associated to the desirable event). Similarly, Ertac (2011) studies the effect of rewarding accurate beliefs and finds no significant difference across compensated and non-compensated sessions in terms of the distribution of priors and the absolute value of the bias (Mann–Whitney test, $p = 0.89$ and $p = 0.79$, respectively). See also Engelmann et al. (2019) for similar negative results regarding the accuracy payment. If people suffer from a positivity bias, in summary, the cost of such bias does not seem to set limits on its size, at least with the parameterizations that have been considered so far. ■

In some models, belief distortion is assumed to be cognitively challenging and hence involving an explicit cost ─Bracha and Brown (2012), Rabin (1994). For instance, Rabin (1994, p. 180) contends that "developing beliefs that differ from this level [of natural, intellectually honest beliefs] is costly because it may intrinsically conflict with other parts of a person's belief system, and reintegrating it can involve laborious intellectual activity."

Prediction 3: Less inflation when it is cognitively costly.

Evidence: To our knowledge, the experimental literature has not dealt thoroughly with this question yet. Coutts (2019b) compares and tests some predictions of BP's model of optimal expectations and the model of affective decision making proposed by Bracha and Brown (2012), which includes explicit mental costs from belief distortion. While Coutts (2019b) finds limited evidence supporting some of the implications of the model by Bracha and Brown (2012), its work focuses on the effects of state-dependent prizes and accuracy prizes on optimistic bias, and not directly on the cognitive aspects of belief distortion.

3. Experimental design

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 –these terms were not used in the subjects' instructions; see Appendix I. The precise rate θ of female balls 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 $1 - \theta$. It follows that the number of female balls $F \in [0, 100]$ is equal to $100 \cdot \theta$; we will make reference generally to θ for consistency, although the instructions were expressed in terms of F. Although the subject does not know θ, the method to determine it is known in advance.[14] 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. 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. Subjects were explained 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. 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 $\theta = 1$. 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 $|\theta - \hat{\theta}|$ is smaller or equal to 0.02, and 0 euros otherwise. The elicitation of the first two estimations of θ, in turn, is not incentivized. Participants are informed about the nature of the 'state prize' before they observe any extractions, whereas 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

[14] 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. The lists, elaborated by the Spanish National Statistics Institute, order the names according to frequency; see https://www.ine.es/en/welcome.shtml. Once θ had been randomly determined for a subject, therefore, we randomly selected $100 \cdot \theta$ girl names and $100 \cdot (1-\theta)$ boy names in the corresponding lists to 'fill' the urn.

probability 0.5 they would get either the 'state prize' or an undefined prize whose nature would be specified later (this design choice is irrelevant to test the theories considered here, but relevant for the analysis in Caballero and López-Pérez (2020b)).

Additional tasks and questions are inserted *between* some extractions. 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. In other words, they indicated a lower and an upper bound for θ, such that they believed that the correct θ was 'almost surely' in the interval determined by those limits. Confidence intervals were not incentivized; as we discuss later, however, our results do not differ much from those in López-Pérez et al. (2020), where subjects were paid for accuracy. After this interval estimation, additionally, we included an incentivized 'recall task' and two questions so as to check whether they expected to recall better female than male extractions, i.e., good than bad news; this data is irrelevant for the test of the theories considered here, but see Caballero and López-Pérez (2020b) for a full description and analysis. Subjects responded, in addition, two questions on statistical knowledge, the LOT-R test on optimism (Scheier and Carver, 1985; Scheier et al., 1994), and a test on disappointment, in this order, thus ending the experiment.

[15] Subjects faced the choice between lottery A with prizes 2 and 1.5 Euros and lottery B with prizes 4 and 0 Euros, with equal probabilities of the larger and lower prize across lotteries. Letting P denote the probability of the larger prize, they had to indicate the threshold value of P such that they always preferred B to A, on a scale from 0 to 100.

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 (translated to English in Appendix I). 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 the screenshot in Appendix I). 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.

Discussion

While appropriate for the test of the theories considered in this paper, the elicitation of a subject's $\hat{\theta}$, i.e., the mode of her posterior beliefs, is a rather unusual feature in the literature on belief updating, specifically on motivated inference, where the subject's posterior probability distribution is often elicited instead –e.g., using the lottery method as in Coutts (2019a), or the crossover method in Möbius et al. (2014). In a sense, we elicit an ordinal instead of a cardinal measure of probability. We introduced this relatively novel aspect for three reasons. First of all, we found the question of how people compute empirical frequencies when they have a preference for some states/values an interesting one in itself. Second, incentive compatible elicitation procedures are often complex to explain to subjects, e.g., Schlag et al. (2015). In contrast, our estimation prize is rather straightforward. Third, we suspected that the computation of the exact probability of any rate was a substantially more demanding problem than the estimation of $\hat{\theta}$, which requires only extrapolating from the sample (see 4.1). As noted by Tversky and Kahneman (1983) and Gigerenzer and Hoffrage (1995), posing problems in frequentist (as opposed to probabilistic) terms may mitigate some errors. In summary, our design attempted to reduce any potential noise due to the subjects' misunderstanding of the elicitation procedures or the statistical nature of the problem. The drawback is that we lose rich information on their posteriors, although the confidence interval estimation offers some insights. Note also that, since we do not elicit the precise posteriors, it is not our research goal to analyze whether people update their probabilistic beliefs in a conservative manner, or display asymmetric updating. Yet these are issues that have received attention before, as we have explained in Section 2.

For another remark, note that subjects are never paid for both their beliefs and the actual state θ. Otherwise we might face hedging problems and hence the subjects' potential misreporting of their beliefs in the incentivized elicitation (Blanco et al., 2010). To understand this, suppose for the sake of the exposition that a subject can get both prizes and believes that $\theta = 0.9$ with probability $p > ½$ and $\theta = 0.4$ with probability $1- p$. If she reports $\hat{\theta} = 0.9$, therefore, she expects with probability p a state prize of $50 \cdot 0.9 = 45$ Euros plus an estimation prize of 10 Euros, and a payoff of $50 \cdot 0.4 = 20$ Euros with probability $1- p$. A report of $\hat{\theta} = 0.4$, on the other hand, generates a lottery with payoffs of 45 and 20 + 10 with respective probabilities p and $1- p$. It follows that a sufficiently risk averse subject would rather report $\hat{\theta} = 0.4$, so as reduce variability. More generally, the Bayesian prediction would be conditional on the subject's degree of risk aversion if there were hedging problems; our design prevents this kind of complexities.[16] In any case,

[16] Subjects also get a payoff in the recall task, but this is introduced after the estimation task.

we note incidentally that the study cited in Section 2 by Ertac (2011) does not find strong evidence in support for hedging in her study.

Finally, the beliefs were incentivized only in one round, i.e., the last one. The rationale under this design choice is multiple. On the one hand, the models of optimism described in Section 4 say that subjects will inflate more when the 'price' of inflation, i.e., the potential monetary loss for being inaccurate, is nil (see Proposition III below). The models therefore predict more overestimation in the non-incentivized rounds. A potential objection against not incentivizing some rounds is that subjects could give little thought to the issue. We note however that these are exactly the type of situations in which BP intuitively predict more optimism and, more substantially, we can compare our results across rounds and hence check whether incentives reduce noise (anticipating somehow our results, the evidence is negative in that respect). The reason to incentivize precisely the *last* round, further, was to avoid ambiguous predictions by the models of optimism. If subjects were paid instead in the first round only, for example, in posterior rounds they would simultaneously wish to believe that (i) the share of female balls is high, but also that (ii) their estimate in the first round was accurate. In these circumstances, it would not be clear whether, according to models of optimism, we should expect optimistic estimates or some degree of anchoring relative to the first estimation. The incentive structure of our design, in contrast, tries to guarantee that only incentive (i) is present in the first two rounds, while both (i) and (iii) a desire for accuracy are relevant in the last round.

4. A test of several theories

In this section, we first introduce in 4.1 the Bayesian model. Afterwards, we apply AD and BP to our experiment and derive a series of predictions, which guide the posterior analysis of the experimental evidence. The discussion concerning models with cognitive costs comes at the end of this section.

4.1 Models and predictions

We start by introducing some general notation, together with the standard Bayesian theory.

General setup & the Bayesian model

An expected payoff-maximizer called Eve must estimate the frequency/rate $\theta \in [0, 1]$ with which a 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. Let p_k denote Eve's priors about rate $\theta_k \in \Theta$. In our experiment, $\Theta = \{0, 0.01,\ldots, 1\}$, whereas the (uniform) prior of any rate θ_k is $p_k = 1/101$.

Eve has observed some realizations of S and hence can use that evidence to update her priors. The number of female observations is denoted as f and that of male ones as m. 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):

$$p_{k|D} = \frac{p_k \cdot \theta_k^f \cdot (1-\theta_k)^m}{\Sigma_{\theta_i \in \Theta}\, p_k \cdot \theta_i^f \cdot (1-\theta_i)^m} = \frac{\theta_k^f \cdot (1-\theta_k)^m}{\Sigma_{\Theta}\, \theta_i^f \cdot (1-\theta_i)^m} \tag{1}$$

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.[17] When the point estimations are not incentivized, finally, the argument is analogous except that no distortion is here expected for any value of f. Our first result is hence direct.

Proposition I (Bayesian): In each estimation round, a Bayesian subject who has observed so far 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 in the last estimation round, where some slight distortion is predicted if the sample observed contains extremely few or extremely many female balls. In average, point estimations do not significantly differ from the average empirical frequency.

Choice of beliefs: Applying AD and BP to our experiment

We maintain the notation introduced in the general setup, and consider an agent called Abel, identical in all respects to Eve except Bayesian updating, i.e., equation (1) above. If Abel participates in our experiment, specifically, any estimation round is

[17] To clarify, think of the case in which the sample contains 30 female balls. The most likely value of θ is 1. If Eve reports an estimate of 1, however, she would eventually earn prize 2 only if the true rate is 0.98, 0.99 or 1. On the other hand, if her estimate is 0.98, she earns prize 2 if the true rate is between 0.96 and 1, both included. A further subtlety is that the distribution of posteriors is not symmetric in general, and particularly for the samples considered here. If Eve observes 1 female ball, specifically, the mode is 0.03, but interval [0.01, 0.05] has less aggregate probability than [0.02, 0.06]. In this case, therefore, Eve should report $\hat{\theta} = 0.04$. A similar argument applies when there are 29 female balls in the sample. There are no other cases where a rational Bayesian should report an estimate different than the mode. These "distortions" could be prevented if the estimation prize required an absolutely correct estimate of θ. Since this could reduce a subject's incentive to exert attention on this task, however, we tried to achieve an equilibrium. Note also that the optimal estimation of θ depends on the structure of the estimation prize. A different set of incentives could imply that the optimal point estimate is a different statistic than the mode, like the mean, the median, etc.

conceived as divided in two periods; in period 1, Abel chooses his subjective beliefs about θ,[18] while he opts for the corresponding estimate $\hat{\theta} \in \Theta$ in period 2, based on those beliefs. Let $\hat{p}_k$ denote the subjective probability of state θ_k as chosen in period 1. Abel can choose any set of subjective probabilities, provided that they satisfy Kolmogorov's probability axioms.

Abel's problem in any round can be solved recursively. In period 2, he chooses the estimate that maximizes his expected monetary payoff, based on his subjective beliefs. If we abstract for simplicity from the payoff in the recall task, Abel's payoff equals either the state prize or the estimation prize; both prizes have equal probability. The state prize, recall, is proportional to the share of female balls, i.e., equal to $M \cdot \theta$, where $M = 50$ in our experiment. With respect to the estimation prize, we simplify matters by assuming that it amounts to zero unless estimation $\hat{\theta}$ exactly matches the actual state of the world (i.e. $\hat{\theta} = \theta$), in which case it equals $\bar{v}$ (implicitly, $\bar{v} = 10$ in the third, incentivized round, but $\bar{v} = 0$ in the first two rounds). Further, let $u(x)$ denote the utility function of money, where x indicates the monetary gain; we posit $u(0) = 0$. In period 2, to sum up, Abel chooses $\hat{\theta}$ so as to maximize expected utility function

$$\frac{1}{2}\sum_{\theta_k \in \Theta} u(M \cdot \theta_k) \cdot \hat{p}_k + \frac{1}{2}u(\bar{v}) \cdot \hat{p}_k(\theta_k = \hat{\theta}) \qquad (2)$$

It comes straightforward that (2) is maximized by choosing the $\hat{\theta}$ that matches the most likely state of nature according to the subjective probabilities, i.e. the subjective mode θ_{sm} with subjective probability $\hat{p}_{sm}$ such that $\hat{p}_{sm} \geq \hat{p}_k$ for any $\theta_k \in \Theta$. If the posterior subjective distribution has several modes, Abel is indifferent between them; in

[18] To reduce degrees of freedom and for simplicity, we posit that subjects in our experiment trust the experimenter's instructions although, formally speaking, the models here allow subjects to choose their beliefs in this respect as well –e.g., believing that the probabilities of earning either the state or the estimation prize are not the same.

this case, θ_{sm} denotes the mode chosen in period 2. Let p_{sm} denote the objective probability of θ_{sm} –note well that p_{sm} does not represent the objective probability of the objective mode θ_{om}, since θ_{sm} is not necessarily equal to θ_{om}. The corresponding probabilities $\hat{p}_{om}$ and p_{om} are analogously defined for the objective mode, which in our problem is unique for any data D observed (see Proposition I).

In period 1, optimal beliefs are chosen. Following AD and BP, we assume that Abel may experience some anticipatory utility at the end of period 1. That is, he gets utility from thinking about his future material payoff. This anticipation utility depends on the specific parameters described above and his current beliefs, i.e. the subjective probabilities. In period 1, that is, Abel chooses the beliefs that maximize

$$a\tfrac{1}{2}\left[\textstyle\sum_{\Theta} u(\mathrm{M}\theta_{\mathrm{k}}) \cdot \hat{p}_{\mathrm{k}} + u(\bar{v})\hat{p}_{sm}\right] + \tfrac{1}{2}\left[\textstyle\sum_{\Theta} u(\mathrm{M}\theta_{\mathrm{k}}) \cdot p_{\mathrm{k}} + u(\bar{v})p_{sm}\right] =$$

$$\tfrac{1}{2}\textstyle\sum_{\Theta} u(\mathrm{M}\theta_{\mathrm{k}})(a\hat{p}_{\mathrm{k}} + p_{\mathrm{k}}) + \tfrac{1}{2}u(\bar{v})[a\hat{p}_{sm} + p_{sm}] \qquad (3)$$

where $a \geq 0$ reflects the intensity of the anticipatory utility. Note that expression (3) has four components. Two of them refer to the anticipatory utility from the state and estimation prizes, i.e., $\sum_{\Theta} u(\mathrm{M}\theta_{\mathrm{k}}) \cdot \hat{p}_{\mathrm{k}}$ and $u(\bar{v})\hat{p}_{sm}$, respectively; they depend on the beliefs chosen in Period 1. Intuitively, the first component is maximized when beliefs put the whole probability mass on $\theta_{\mathrm{k}} = 1$, while the second one is maximized when a single rate receives all the probability mass; it follows that anticipatory utility is maximized when rate $\theta_{\mathrm{k}} = 1$ is certain, i.e., $\hat{p}_k = 1$ for $\theta_{\mathrm{k}} = 1$. In turn, a third component of (3) is the objective expectation of the state prize, $\sum_{\Theta} u(\mathrm{M}\theta_{\mathrm{k}}) \cdot p_{\mathrm{k}}$, which cannot be altered by Abel. Finally, the objective expectation of the estimation prize, $u(\bar{v})p_{sm}$, depends on Abel's estimate, which in turn depends on his beliefs. The size of this component decreases as θ_{sm} moves further from θ_{om}, as this reduces p_{sm}. Several implications follow from this optimization problem. The proofs can be consulted in Appendix III.

Proposition II. If $a = 0$, then any subjective belief such that θ_{sm} coincides with the objective mode θ_{om} is optimal.

Proposition II implies that, in the absence of anticipation utility, only beliefs that do not alter the optimal action in the second period are optimal. Note the difference with Eve's case: Since Abel can choose his beliefs, the optimal ones when $a = 0$ are indeterminate, except that the mode of Eve's and Abel's must coincide.

Proposition III. If $\bar{v} = 0$, then the optimal beliefs are characterized by $\hat{p}_{\mathrm{k}} = 1$ for $\theta_{\mathrm{k}} = 1$ and $\hat{p}_{\mathrm{k}} = 0$ for any $\theta_{\mathrm{k}} < 1$.

This simply states that, if there is no potential loss in keeping distorted beliefs, then it is optimal for Abel to believe that the only possible state is the most favorable. This prediction is relevant for the first two estimation rounds, where there were no incentives for accuracy. The following results offer insights on the third, incentivized round.

Proposition IV. The optimal subjective mode θ^*_{sm} is at least equal to the objective mode. If $a > 0$, further, optimal beliefs are characterized by $\hat{p}_{\mathrm{k}} = 0$ for any $\theta_{\mathrm{k}} < \theta^*_{\mathrm{sm}}$.

In our experiment, Proposition IV implies that Abel will never underestimate the number of female balls in the urn and that posteriors will be extremely skewed about its mode, with clear implications on Abel's (subjective) 95% confidence intervals. The rationale for this prediction is quite intuitive. On one hand, the assignment of non-nil probability $\hat{p}_{\mathrm{k}} < \hat{p}_{sm}$ to any rate $\theta_{\mathrm{k}} < \theta^*_{\mathrm{sm}}$ is 'useless': it does not affect the estimate at period 2 and hence the objective probability of getting the estimation prize, and has an opportunity cost, in that $\hat{p}_{\mathrm{k}}$ could be assigned instead to θ^*_{sm}, thus increasing anticipatory utility from both prizes (if $a > 0$). In addition, when Abel chooses beliefs where the subjective mode is different from the objective mode, this comes at the cost of reducing the objective probability of getting the estimation prize. The only incentive for such a choice, therefore, is to sufficiently increase the subjectively expected payoff from the state prize, which requires choosing beliefs with a subjective mode at least equal to θ_{om}.

The remaining propositions study conditions for overestimation, i.e., the subjective mode being higher than the objective one, and for overprecision, which refers in our context to the length of the 95% subjective confidence interval. In very general terms, they express that the position chosen by Abel for the subjective mode has implications in the optimum regarding the level of overprecision. We distinguish two different situations.

Proposition V. Consider beliefs with subjective mode θ_{sm}. If $u(\bar{v}) \geq u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{sm}})$, a necessary (but not sufficient) condition for these beliefs to be optimal is that they assign $\hat{p}_{\mathrm{k}} = 0$ to any $\theta_{\mathrm{k}} \neq \theta_{\mathrm{sm}}$.

Proposition V implies that any beliefs with subjective mode 'close' to $\theta_{\mathrm{k}} = 1$ cannot be optimal if they do not concentrate all the probability mass in that mode. In these circumstances, intuitively, anticipatory utility is maximized when Abel is certain to get the estimation prize. A problem of Proposition V is that it says little about the optimal value θ^{*}_{sm} and cannot be tested using directly observable data. However, Propositions IV and V imply the following corollary, which states a sufficient (although not necessary) condition in the optimum for maximal overprecision, i.e., a degenerate belief distribution. This condition is based on the value of the objective mode θ_{om}, which is determined by the evidence available to Abel and hence observable.

Corollary 1: If $u(\bar{v}) \geq u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{om}})$, Abel assigns all the probability mass to one single rate. Its specific value depends on the curvature of the utility function of money, u(x), but also on a. In particular, Abel is Bayesian, i.e., $\theta_{\mathrm{sm}} = \theta_{\mathrm{om}}$, if a is low enough or if the utility of money increases at a sufficiently lower rate than the posterior beliefs (as we move towards the objective mode).

To clarify how stringent condition $u(\bar{v}) \geq u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{om}})$ is, note that $\bar{v}$ = 10 and M = 50 imply that the condition is necessarily satisfied if Abel is risk-averse and $\theta_{\mathrm{om}} > 0.8$. However, the condition can be also fulfilled for much lower values of θ_{om} if Abel displays sufficiently high levels of risk-aversion. Let θ_{A} denote the rate such that $u(\bar{v}) = u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{A}})$. Corollary 1 says that, if $\theta_{\mathrm{om}} \geq \theta_{\mathrm{A}}$, Abel is extremely overprecise, in the sense that he believes that there is only one possible state of nature. When $\theta_{\mathrm{om}} < \theta_{\mathrm{A}}$, in contrast, the following prediction shows that Abel's posteriors can be more spread. In other words, beliefs can be nondegenerate only if $\theta_{\mathrm{om}} < \theta_{\mathrm{A}}$. The degree of overprecision, in other words, is conditional on the evidence received. In average, confidence intervals will be larger when θ_{om} is relatively small, in particular when θ_{om} < ½.

Proposition VI. Consider beliefs with subjective mode θ_{sm} such that $u(\bar{v}) < u(\mathrm{M}) - u\big(\mathrm{M}\theta_{\mathrm{sm}}\big)$. Conditional on risk aversion, a necessary (but not sufficient) condition for these beliefs to be optimal is that they assign evenly as much probability as possible to some of the largest state(s), subject to $\hat{p}_{\mathrm{sm}} > \hat{p}_{\mathrm{k}}$ for any θ_{k}.

While Proposition VI is not very specific about the optimal θ^{*}_{sm}, a similar argument to that in Corollary 1 seems to apply as well for low values of θ_{om}. That is, again, $\theta^{*}_{\mathrm{sm}} = \theta_{\mathrm{om}}$ if a is low enough or if the utility of money increases at a sufficiently lower rate than the posterior beliefs (as we move towards the objective mode). In summary, risk aversion correlates with more Bayesian estimations (assuming that Abel's degree of risk aversion is independent of his a). Intuitively, a subjective mode larger than θ_{om} increases the anticipatory utility from the state prize but reduces the chances of earning anything if the estimation prize is finally the selected one. Obviously, a very risk-averse Abel strongly dislikes such possibility. Note also that, since θ_A depends on Abel's degree of risk aversion, a relatively risk-averse Abel would also be less likely to be over-precise, given some evidence characterized by θ_{om} , although this prediction is more complex to test and hence will not be considered in our posterior data analysis.

4.2. Data analysis

For starters, we consider the Bayesian model. Proposition I above states that, in any round, subjects should report estimations of θ that track the (rounded) frequency *f* of female balls in the sample observed by them so far (leaving aside extreme samples in the last estimation round).

Hypothesis I: In average, point estimations do not significantly differ from the average *f*.

Evidence: In average, the subjects' urns have around 56.7 female balls, i.e., the mean θ equals 0.567. In the samples corresponding to the first 15, 22, and 30 extractions, furthermore, the mean (non-rounded) *f* is 0.584, 0.577 and 0.578, respectively. Since the subjects' actual estimates of θ have averages equal to 0.517, 0.522 and 0.530, respectively, we observe a systematic (although small) underestimation of the number of female balls in all rounds. In this respect, the differences between the mean Bayesian and subjects' estimates are significant in the first two rounds (paired t-test, p-value = 0.018 and 0.021) but not so much in the final round (paired t-test, p-value = 0.068).

Observed frequency	Estimation round		
	1	2	3
[0, 0.2)	0,010	0,060	0,030
[0.2, 0.4)	0,060	-0,005	-0,025
[0.4, 0.6)	0,000	-0,010	0,000
[0.6, 0.8)	0,005	-0,020	0,000
[0.8, 1)	-0,140	-0,110	-0,020
Aggregated	-0,005	-0,020	-0,010

Table 1: Median deviation from the Bayesian estimation, conditional on the observed frequency

For a more disaggregate analysis, we define a subject's deviation in a round as the difference between her actual estimate of θ and the predicted Bayesian estimate. In Table 1, each column corresponds to one of the three estimation rounds, and there is one row for each of the intervals [0, 0.2); [0.2, 0.4), etc. Each cell indicates the subjects' median deviation, conditional on the estimation round and the value of f observed so far. Intuitively, the table compares the subjects' biases when most or the majority of news are bad, i.e., f low, and when most news are good. The Bayesian model predicts a nil deviation in each cell. In this respect, we see that the median deviation is practically zero in several cells of the table. Interestingly, overestimation seems to be more systematic when f is low, whereas underestimation tends to occur when f is large. We will return to this point later.[19]

When we consider average, not median, deviations, the differences with the Bayesian prediction are a bit more pronounced –see Table A in Appendix II. Some inflation is observed in this case, particularly when f is low. However, it is far from

[19] Does this evidence signal a bias "towards 50%", maybe because people have (inaccurate) priors assigning non-uniform probability to θ = 0.5? We note in this respect that our control questionnaire explicitly asked whether priors were uniform (see Appendix I) and that Figure 1 is hardly consistent with such hypothesis, although we cannot exclude this possibility for a few subjects.

systematic. Indeed, recall, the overall mean deviation is negative in any round. Further, in those contingencies where there is significant inflation or alternatively deflation, it is largely run by the presence of some outliers. This is illustrated by Figure 1, where each dot corresponds to a subject, placed according to her actual estimate in the third round, and her predicted, Bayesian estimate. As we see, the majority of the points are close to the diagonal. That is, most subjects' estimates track the empirical frequency, except for a few cases. [20]

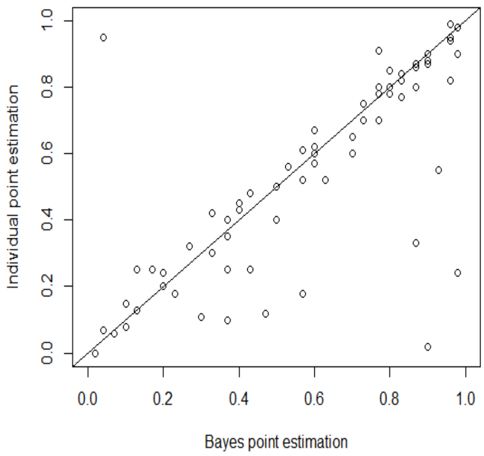


Figure 1: Subjects' estimates in the last round vs. Bayesian prediction

Result I: The average and median subjects slightly underestimate θ in any round. Conditional on the actual frequency of female balls observed by the individual, *f*, inflation is observed at a systematic level only when *f* is low and is rather small in size.

[20] As the figure shows, the underestimation observed in the third round is not *exclusively* due to the 'distortions' described in Footnote 8. Out of the 68 subjects, 10 of them faced in the last round a sample with 0, 1, 29, or 30 female balls, and just 7 of them observed either 29 or 30 balls, which are the only cases where some underestimation is predicted by the Bayesian model, although never higher than 2 balls. Yet the mean deviation among these subjects was -0.1495, i.e., around 15 balls, while the median one was -0.0267. Note though that one of these 7 subjects deviated in 76 balls from the Bayesian estimation; the mean deviation among the remaining 6 subjects is -0.0478.

Consider now the models by AD and BP described in Section 4.1. As we proved there, these models imply that subjects should inflate the estimate of θ, that is, the average/median estimate should be significantly higher than the average/median observed *f* in any estimation round. Further, inflation should not be conditional on *f*. While the evidence described in Result I is not very encouraging, Figure 1 above also shows that some people overestimate. This warrants further tests of those models. In what follows, therefore, we consider additional hypotheses based on Propositions II to VI. The next one follows from Propositions II and IV. Intuitively, the level of inflation depends on whether the estimate of θ is incentivized. Without incentives for accuracy, there is no risk of a material loss for having a mistaken belief about θ. Hence, utility is maximized if the belief is as optimistic as possible, which means that Abel should report $\hat{\theta} = 1$ in the first two, non-incentivized estimations. When there is some risk, in contrast, Abel should infer in a more Bayesian manner, so as to reduce the chance of a mistake (this is particularly true if Abel is very risk averse or cares little about anticipatory utility, i.e., has a small a). In the third estimation, that is, the average $\hat{\theta}$ should be strictly lower than 1 but higher than the average *f*, assuming that subjects display enough heterogeneity in a and risk aversion.

Hypothesis II: The prevalence and extent of inflation is higher in the first two rounds. In the third round, an optimistic subject's accuracy depends on her degree of risk aversion.

Evidence: Several findings speak against the first part of the hypothesis. To start, the share of subjects who give inflated estimates in the first, second, and third rounds equals 36.76, 35.29 and 33.82, respectively. In theory, there should be less people inflating in the third round, but the effect seems negligible (McNemar's test, p-value = 0.6171 and 0.7815 for the comparison of the first and second round with the third one, respectively). Second, people do not become more accurate as a result of the introduction of incentives in the third round. To check this, we consider the *absolute* value of a subject's deviation in a round, which measures the extent of her error. At first sight, we observe increased accuracy, as the mean of the absolute deviation after 15, 22 and 30 extractions is 0.1378, 0.1159 and 0.1060, respectively. An important question, however, is whether this increased accuracy is the result of a learning process or caused by the incentives introduced in the last estimation, as Hypothesis II contends. Thus, we estimate a linear panel data model where the Y variable is a subject's deviation from f (in absolute terms) and the explanatory variables are (i) the round number (1, 2 or 3), in order to measure learning effects; and (ii) a dummy for the third stage, to capture any additional effect due to the incentives provided. Note that learning effects should improve accuracy in the second and third rounds relative to the previous one. In this respect, we find that the coefficient of variable (i) is negative and significant (-0.0219, p-value = 0.023), although quantitatively modest, while that of variable (ii) is not significantly different from zero (p-value = 0.488). There is hence some limited learning effect, while the incentives introduced in the last estimation do not increase accuracy. We have also explored whether incentives affect the direction, if not the extent, of the bias. For this, we run a simple linear regression where the dependent variable is a subject's deviation from f, i.e., not in absolute value, and the X-variable is a dummy taking value 1 when the estimation is made in the third, incentivized round. The coefficient of this variable is positive but non-significant (p-value = 0.698; results are similar with a panel data model). Observe that the positive sign, although non-significant, means that the deviations tend to move towards the positive side in the last round, something unpredicted by the models (people should inflate less frequently then).

We move now to the second part of Hypothesis II, together with a more thorough study of heterogeneity. As we have said, the overall evidence points out that most subjects do not exhibit a substantial bias. Yet averages can be misleading if some people inflate and others deflate. To check for heterogeneity, we compute the share of subjects who under-estimate θ never, once, twice or thrice across all rounds. The respective figures are 26.47, 20.59, 17.65 and 35.29. It can be worth to analyze what characterizes the subjects in this former group, i.e., the most systematically 'optimistic' ones, who never report an estimate lower than the Bayesian one (see Proposition IV above).[21] Before answering this question, however, it must be noted that these subjects do not exhibit very large deviations from *f*; the median deviation is 0.05 and the average one is 0.10.[22] For the sake of comparison, the median *absolute* deviation is 0.03, 0.08, and 0.13 among the subjects who under-estimate one, twice, and thrice, respectively.

A first thing we observe among the optimistic subjects is that they do not act more Bayesian in the third round, i.e., more accurate. This is indicated by the linear panel data model explored above: if we add a dummy for these subjects, they are not significantly more accurate in the third round (p-value = 0.482). A second thing is that there is a clear correlation between the observed frequency *f* and the degree of optimism, i.e., the number of rounds in which a subject does not under-estimate. Figure 2 below provides a comprehensive picture. Each box corresponds to a different group of subjects, i.e., those who underestimated in 0, 1, 2, or 3 rounds, and gives information about the mean *f* observed by those subjects across rounds. Specifically, the length of each box represents the inter-quartile range (IQR) in the corresponding distribution, whereas the vertical lines extend above (below) so as to include all data points within 1.5 IQR of the upper (lower) quartile, stopping at the largest such value. The horizontal line within each box, in turn,

[21] Our conclusions below are similar if we instead define an optimistic subject as one who gives an estimate strictly larger than the Bayesian one in all rounds.

[22] This average deviation equals 0.06 if we remove one subject from the group whose average deviation was around 0.93.

indicates the median value of the mean f observed in the associated distribution. As we can see, half of the participants that never underestimate θ observed a mean frequency lower than 0.345 while half of the participants who were consistently pessimistic across all rounds observed a mean f greater than 0.785.

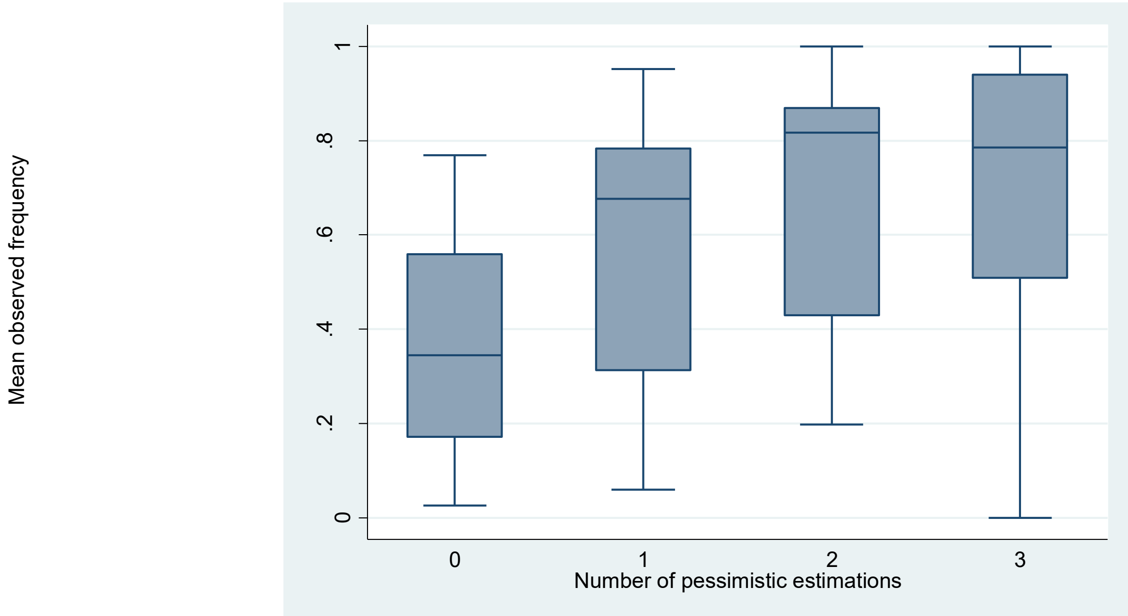


Figure 2: Mean observed frequency conditional on number of underestimations of θ

Therefore the 'optimistic' subjects tend to observe samples with relatively few female, i.e., good signals. Are they different from the other subjects in other respects? To study this issue, we run two regressions. To start, a logit regression finds no significant correlation between a binary variable taking value 1 when the subject strictly inflates in all rounds, and (i) any of our socio-demographic variables, (ii) the subject's degree of risk aversion, (iii) the number of correctly recalled names (net of errors), i.e., with the subject's memory capacity, and (iv) her knowledge of Statistics.[23] The only exception is

[23] As a measure of their statistical knowledge, participants first answered the following question: "In an electoral survey with a sample of 10 voters randomly chosen, 40 percent of them stated they were voting for Party A. From this data and assuming that there are 1000 voters in the country, how many of them do you think will vote for Party A? Provide your best estimate, which must be a number between 0 and 1000." In addition, they were also asked how many ECTS on Statistics and related subjects (Econometrics, Psychometrics, etc) they had passed in the last five years.

the CRT score: More reflexive people are significantly ($p = 0.041$) *more* likely to overestimate in all rounds.[24] Similar results are obtained in a panel data model where the dependent variable is the subject's deviation in each round: The only significant X-variables are the observed frequency (p-value < 0.001) and the CRT score (p-value $= 0.008$), which predict a negative and a positive effect, respectively. We stress that our index of risk aversion has no significant predictive power in the econometric models that we have specified, including one focused on the third round and the optimistic subjects: risk aversion does not correlate with a lower deviation, i.e., 'more' Bayesian estimates.

To further explore the relationship between the CRT score and inflation, Figure 3 below represents the average deviation (grey bars) and the average absolute deviation (white bars), conditional on the subject's CRT score. Two things are worth mentioning here. First, the size of the errors, i.e., the absolute values, tends to be higher for those subjects with low CRT scores, although the effect is not entirely systematic (subjects with a score of 4 have relatively large errors). Second, the CRT score is apparently related with the sign of the errors, as reflexive subjects tend to inflate; note yet that the degree of inflation is in average very small: these subjects tend to be optimistic, but very little. The following result summarizes our key findings so far.

Result II: The size and direction of the deviations does not depend on the round, and hence on the risk of a loss. The share of subjects who overestimate in all rounds is relatively small; moreover, these subjects deviate little from the Bayesian benchmark and do not appear different from other subjects, except in their CRT score and the sample observed (relatively few positive signals). Risk aversion does not predict more

[24] We have also included the amount of time that each subject takes to complete each estimation round, i.e., from the moment that the corresponding screen appears until the subject enters the estimate and proceeds to the next screen. Our hypothesis here is that optimistic subjects might respond relatively fast, without much thought, in these rounds. Yet none of these three variables, i.e., one per round, is significant.

Bayesianism among the optimistic subjects in the third round. Overall, the evidence seems hardly consistent with models of optimism like AD and BP.

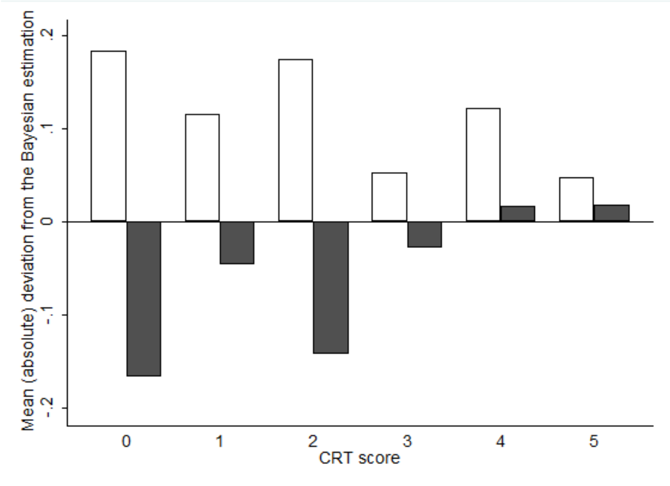


Figure 3: Mean deviation (gray) and mean absolute deviation (white) from the Bayesian estimation, conditional on CRT score

The following hypothesis explores the subjects' degree of doubt in their inferences, as measured by the 95 percent confidence interval elicited after the third estimation round. If subjects can choose their beliefs, it seems at first sight natural that they should express little doubt, i.e., very narrow confidence intervals, particularly since the interval estimations are never incentivized and hence entail no risk –see Möbius et al. (2014) for an application of these ideas to financial markets. Specifically, optimists might assign all the probability mass to one single rate. As we have shown in Section 4.1, however, this kind of extreme over-precision must be present when $\theta_{om} \geq \theta_A$, but not necessarily when $\theta_{om} < \theta_A$, in which case posteriors might be more spread, conditional on risk aversion. In addition, Proposition IV says that Abel will never assign positive probability to any rate below the subjective mode, i.e., the point estimation.

Hypothesis III: For an optimistic subject, confidence intervals are asymmetric, assigning in particular nil probability to any rate below $\hat{\theta}$. Further, they are larger when θ_{om} is relatively small, e.g., when $\theta_{om} \leq ½$.

Evidence: Contrary to the hypothesis, the confidence intervals were generally symmetric around the last point estimation. Specifically, there are no significant differences between the mean last estimate of θ and the mean center of the confidence intervals (paired t-test, p = 0.1014). Figure 4 further illustrates this point. Note also that most subjects lie below the diagonal. Hence subjects report intervals whose midpoints tend to be slightly lower than the estimate of θ, contrary to the predictions by the optimism models.

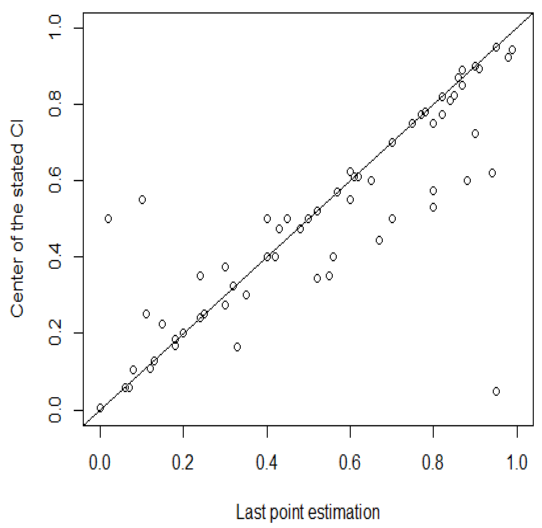


Figure 4: Subject's last estimate of θ vs. midpoint of the stated confidence interval

Note yet that Hypothesis III explicitly refers to the optimistic subjects. Hence, a relevant question is whether the people who never under-estimate θ report also asymmetric intervals. We check this first with a simple linear regression where the dependent variable is the difference between the midpoint of the interval reported by the subject and her last estimate (this difference is called D below), while the X-variable is a binary one taking value 1 when the subject never gives a deflated estimate. The coefficient happens to be negative, although non-significant (p-value = 0.287). Since the estimated constant is negative (and non-significant) as well, midpoints are even lower among the optimistic types, which hardly fits with Hypothesis III.[25]

For further illustration, Figure 5 represents subjects according to their deviation from the Bayesian prediction in the *last* round (in the X-axis) and variable D. The box is divided in quarters, and the optimistic subjects (in that round at least) are the dots in the right-hand quarters. Those in the lower right-hand quarter, further, indicate an interval such that D is negative, contrary to what Hypothesis III predicts. We can see that many optimistic subjects are placed in such quarter and, in any case, the value of D is rarely large for any of those subjects.

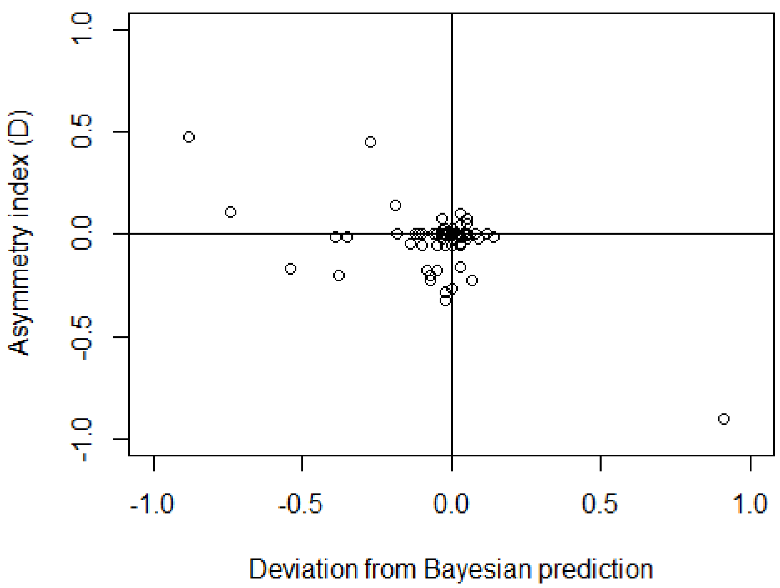


Figure 5: Subject's deviation from Bayesian prediction vs. asymmetry index D

[25] If we control in the regression for the effect of the observed frequency *f*, closely correlated with optimism in our experiment, the coefficient of the dummy becomes marginally significant (p-value = 0.057), but it is still negative.

With respect to the second half of Hypothesis III, it basically says that intervals should be of a relatively larger size when $f \leq 0.5$. Note first that the mean and median size of the elicited interval is 0.223 and 0.150, respectively. More specific to our hypothesis, further, the mean and median size of the elicited interval is 0.179 and 0.140 among participants with $f \leq 0.5$; and 0.253 and 0.180 respectively among participants with $f > 0.5$. Contrary to Hypothesis III, therefore, intervals are larger when $f > 0.5$, although this difference is not significant (t-test, p-value = 0.1576). For further detail, the dark circles in Figure 6 show the observed frequency f and the size of the elicited interval for each individual. Again, we observe that Hypothesis III is not satisfied, as the size of the intervals is larger when f is large, although the difference is not statistically significant, as we have observed. For comparison, Figure 6 also depicts the size of the 95 percent confidence interval of a Bayesian agent for each possible value of f (see the hollow circles). From this figure, it seems quite clear that most participants were overprecise, in the sense that their stated confidence interval were too tight relative to the Bayesian confidence interval. On the other hand, most underprecise individuals observed larger proportions of female names. We stress that similar findings have been also obtained in the study by López-Pérez et al. (2020), where the elicited intervals were incentivized. This suggests that our results here are not an artifact of our experimental design.

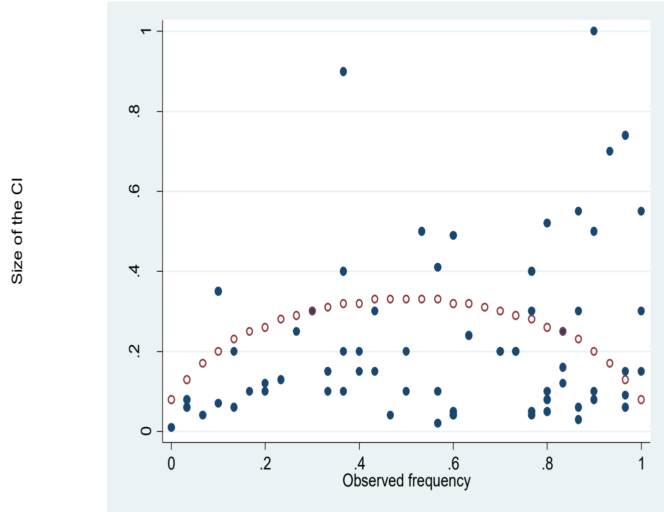


Figure 6: Size of the stated (dark circles) and Bayesian (hollow circles) CIs, by frequency.

Result III: The average subject does not systematically report 'positively skewed' intervals in the last round. This is also true in particular for those subjects who overestimate θ. The size of the intervals does not depend on the evidence observed.

Miscellaneous remarks

We briefly consider here three unrelated issues. The first once concerns model of optimism with cognitive costs. According to the model of choice of beliefs that we have considered, based on AD and BP, the only cost associated to optimistic beliefs is monetary. In the last estimation, inaccurate beliefs reduce the chance of winning the estimation prize, which according to the model should alleviate the participants' estimation bias relative to the previous rounds. Yet, as alternative models suggest –see, e.g. Rabin (1994)–, avoiding evidence that is easily available or repressing unfavorable information are cognitively costly tasks. In our experiment, participants are in fact compelled to observe the evidence. Further, the information provided is easily understandable and the sample size is quite large. Considering these factors altogether, participants may find cognitively hard to ignore the evidence and to get swept up in optimism.

In a nutshell, models of optimism with cognitive costs predict that subjects will overestimate θ in our experiment, but only to a limited extent, similar in all rounds. Although some of our findings go well in line with these predictions, the evidence overall does not seem very supporting. On the negative side, we find that most subjects underestimate at least in one round, i.e., the share of subjects who never underestimate θ is relatively small. Moreover, these optimistic subjects tend to face samples with a reduced share of female balls, and it is not clear why cognitive costs of delusion should be lower in that case. We have also seen that overestimation is somehow correlated with the subject's CRT score. Is this because more reflective subjects tend to be relatively more successful in repressing or avoiding the negative evidence? We find this conjecture a bit puzzling. On the positive side, it is true that the people who inflate θ always often do it to a reduced extent: As we have seen, in fact, the median deviation from the Bayesian estimate among these participants amounts to 5 balls. Further, the introduction of a payoff for accuracy in round 3 has little or no effect on the magnitude of the bias, which again is favorable to these models. In this sense, the models with cognitive costs seem to fare *relatively* better than models like BP.

A second issue that deserves some exploration concerns the intrinsic utility function that we have considered in Section 4.1, which depends just on the agent's monetary gain. It might be also the case, though, that (some) subjects do not like to be disappointed if their beliefs point too high and the reality happens to be mediocre. This is somehow the opposite of anticipatory utility, in that people deflate their expectations, and could be particularly important in our design, where subjects learn the actual state of the world at the end of the experiment.[26] To check this point, we included a test so as to measure a subject's concern with disappointment and regret. Specifically, participants were asked to think of some experience in which their expectations had not been fulfilled. To illustrate the nature of the problem, we included several examples in which they could think, like their performance in a test, the behavior of a beloved person or the result of some bet. Then, they were asked to assess the following statements: (i) In this kind of situations, I tend towards anger or rage; (ii) In this kind of situations I tend towards sadness and/or to mull over the issue for a long time; (iii) I tend to prevent these situations by adjusting downwards my expectations; and (iv) When I live one of these situations, I find it hard to focus or think on other things. Subjects reported their agreement with each statement in a scale from 1 (totally disagree) to 5 (totally agree). The mean (median) answer to questions (i) to (iv) was, respectively, 3 (3), 3.90 (4), 3.09 (3), and 3.41 (4). When included in the panel data model introduced in the discussion of the evidence on Hypothesis II, these measures of disappointment and regret do not predict the observed bias, except for statement (iii), whose associated coefficient turns out to be positive and significant (0.051, p-value = 0.011). This, however, would mean that participants who claim to avoid disappointment by adjusting downwards their expectations are in fact more prone to provide less pessimistic or even optimistic estimations, other things being equal. Although it is unnecessary for our purposes to find an explanation for this paradoxical result, it might be the case that the participants' answers to statement (iii) give information about their *awareness* of their downward bias. If so, it is plausible that those participants who are conscious of their bias try to correct them, therefore providing less pessimistic or even optimistic estimations.

[26] This is not unusual in the literature; see for instance Gotthard –Real (2017) and Barron (2020).

A third issue is that, to get a better insight of the participants' heterogeneity in terms of optimism, we also conducted the Revised Life Orientation Test (LOT-R), a widely used instrument in psychology to assess the level of optimism. The LOT-R comprises ten statements and the respondents must indicate their agreement to each one using a scale from 1 (strongly disagree) to 5 (strongly agree). Three of the statements measure optimism directly, while other three statements measure pessimism. The four remaining statements are fillers and they are not considered in the calculation of the LOT-R score, which is computed as the sum of scores in the statements about optimism and pessimism (for the latter, the scores are previously reversed) and it is comprised between 6 (strongly pessimistic) and 30 (strongly optimistic). The mean and median LOT-R score were 20.71 and 22, respectively.[27] When included in the panel data model introduced in the discussion of the evidence on Hypothesis II, neither the aggregate score or the scores of the different statements were found to have a significant effect on the observed bias.

Result IV: Models of optimism with cognitive costs are inconsistent with: (a) most of the participants underestimate θ at least once and (b) overestimation is more prevalent among individuals who observed relatively few female balls. Measures of the participants' degree of optimism, pessimism and disappointment aversion hardly explain the observed biases.

5. Conclusion

Subjects in our experiment face a rather simple problem of inference, that is, estimating the mode of their posterior beliefs by extrapolation from a sample. Moreover, we have a strong control over the subjects' priors and the signals they observe. While

[27] The data relative to one of the statements of pessimism were corrupted for participants in the first sessions. The results provided correspond to the remaining subsample (n=21).

most subjects track rather closely the Bayesian prediction, Figure 1 shows that a fraction of them deviate. However, they tend to underestimate the mode, not inflate it. Further, underestimation is equally prevalent when there is no prize for accuracy, it is unrelated to personal characteristics like risk aversion, and subjects rarely report skewed confidence intervals. The preference for the state prize to be high, we conclude, hardly motivates deviations from Bayes' rule in our context. Overall, the evidence is not supportive of the models by AD and BP, whereas models with cognitive costs are empirically more relevant, but only in relative terms.

The two families of models state sufficient conditions for a positivity bias, but our findings in this respect are mostly negative. On one hand, models like AD and BP predict that the prevalence or 'demand' of the bias should be maximal when its price is nil or low. In our experiment, however, we find always a very small, arguably negligible bias. Relatedly, these models also say that the prevalence of the bias should decrease as its material cost increases. Our findings do not support this idea, at least within our payoff constellation,. Taking into account additional evidence surveyed in Section 2, the relation between the bias and its 'price', if any, seems hardly linear. While the demand for the bias could be further explored in a design similar to ours, but with a very large estimation prize, e.g., 100 euros if the estimate is accurate, we are however unconvinced that this possibility is worth the cost given the extremely limited 'demand' of optimism in our setting, where the price is low.[28] On the other hand, models with cognitive costs say that the extent of the bias is a function of the proportion of individuals with low costs. The models are not very specific about the determinants of those costs, but the limited evidence for optimism that we find suggests either that (i) our sample was biased towards

[28] Alternatively, one could explore further whether the demand of optimism depends on how desirable the positive beliefs are, e.g., using a state prize of 20 euros per female ball.

agents with high costs or (ii) cognitive costs are not so essential for the occurrence of optimism in our setting. Perhaps these costs are negatively related to the complexity of the inference problem which, arguably, was low in our experiment. This suggests a potential line of investigation.

In this line, our plan for future research is to propose alternative sufficient environmental or personal conditions for optimism and extend our experimental design to test them. For instance, the models with cognitive costs are sometimes not specific about the determinants of these costs. One could guess however that the size of the sample and its informativeness are relevant in this regard. In particular, inflation might be more prevalent and acute when the sample size is small, or the posteriors of several beliefs are similar. For a second line of research, one of our conjectures is that optimism failed to appear in our setting because learning was too ‘transparent’ and hence did not lead to a crowd-out of attention, i.e., to a focus on those aspects of the problem that were more beneficial to the decider; Epley and Gilovich (2016). According to this conjecture, therefore, a sufficient condition for optimism is complexity coupled with incentives to pay attention on the stories or details most beneficial. More research is warranted.

References

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

Aliber, R. Z., & Kindleberger, C. P. (2015). *Manias, Panics, and Crashes*. London: Palgrave Macmillan UK. https://doi.org/10.1007/978-1-137-52574-1

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*. US: American Psychological Association. https://doi.org/10.1037/0096-3445.108.4.441

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., & 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

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, *2*, 69–186. https://doi.org/10.1016/BS.HESBE.2018.11.002

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

Bracha, A., & Brown, D. J. (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, *75*(1), 67–80. https://doi.org/https://doi.org/10.1016/j.geb.2011.11.004

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

Buser, T., Gerhards, L., & van der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, *56*(2), 165–192. https://doi.org/10.1007/s11166-018-9277-3

Caballero, A., & López-Pérez, R. (2020a). Economic Models of Optimism: What Does the Evidence Say? *Cuadernos Económicos de ICE*, *1*(99). https://doi.org/10.32796/cice.2020.99.7008

Caballero, A., & López-Pérez, R. (2020b). *Self-serving recall is not a sufficient cause of optimism: An experiment*.

Case, K. E., & Shiller, R. J. (2003). Is There a Bubble in the Housing Market? *Brookings Papers on Economic Activity*, *2003*(2), 299–362. https://doi.org/10.1353/eca.2004.0004

Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior*, *104*, 1–23. https://doi.org/https://doi.org/10.1016/j.geb.2017.02.015

Coutts, A. (2019a). 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

Coutts, A. (2019b). Testing models of belief bias: An experiment. *Games and Economic Behavior*, *113*, 549–565. https://doi.org/https://doi.org/10.1016/j.geb.2018.11.001

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

Engelmann, J., Lebreton, M., Schwardmann, P., van der Weele, J. J., & Chang, L.-A. (2019). Anticipatory Anxiety and Wishful Thinking. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3408017

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

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

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

Grossman, Z., & Owens, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization*, *84*(2), 510–524. https://doi.org/https://doi.org/10.1016/j.jebo.2012.08.006

Heger, S. A., & Papageorge, N. W. (2018). We should totally open a restaurant: How optimism and overconfidence affect beliefs. *Journal of Economic Psychology*, *67*, 177–190. https://doi.org/https://doi.org/10.1016/j.joep.2018.06.006

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

López-Pérez, R., Rodríguez-Moral, A., & Vorsatz, M. (2020). *Simplified mental representations as a cause of overprecision (manuscript)*.

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

Mayraz, G. (2013). *Wishful thinking in predictions of asset prices*.

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

Rabin, M. (1994). Cognitive dissonance and social change. *Journal of Economic Behavior & Organization*, *23*(2), 177–194. https://doi.org/10.1016/0167-2681(94)90066-3

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

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

Schlag, K. H., Tremewan, J., & van der Weele, J. J. (2015). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics*, *18*(3), 457–490. https://doi.org/10.1007/s10683-014-9416-x

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

Shiller, R. J. (2000). *Irrational Exuberance*. Princeton University Press.

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

Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C., & Sharot, T. (2016). How people update beliefs about climate change: Good news and bad news. *Cornell L. Rev.*, *102*, 1431.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, *34*(1), 1–16. https://doi.org/DOI: 10.1017/S0140525X10001354

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

Wiswall, M., & Zafar, B. (2015). How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital*, *9*(2), 117–169. https://doi.org/10.1086/681542

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

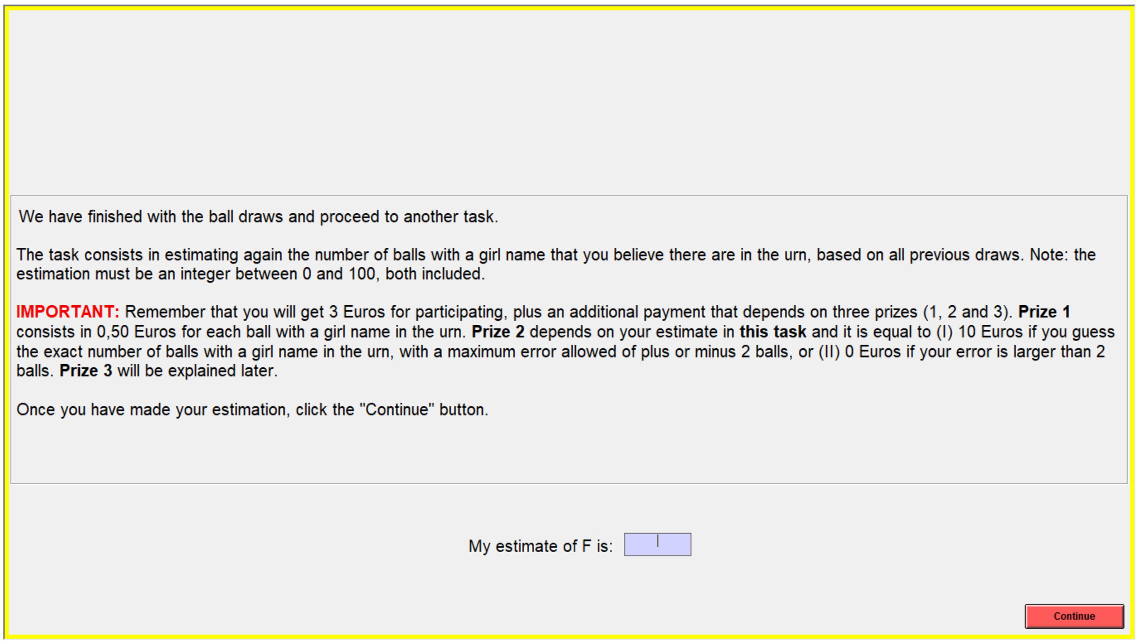
There is a 'virtual urn' with 100 balls. Each ball has written a different name of girl or boy; any of these names is used with a relatively high frequency in Spain. Let us call F the actual number of balls with a female name in your urn. You do not know either F or the number of balls in your urn with 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.

Examples of screenshots



We have finished with the ball draws and proceed to another task.

The task consists in estimating again the number of balls with a girl name that you believe there are in the urn, based on all previous draws. Note: the estimation must be an integer between 0 and 100, both included.

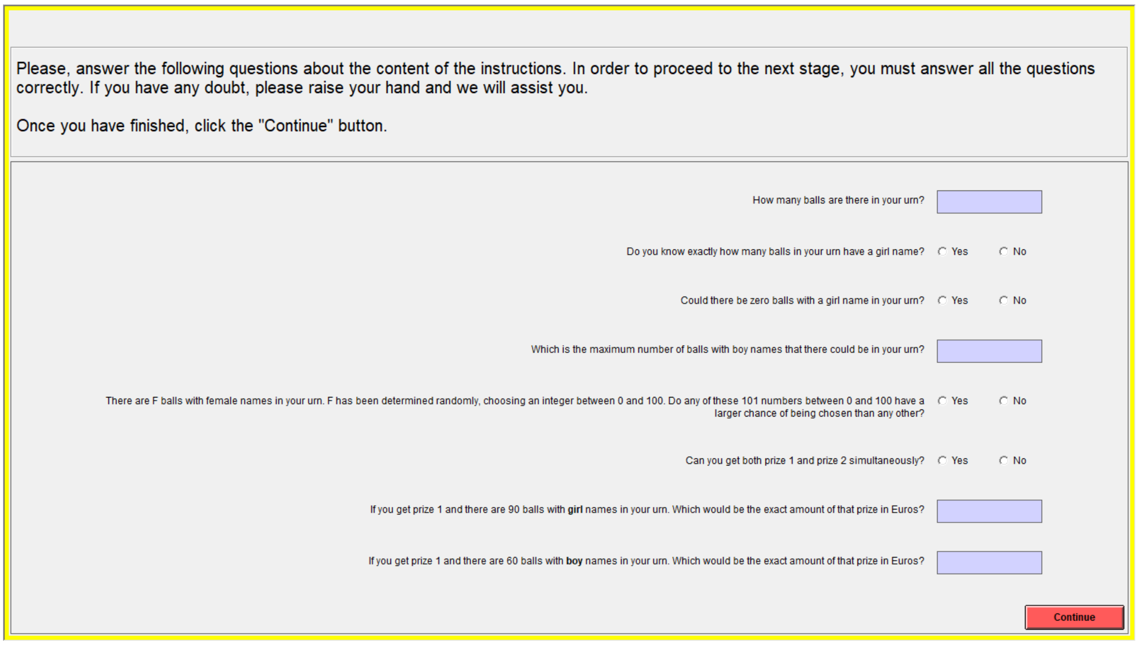
IMPORTANT: Remember that you will get 3 Euros for participating, plus an additional payment that depends on three prizes (1, 2 and 3). **Prize 1** consists in 0,50 Euros for each ball with a girl name in the urn. **Prize 2** depends on your estimate in **this task** and it is equal to (I) 10 Euros if you guess the exact number of balls with a girl name in the urn, with a maximum error allowed of plus or minus 2 balls, or (II) 0 Euros if your error is larger than 2 balls. **Prize 3** will be explained later.

Once you have made your estimation, click the "Continue" button.

My estimate of F is:

Continue

Screenshot for the third estimation task, i.e., the incentivized one



Please, answer the following questions about the content of the instructions. In order to proceed to the next stage, you must answer all the questions correctly. If you have any doubt, please raise your hand and we will assist you.

Once you have finished, click the "Continue" button.

How many balls are there in your urn?

Do you know exactly how many balls in your urn have a girl name? Yes No

Could there be zero balls with a girl name in your urn? Yes No

Which is the maximum number of balls with boy names that there could be in your urn?

There are F balls with female names in your urn. F has been determined randomly, choosing an integer between 0 and 100. Do any of these 101 numbers between 0 and 100 have a larger chance of being chosen than any other? Yes No

Can you get both prize 1 and prize 2 simultaneously? Yes No

If you get prize 1 and there are 90 balls with **girl** names in your urn. Which would be the exact amount of that prize in Euros?

If you get prize 1 and there are 60 balls with **boy** names in your urn. Which would be the exact amount of that prize in Euros?

Continue

Screenshot for the control questions

Note: In the last two questions in this screenshot, the respective numbers of balls with girl and boy name in the urn were determined randomly for each subject; the selected numbers could be any multiple of 10 between 0 and 100, so as to facilitate mental computations.

Appendix II: Additional data

Observed frequency	First estimation	Second estimation	Third estimation
[0, 0.2)	0,154	0,138	0,127
[0.2, 0.4)	0,024	-0,055	-0,042
[0.4, 0.6)	-0,085	-0,061	-0,079
[0.6, 0.8)	-0,015	-0,018	-0,008
[0.8, 1)	-0,174	-0,146	-0,127
Aggregated	-0,066	-0,054	-0,048

Table A: Mean deviation from the Bayesian estimation, conditional on the observed frequency

Appendix III: Proofs

Proposition II. If $a = 0$, then any subjective belief such that θ_{sm} coincides with the objective mode θ_{om} is optimal.

Proof: if $a = 0$, Abel chooses beliefs that maximize $\frac{1}{2}[\sum_{\Theta} \mathrm{u}(\mathrm{M}\theta_{\mathrm{k}}) \cdot p_{\mathrm{k}} + \mathrm{u}(\bar{v})p_{sm}]$. Since p_{sm} is the objective probability of the subjective mode, and it is maximized when the subjective mode coincides with the objective mode, any distribution of subjective probabilities such that $\theta_{\mathrm{sm}} = \theta_{\mathrm{om}}$ is optimal. ■

Proposition III. If $\bar{v} = 0$, then the optimal beliefs are characterized by $\hat{p}_{\mathrm{k}} = 1$ for $\theta_{\mathrm{k}} = 1$ and $\hat{p}_{\mathrm{k}} = 0$ for any $\theta_{\mathrm{k}} < 1$.

Proof: if $\bar{v} = 0$, Abel chooses beliefs that maximize $\frac{1}{2}\sum_{\Theta} \mathrm{u}(\mathrm{M}\theta_{\mathrm{k}})(a\hat{p}_{\mathrm{k}} + p_{\mathrm{k}})$. It is straightforward that this expression is maximized for beliefs such that $\hat{p}_{\mathrm{k}} = 1$ for $\theta_{\mathrm{k}} = 1$ and $\hat{p}_{\mathrm{k}} = 0$ for any $\theta_{\mathrm{k}} < 1$. ■

Proposition IV. The optimal subjective mode θ^*_{sm} is at least equal to the objective mode. If $a > 0$, further, optimal beliefs are characterized by $\hat{p}_{\mathrm{k}} = 0$ for any $\theta_{\mathrm{k}} < \theta^*_{\mathrm{sm}}$.

Proof: consider subjective beliefs (A) such that $\theta_{\mathrm{sm}} = \theta_{\mathrm{k}} < \theta_{\mathrm{om}}$, $\hat{p}_k = x$ and $\hat{p}_{om} = y$ ($x \geq y$). Now consider beliefs (B), identical to (A) except by the fact that $\hat{p}_k = y$ and $\hat{p}_{om} = x$ so that $\theta_{\mathrm{sm}} = \theta_{\mathrm{om}}$ (if $x = y$ and there are hence multiple modes, condition $\theta_{\mathrm{sm}} = \theta_{\mathrm{om}}$ means that the estimate chosen in period 2 is θ_{om} when beliefs are (B)). The expected utility with beliefs (B) is larger than with (A) if

$$a\frac{1}{2}\left[(x-y)\big(u(\mathrm{M}\theta_{\mathrm{om}}) - u(\mathrm{M}\theta_{\mathrm{k}})\big)\right] + \frac{1}{2}u(\bar{v})(p_{om} - p_k) > 0$$

which holds for any $\theta_{\mathrm{k}} < \theta_{\mathrm{om}}$, since $p_{om} > p_k$ and $x \geq y$. Therefore, the optimal subjective mode cannot be lower than θ_{om}. For the second part of the proposition, consider beliefs such that $\hat{p}_k > 0$ for some state $\theta_{\mathrm{k}} < \theta_{\mathrm{sm}}$. If Abel transfers all the subjective probability from state θ_{k} to θ_{sm}, his expected utility increases if

$$a\frac{1}{2}\left[\hat{p}_k\big(u(\mathrm{M}\theta_{\mathrm{sm}}) + u(\bar{v}) - u(\mathrm{M}\theta_{\mathrm{k}})\big)\right] > 0$$

which holds necessarily since $\theta_{\mathrm{sm}} > \theta_{\mathrm{k}}$ and $a > 0$. Therefore, any beliefs such that $\hat{p}_k > 0$ for any state $\theta_{\mathrm{k}} < \theta_{\mathrm{sm}}$ are suboptimal. ■

Proposition V. Consider beliefs with subjective mode θ_{sm}. If $u(\bar{v}) \geq u(\mathrm{M}) - u\big(\mathrm{M}\theta_{\mathrm{sm}}\big)$, a necessary (but not sufficient) condition for these beliefs to be optimal is that they assign $\hat{p}_{\mathrm{k}} > 0$ to any $\theta_{\mathrm{k}} \neq \theta_{\mathrm{sm}}$.

Proof: the proof of Proposition IV above shows that it is not optimal to assign strictly positive probability to any rate lower than θ_{sm}. Assume now, without loss of generality, that the posteriors assign strictly positive probability to some states larger than θ_{sm} (if any) and in particular to the largest state, i.e., $\theta_{\mathrm{k}} = 1$ in our experiment. Consider also posteriors (B), identical to (A) except that some probability mass $\varepsilon \geq 0$ is transferred from $\theta_k = 1$ to θ_{sm}. Taking (3) into account, the expected utility of (B) will be higher than that of (A) if

$$\varepsilon \cdot a \cdot u(\mathrm{M}\theta_{sm}) - \varepsilon \cdot a \cdot u(\mathrm{M}) + \varepsilon \cdot a \cdot u(\bar{v}) \geq 0 \leftrightarrow u(\bar{v}) \geq u(\mathrm{M}) - u(\mathrm{M}\theta_{sm}) \qquad (4)$$

While condition (4) is obtained given a transfer from $\theta_k = 1$ to θ_{sm}, it also ensures that a transfer of probability from any state between 1 and θ_{sm} (if any) to θ_{sm} improves Abel's utility. If (4) holds, therefore, Abel is better if he concentrates the probability mass in θ_{sm}, instead of spreading part of it among some larger states (keeping at the same time the mode in θ_{sm}). ■

Corollary 1: If $u(\bar{v}) \geq u(\mathrm{M}) - u(\mathrm{M}\theta_{om})$, Abel assigns all the probability mass to one single rate. Its specific value depends on the curvature of the utility function of money, $u(x)$, but also on a. In particular, Abel is Bayesian, i.e., $\theta_{sm} = \theta_{om}$, if a is low enough or if the utility of money increases at a sufficiently lower rate than the posterior beliefs (as we move towards the objective mode).

Proof: Proposition IV says that, in the optimum, the subjective mode must be some rate between θ_{om} and 1. When condition (4) is satisfied for $\theta_{sm} = \theta_{om}$, therefore, Abel's optimal beliefs must necessarily concentrate all the mass in some rate. It follows that the optimal beliefs, i.e., the value of θ^*_{sm}, can be determined by comparing the utility of the potential degenerate distributions. For example, it turns out that a probability distribution with the whole mass in θ_{om} is optimal if

$$au(\mathrm{M}\theta_{om}) + p_{om} \cdot u(\bar{v}) \geq au(\mathrm{M}\theta_k) + p_k \cdot u(\bar{v}) \qquad \forall \theta_k > \theta_{om}$$

Rearranging terms, this can be expressed as

$$p_{om} - p_k \geq \frac{a[u(\mathrm{M}\theta_k) - u(\mathrm{M}\theta_{om})]}{u(\bar{v})} \qquad \forall \theta_k > \theta_{om} \qquad (5)$$

which holds true as far as a is low enough and/or the utility of money increases at a sufficiently lower rate than the posterior beliefs (as we move towards the objective mode).

Proposition VI. Consider beliefs with subjective mode θ_{sm} such that $u(\bar{v}) < u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{sm}})$. Conditional on risk aversion, a necessary (but not sufficient) condition for these beliefs to be optimal is that they assign evenly as much probability as possible to some of the largest state(s), subject to $\hat{p}_{\mathrm{sm}} > \hat{p}_{\mathrm{k}}$ for any θ_{k}.

Proof: consider again beliefs (A) and (B), described in the proof of Proposition V. For the sake of the exposition, assume $\theta_{\mathrm{sm}} < 1$, so that the probability mass is not entirely assigned to $\theta_{\mathrm{k}} = 1$. If $u(\bar{v}) < u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{sm}})$, we have shown that Abel is better by transferring some probability from θ_{sm} to $\theta_{\mathrm{k}} = 1$ or in fact from any θ_{k} such that $\theta_{\mathrm{sm}} < \theta_{\mathrm{k}} < 1$, at least as far as θ_{sm} remains the only subjective mode (or the one selected as an estimate in period 2 if there are several modes). A relevant question is therefore if it can be optimal to transfer all probability to $\theta_{\mathrm{k}} = 1$, and the answer is not affirmative in general. To check this point, it suffices to compare the expected utility of two degenerate beliefs: one where all the probability mass is in $\theta_{\mathrm{sm}} < 1$, and another where the mass is concentrated in $\theta_{\mathrm{k}} = 1$, with objective probability p_1. The latter beliefs give higher expected utility if

$$a\frac{1}{2}[u(\mathrm{M}) + u(\bar{v})] + \frac{1}{2}u(\bar{v})p_1 > a\frac{1}{2}[u(\mathrm{M}\theta_{\mathrm{sm}}) + u(\bar{v})] + \frac{1}{2}[u(\bar{v})p_{sm}]$$

that is, if

$$a[u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{sm}})] > u(\bar{v})(p_{sm} - p_1)$$

which is not necessarily true under our conditions, because a can be low and $p_{sm} - p_1$ large relative to $u(\mathrm{M}) - u(\mathrm{M}\theta_{\mathrm{sm}})$, which depends on the curvature of the utility function.

Although we cannot find a closed-form solution for θ^{*}_{sm}, hence, we can at least say that it can be lower than 1 in some circumstances, and the probability mass will be in that case spread. One possibility is that optimal beliefs entail subjective probabilities $\hat{p}_{sm} = 0.5$ and $\hat{p}_1 = 0.5$ (if Abel expects $\theta_{\mathrm{k}} = 1$ to be chosen in period 2, he should assign infinitesimally less probability to $\theta_{\mathrm{k}} = 1$). Alternatively, Abel might find optimal to assign some positive probability ε also to the second largest state $\theta = 0.99$. In fact, this is better than assigning positive probability only to the subjective mode and the largest state if

$$a\frac{1}{2}\left[\left(\frac{1}{2}-\frac{\varepsilon}{2}\right)u\left(M\theta_{\mathrm{sm}}\right)+\left(\frac{1}{2}-\frac{\varepsilon}{2}\right)u(M)+\varepsilon u(0.99M)+\left(\frac{1}{2}-\frac{\varepsilon}{2}\right)u(\bar{v})\right]$$
$$> a\frac{1}{2}\left[\frac{1}{2}u\left(M\theta_{\mathrm{sm}}\right)+\frac{1}{2}u(M)+\frac{1}{2}u(\bar{v})\right]$$

That is, if

$$u(0.99M) > \frac{1}{2}\left[u\left(M\theta_{\mathrm{sm}}\right)+u(M)+u(\bar{v})\right]$$

We are assuming that Abel transfers the same mass of probability from the subjective mode and the largest state to the second largest state. Note that Abel can never be better off by transferring probability from the largest state only. However, given the previous probabilities $\hat{p}_{sm} = 0.5$ and $\hat{p}_1 = 0.5$, the restriction that the subjective mode is unchanged implies that the probability must be transferred in the same amount from θ_{sm} and the largest state. We can generalize the last condition so that Abel will assign positive probability to the n-th largest state θ_{k} if

$$M\theta_k > \frac{M\theta_{sm} + v + M\sum_{k+1}^{100}\theta_i}{n}$$

Chapter 3

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

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".[29] 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.[30]

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

[29] 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".

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

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.[31]

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

[31] 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.

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 θ = 1, i.e., in the '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.[32] 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.

[32] 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.

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 recalls are incentivized.[33] 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%).[34] 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

[33] 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.

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

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).[35] 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.[36] In Zimmermann (2020), subjects solved 10 Raven matrices and were then randomly matched into a group with nine other subjects.

[35] 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.

[36] 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.

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 directions, irrespectively of the feedback. Yet beliefs elicited one month later substantially underweight negative feedback.[37] 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

[37] 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.

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 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 and 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 ∊ {f, m}, and such that probability (S = f) = θ –for expositional purposes, we sometimes refer to f as female, and m as male. Eve does not know the exact value of θ. Let Θ ⊆ [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 ⊆ Θ to a probability p(E). Let p_k denote Eve's priors about rate $\theta_k \in \Theta$. In our experiment, to illustrate, Θ = {0, 0.01,…, 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):

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

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 ∈ {f, m}, one gets different specifications of the model. The Bayesian model of inference assumes of course that all agents are like Eve, with $\pi(I_{tf}) = \pi(I_{tm}) = 1$ ∀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).[38] 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.[39] 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

[38] 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.

[39] 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,

male realizations that Adam recalls accurately, i.e., $\pi(I_{tm}) = 1$ for any such observation. Further, all female realizations are 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" $M(\lambda)$, with $M(\lambda_N) = 0$, and $\frac{dM}{d\lambda} \leq 0$ for $\lambda < \lambda_N$.[40]

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 and data analysis

4.1. Experimental design and procedures

Succinctly, the experiment consisted of an 'estimation task' followed by a 'recall task'. The experimental design is the same used in the previous chapter, which was focused on the 'estimation task'. Therefore, we provide here only a brief summary of the experiment. The reader can find a detailed description of the experiment -particularly in relation to the 'estimation task'- in Chapter 2.

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

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

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.[41] 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.[42] 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 each estimation task only immediately after observing the corresponding extractions and did not receive any feedback about prior extractions. Further, additional tasks and questions are inserted *between* some extractions (see Chapter 2 for details).

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

[41] 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.

[42] 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.

deducted 0.20 euros for each 'incorrect', non-extracted name, so as to prevent subjects from introducing well-known, common names that had not been extracted.[43] 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

[43] 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.

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.[44]

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., $f = \frac{\mathrm{f}}{\mathrm{m+f}}$. 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.

[44] 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.

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.[45] 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 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).

[45] 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.

Hence, the under-estimation observed does not seem an artifact of the estimation prize. Overall, the evidence observed is rather favorable to Hypothesis I. ■

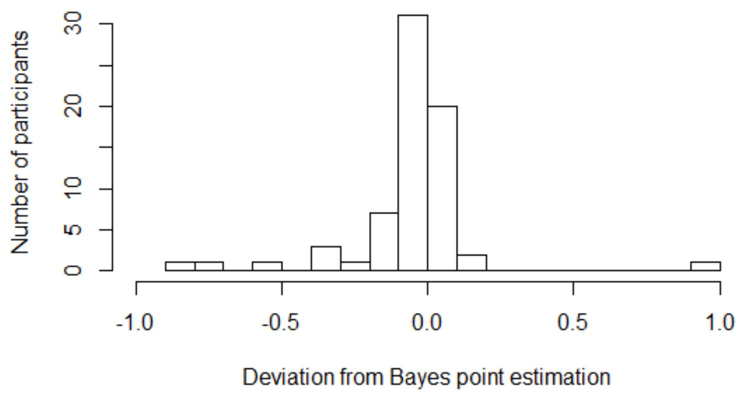


Figure 1: Distribution of subjects' deviations in the last, incentivized estimation round

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 $f > 0.5$.[46] 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.

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-

[46] 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.

observed name was introduced or because a name was written more times than it had been sampled.

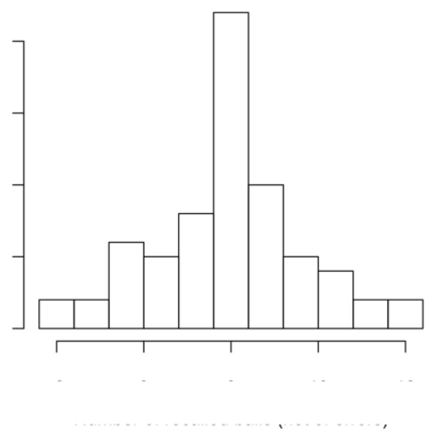


Figure 2: Distribution of subjects' correct insertions in the recall task, net of errors

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[47] 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.

[47] 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.

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 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).[48]

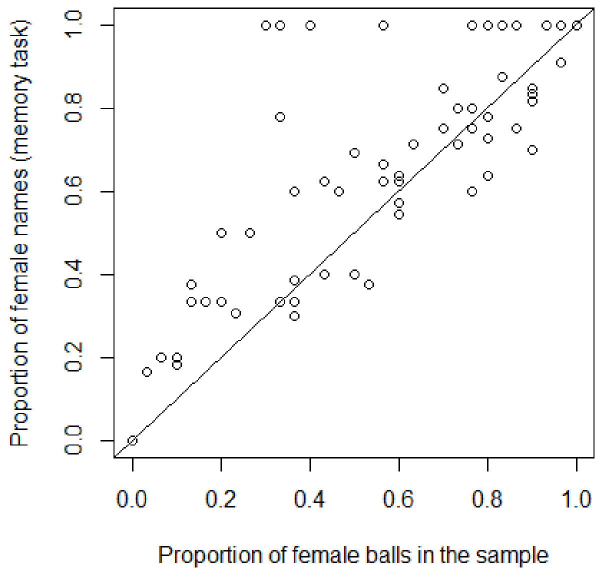


Figure 3: share of recalls that are female vs. objective sample

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

[48] 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.

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.

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.[49] 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,

[49] 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.

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 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).[50]

[50] 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.

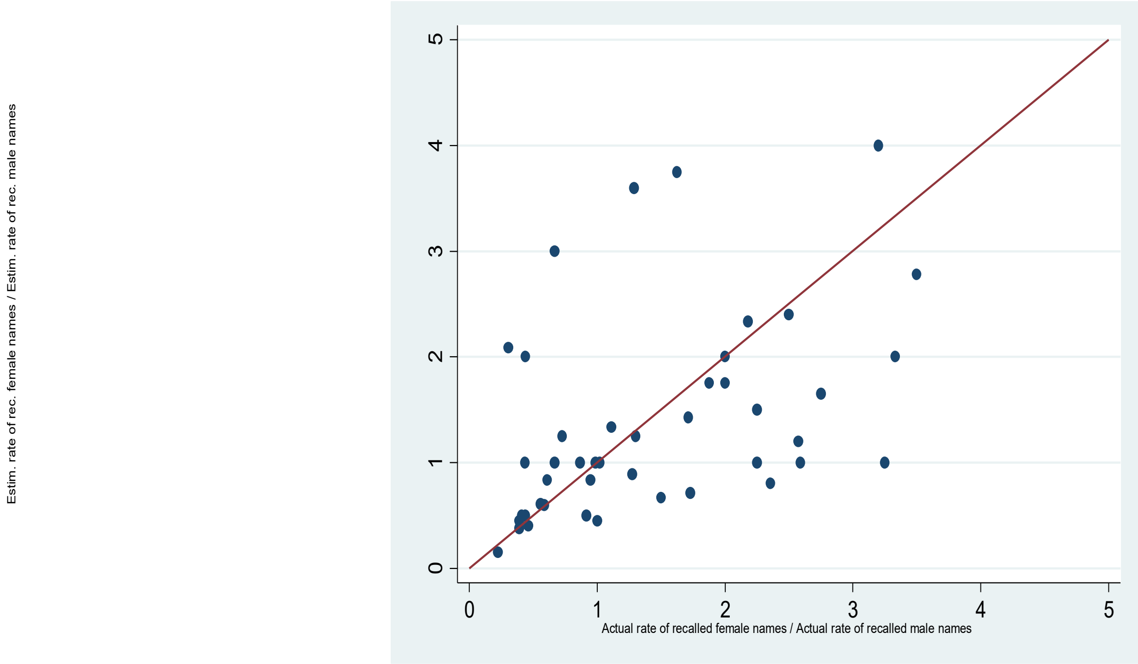


Figure 4: Subjects' actual and perceived recall bias (favoring 'good news')

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 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.[51] 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

[51] 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.

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.[52]

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 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.[53] 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

[52] 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.

[53] 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.

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 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.[54] 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.

[54] 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.

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.

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.

References

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

Chapter 4

Omitting the past

1. Introduction

In a more or less refined form, our beliefs about the likelihood of relevant events appear to play a key role in numerous choice problems, including individual decision problems but also social interactions, e.g., in Bayesian games. Further, beliefs often evolve and change, in a process mediated by the evidence available. That is, belief updating frequently takes the form of sampling from experience. If the observed empirical frequency of an event (e.g. rain in the Atacama Desert) is very low, for instance, people typically respond accordingly and put low probability on that event. In fact, people seem to apply experience-based-inference in many real-life decisions: Jury verdicts (Davis, 1984; Pennington and Hastie, 1986), physicians' recommendations of vaccination or surgery (Hertwig et al., 2004; Barron and Ursino, 2013), the daily decision to use safety devices (Erev, 2007), the share of copyrighted material without authorization (Barron, Leider and Stack, 2008), submissions of papers to top journals (Hilbig and Glöckner, 2011), evaluation of innovations (Rakow and Miler, 2009), the purchase of insurance for disaster (floods, accident, etc.), or the response to terrorist attacks (Yechiam et al., 2005).

In the standard Bayesian model of inference, e.g., Samuelson (2004), agents can assign prior probabilities to each element of the state-space. If they repeatedly observe some random variable or signal S, further, they use *all* this evidence to update their priors by means of Bayes' rule. While this learning model has appeal from a normative point of view, its empirical validity has been contested (e.g., Tversky and Kahneman, 1973, 1974; Hogarth and Einhorn, 1992; McFadden, 1999; Hertwig et al, 2004; Barron and Ursino, 2013). Objections take many shapes, but the one that interests us here is that, when people sample from experience, they are likely to omit part of that evidence, particularly if they lack records and receive no feedback. This idea is consistent with some non-Bayesian models of inference like Mullainathan (2002), Gabaix (2014) and Schwartzstein (2014). Motivated by this literature, we use here lab experiments and a theoretical framework to explore inference about event frequencies when people omit or fail to consider some observations of S, and more specifically research questions **I** to **III**, presented below.

Before we present those questions, however, consider for expositional purposes our experiment, which has two phases. In both, each subject observes one by one an undetermined number (in fact, 50) of random draws with replacement from one urn containing 100 balls, blue or red. Although the proportion or rate θ of red balls in her/his urn is uncertain, the subject knows that θ is randomly selected at the start of each phase with uniform probability from the same set of three rates –we consider two treatments (T1 and T2), with respective sets (0.3, 0.5, 0.7) and (0.4, 0.5, 0.6). After subjects observe the 50 drawings in a phase, they are asked their point estimation of θ in an incentive-compatible manner. The only difference between phases, leaving aside that the urn is renewed, is that subjects know in advance in Phase 2 that they will have to estimate θ and will earn a prize if they are accurate.

We think that our experiment offers insights on data omission, for at least four reasons. To start, the estimation of θ clearly requires extrapolation. *Second*, subjects (i) cannot keep records, (ii) are never given feedback about previous extractions, and (iii) do not know in advance the total number of draws (i.e., 50). On top of that, (iv) extractions are paused during 90 seconds between the 42^{nd} and the 43^{rd} draws in each phase and (v) the estimation task is unexpected in Phase 1. It follows that attention and memory failures can lead to data omission, particularly in Phase 1. *Third*, the estimation of θ is computationally undemanding, which minimizes confounds. Indeed, its optimal resolution just involves *counting* the number of blue and red extractions in the phase. If the frequency of red balls in the sample observed equals $f \in [0, 1]$, in effect, the Bayesian estimate of θ is simply the proportion closest to f–thus, if the feasible rates are 0.3, 0.5, and 0.7 and $f = 0.47$, the Bayesian estimate is 0.5. In plain terms: If a relatively large number of extractions are red (blue), a Bayesian chooses the highest (lowest) rate, while she chooses the intermediate rate when there is a similar number of blue and red extractions. *Fourth*, the subjects' estimates of θ allow us to make inferences on what data points they omit (assuming that they estimate by extrapolation).

To clarify our methodology, we embed it into a tractable and general theoretical framework. In this stylized behavioral model, an agent called Adam has to estimate the frequency or rate of occurrence θ of some outcome x of a binary i.i.d. signal S, based on a number of realizations of S, each one possibly observed at a different point in time. Perhaps due to limited attention and memory, Adam considers a subsample of the realizations, but otherwise uses Bayes' rule. Compared with a Bayesian agent, therefore, his estimate of θ tracks the empirical frequency of x in the subsample, often resulting in under or overestimation. Our framework allows for heterogeneity, in that different Adams may focus on different subsamples (or even the whole sample, which is the Bayesian case). When applied to our experiment, the model says that Adam's estimate of the rate of red balls in his urn will track the frequency of red balls in the sub-sample considered by him. From the subjects' estimates, hence, we can make inferences about data omission and explore questions **I** to **III** below.

I (Heterogeneity): Are people heterogeneous in their data omission patterns? Guided by our theoretical framework, we employ a novel fit analysis to infer what data points people omit. For simplicity, we assume that there are *at maximum* two types of Adams, and that each one either considers some of the first observations or some of the last observations. In one type pair, for instance, the first type considers just the first five observations, while the second one considers the whole sample (as a Bayesian). Briefly speaking, we posit that people exhibit either primacy *or* recency effects; both patterns have been extensively documented –e.g., Ebbinghaus, 1911; Murdock, 1962; Hogarth and Einhorn, 1992.[55] We then compare thousands of type pairs to compute the pair that best replicates our aggregate data, additionally classifying our subjects into those two types. We find several things. First, the assumption that there is just one single type performs relatively badly. Second, not every heterogeneous pair performs equally well. Specifically, pairs where one type considers many observations while the other one considers relatively few observations work much better; in fact, these pairs often are among the top 5%. In the optimal pair, in fact, one type focuses on very few *initial* observations, whereas the other type considers a large streak of the last observations (or maybe all of them, as a Bayesian). It seems therefore that people may experience primacy and recency effects to different degrees, which to our knowledge is a novel contribution. Although there are possibly other explanations and a full exploration of this issue is out of the scope of this paper, the phenomenon might be due to subjects' heterogeneity in attentiveness or reflectiveness. In our experiment, that is, (relatively) inattentive types might pay attention mostly to the first signal observations (perhaps losing motivation afterwards), which as a result leave a stronger memory trace. These types hence exhibit primacy effects. In contrast, attentive types pay in average more attention to all observations.[56] In this vein, a regression analysis shows that a subject's CRT score (Frederick, 2005) correlates with his type in the optimal pair derived from the

[55] We have also conducted a computationally more demanding fit analysis that allows for Adam to display both primacy and recency effects, but it gives few additional insights.

[56] This can include subjects who counted the number of extracted red balls from the outset, even when they were given no explicit clues that they should do so. Note that people are unlikely to keep such detailed mental records for every event out of the lab. Hence, we believe that our data is a conservative estimate of the extent of data omission in the field.

classification analysis. A subject's performance in a memory task, though, is unrelated with his type. It *seems* therefore that data omission in our design is more related to inattention than memory.

II (Significance): How relevant is data omission? While all of us probably agree that humans have limited attention and memory, skeptic readers may object that in practice data omission is either minor, non-systematic, or relevant only in complex inference problems characterized by multiple signals with numerous outcomes each, large state spaces, etc. In this respect, the evidence for data omission in our stylized experimental setting strongly suggests that the phenomenon is often relevant, and not only in complex inference problems.[57] Indeed, the fit analysis shows that the (homogeneous) Bayesian model explains around 62% of the subjects' estimations, whereas the optimal pair performs statistically better ($p < 0.001$), explaining around 74.81 % of the choices.[58] We interpret this significant difference, not warranted a priori at all, as a signal of the relevance of data omission, even in highly stylized settings.

III (Ambiguity & Complexity): What environmental factors influence accuracy when people can omit data? We know from studies like Eysenck and Eysenck (1980) and Caplin and Dean (2015) that performance on information tasks (including counting tasks) responds to incentives. In many occasions, however, incentives are simply ambiguous. Does accuracy improve when agents have some experience and learn the true payoffs

[57] Regardless, data omission is likely to be more prevalent and severe in more complex and natural scenarios, where there are multiple signals (i.e., urns), each one with its own rate. In comparison with these settings, again, our experiment possibly under-estimates the general effect of data omission on inference.

[58] Two remarks are due. First, our primary goals are questions **I** to **III** above, not to test the Bayesian model (a point which has received a lot of attention in the literature). A comparison with the Bayesian model seems however necessary to ascertain the relevance of data omission. In this regard, to be clear, we do not claim that the Bayesian standard does badly in our treatments, but that incorporating data omission increases accuracy even in a relatively 'trivial' problem like the one we consider. This is a signal of the relevance of the ensuing biases. Second, a regression analysis indicates that our theory is more accurate when the frequency observed by the subject is not close to the midpoint between two rates, i.e., when the inference problem is cognitively less complex. Hence, other factors aside from data omission, e.g., computation errors, seem to generate mistakes in our experiment.

from learning? We check this point with a comparison of the two phases of our experiment. That is, since incentives are clear from the outset in the second phase of each treatment, but not in the first one, we surmise that subjects should be less likely to commit large errors in the second phase. We also explore whether this potential effect is mediated by the complexity of the problem, which explain why we consider two treatments, including the more 'difficult' T2 (in a sense, we use a 2x2 design which changes the difficulty of the problem across treatments and the ambiguity of incentives within treatments across phases). We find that the probability of a Bayesian prediction is significantly higher in the second phase in T1 but not in T2. Hence, complexity mediates the role of ambiguity on accuracy. Our regression analysis also shows that the overall effect is mediated by a change in the share of types: When incentives are clear, there are significantly less subjects who estimate as if they focused on a small subsample.

The rest of the paper is organized as follows. The next section reviews in more detail some related literature, particularly on quasi-Bayesian inference, to contextualize our work. Section 3 describes our experimental design. In Section 4 we first present the Bayesian model, a natural benchmark, and then our framework to analyze data omission. Afterwards, we explore questions **I** to **III** above in light of our experimental data. The paper concludes by mentioning future venues of research.

2. Related literature

Our paper is closely related to the heuristics and biases literature. In particular, Tversky and Kahneman (1973) suggest that people often resort to the *availability heuristic*, evaluating the frequency of an event or category based on the ease and fluency with which instances of that event come to mind. They stress that the use of this mental shortcut leads to systematic biases, because there are other factors apart from frequency that make it easy to come up with instances, like their recency, the attention one pays to them, or how familiar they are. In a famous experiment, participants listened to a list of names containing 19 famous women (or men) and 20 less famous men (or women). Some participants were subsequently asked to recall as many names as possible; the names of the famous celebrities were recalled relatively more frequently. Others had to estimate whether male or female names were more frequent on the list; most of them incorrectly judged the gender associated with more famous names to be more frequent. This evidence markedly goes in line with the idea that people can omit data in their inferences. In this vein, our theoretical framework incorporates insights of the availability heuristic, although with a quasi-Bayesian structure.

There is a growing theoretical literature on quasi-Bayesian inference, to which our paper is related as well –e.g., Barberis et al., 1998; Rabin and Schrag, 1999; Rabin, 2002; Mullainathan, 2002; Gennaioli and Shleifer, 2010; Gabaix, 2014; Schwartzstein, 2014. One strand of this literature assumes that agents misunderstand the statistical properties of the signal (Barberis et al., 1998; Rabin, 2002) or that they misperceive it (Rabin and Schrag, 1999). In Barberis et al. (1998), an investor believes that the shocks to her earnings are determined by one of two models characterized as Markov processes, while in fact shocks follow a random walk. In Rabin and Schrag (1999), an agent believing that certain state of nature is more likely may misread a challenging signal, believing in fact that it supports her current beliefs. In any of these models, no data points are omitted; the attention one pays to the signal or its timing are hence irrelevant factors. Another strand considers inference with limited attention or recall –see Gabaix (2019) for a comprehensive survey on the theoretical and experimental research on inattention. Most related to our approach, Mullainathan (2002) builds a model of limited memory based on the assumptions of rehearsal and association. Conversely to the previous models, some observations are assumed to be forgotten and any signal is more likely to be recalled when it has been already recalled in the past or when it looks similar to the current observation, so that the timing of each signal plays a crucial role in the formation of beliefs. Similarity-based recall is also present in Bordalo et al. (2017)'s model of norm recalling and choice, in which past events that are similar to the current event or that have taken place more recently or in a similar context are more likely to be recalled. In turn, the models of rational inattention by Caplin and Dean (2015) and Oliveira et al. (2017) take into consideration the trade-off between the benefits and the costs of acquiring information. These models are consistent with the evidence that stronger benefits improve recall –e.g., in Eysenck and Eysenck (1980), words associated to a large monetary incentive are better recalled than those associated to a smaller payoff.

For clarification, we note as well that several papers consider omissions of arguments in the agent's utility function, or elements of the sample space, but not omissions of data points. For instance, Gennaioli and Shleifer (2010) explore the implications of having a simplified mental representation of the state space. In line with the mechanism of anchoring and adjustment proposed by Tversky and Kahneman (1974), Gabaix (2014) considers an individual that rationally decides how much attention she pays to every relevant variable in a decision problem, so that the recalled value of each variable lies between a default value or anchor and the true one. Schwartzstein (2014) extends this analysis to situations in which the individuals must infer which variables provide more information –and therefore are worth more attention– from experience. Consequently, individuals may keep biased beliefs not only about the state of the world but also about the informativeness of the different variables.[59] Observe finally that our

[59] While these aspects are generally important, it must be noted that they are largely irrelevant in our experiment, where the draws from the urn are all of them equally informative.

focus on inattention also relates our study to a wide branch of the empirical and theoretical literature applied to, for example, tax policy (Chetty et al., 2009), health insurance (Abaluck and Adams, 2017) or finance (DellaVigna and Pollet, 2009).

3. Experimental design and procedures

The experiment is computerized and consists of two treatments (T1 and T2). Each treatment has two phases. In the first phase, any subject faces a virtual urn with 100 balls inside. Any ball in the urn is either blue or red. The space of potential rates (of red balls) is $\Theta_1 = \{0.3, 0.5, 0.7\}$ in Treatment 1 and $\Theta_2 = \{0.4, 0.5, 0.6\}$ in Treatment 2. Each subject in treatment $m \in \{1, 2\}$ knows that the computer determines the actual rate θ in her/his box by selecting a rate from Θ_m with uniform probability. Although subjects do not know the actual rate of red balls in their own urn, therefore, their priors are fixed. Each subject then observes the realization of an a priori undetermined number (in fact, 50) of consecutive random draws with replacement from her/his box.[60] Between the 42nd and the 43rd extraction, moreover, extractions stop for 90 seconds; meanwhile the screen just shows a message indicating that extractions will resume later.

After the 50 extractions are observed, the subject is asked to provide an *estimation* $\hat{\theta} \in \Theta_m$ of θ. This elicitation is incentivized, as the subject gets 10 euros if her/his point estimation is totally accurate. Participants cannot keep a record of previous observations and they are explained the estimation task *only after* the 50 drawings have been completed. Immediately afterwards, we asked subjects to indicate with a number from 0 to 100 how sure they were of their prior estimation of θ; this task was not incentivized. The instructions instructed subjects to think of this number as a probability, so that 0 (100)

[60] To ensure that all subjects really 'observe' the extractions, the outcome of each one is shown 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.

means absolute certainty that the response was incorrect (correct). In what follows, we refer to this number as a subject's *confidence*. Then subjects move to the second phase of the treatment, which is identical to the first one, except that they are aware from the beginning that they will have to estimate afterwards the percentage of red balls in the box (randomly determined anew), and can receive a prize if they are correct. We stress that the only difference between treatments lies in the set Θ of feasible rates.

In general, subjects in our experiment face a situation in which inference is cognitively easy, as it only requires counting the number of blue and red extractions, but where memory limits and inattention still introduce distortions due to data omissions. This explains some of our design choices: the time gap between the 42^{nd} and 43^{rd} extractions, the relatively large number of draws, the lack of feedback or records about previous extractions, and the subjects' ex ante unawareness about the belief elicitation (in the first phase) and the total number of extractions, i.e., 50. Reasoning as a Bayesian in our experiment, in other words, is arguably an undemanding task, particularly if one has good memory and pays sufficient attention, e.g., by counting the red and blue extractions. Deviations from the Bayesian standard (if any), therefore, must be largely explained by data omission, particularly taking into account that subjects had to show enough comprehension of the instructions (see below).

The experiment consisted of eleven computerized sessions, with a total of 133 participants. The software used was z-Tree (Fischbacher, 2007). Participants were undergraduate students at Universidad Autónoma de Madrid; they were not students of the experimenters. After being seated at a visually isolated computer terminal, each participant received written instructions for the first phase (see Appendix I). Subjects could read the instructions at their own pace and their questions were answered in private. Understanding of the instructions was checked with a control questionnaire that all subjects had to answer correctly before they could start making choices (see Appendix II). Instructions for the second phase (see Appendix I) were given only after the first belief elicitation.

The first two draws in each phase were shown together with a concise summary of the instructions, while for the rest of the draws the color of the ball drawn was the only information shown in the screen. Subjects did not observe others' samples and were never given any feedback about prior extractions. After the subjects' point estimations in the second phase had been elicited, they answered a brief questionnaire where we gathered personal information on socio-demographic characteristics (including gender, income level, major, religiosity, and political ideology), a risk aversion index,[61] and a cognitive reflection test or CRT (Frederick, 2005). We also checked their knowledge on Statistics (more precisely, the last year they had attended a course in Statistics). After the CRT, further, subjects performed a memory task: A list with 15 four-digit numbers was presented and they were given 60 seconds to memorize as many as possible. In the next screen they had to answer correctly two arithmetic problems: (a) (14*10) – 25 = ? and (b) (5*8) + 39 = ? Only when they solved these two problems, a new screen allowed them to introduce the recalled numbers; they were paid 0.50 euro for each correct number. After an additional, but unrelated set of questions, subjects were informed of their final payoff, thus ending the experiment. They were paid in private. Each session lasted approximately 1 hour (including payment), and on average subjects earned 17.09 euros, including a show-up fee of 5 euros.

[61] Subjects faced the choice between lottery A with prizes 2 and 1.6 Euros and lottery B with prizes 3.85 and 0.1 Euros, with equal probabilities of the larger and lower prize across lotteries. Letting P denote the probability of the larger prize, they had to indicate the threshold value of P such that they always preferred B to A, on a scale from 0 to 100.

4. Inference with omitted data: Evidence and theory

This section first introduces a basic analytical framework to understand learning about the rate of occurrence of some event when some signal observations can be omitted. The Bayesian model happens to be a particular specification, and we use several of its implications to organize the presentation of some aggregate results in Section 4.2. In 4.3, we explain how our framework and experimental data shed light on heterogeneous data omission and the associated biases. Then we employ a classification analysis to study heterogeneity. In addition, a regression analysis further explores the determinants of the subjects' estimations, confidence level, and other variables.

4.1. Estimating the frequency of an event: A theoretical framework

Let time be indexed as $t = 1, 2\ldots$. At period $T \geq 1$, an agent must solve an inference problem given some data. The outcome will be a probability distribution (the *beliefs*) over the set of frequencies or *rates* with which some phenomenon *x* can occur –or, alternatively, over the potential degrees of confidence that hypothesis *x* is true. More formally, let $\theta \in [0, 1]$ denote a generic rate and $\Theta = \{\theta_1, \theta_2, \ldots, \theta_K\} \subseteq [0, 1]$ denote the space of potential values of θ, with cardinality K. For expositional convenience, we assume unless otherwise noted that Θ is finite, although it is immediate to extend the model to some infinite Θ like the interval [0, 1]. A Bayesian agent called Eve has prior beliefs over Θ, quantified by a standard probability measure p. Let p_k denote Eve's priors about rate $\theta_k \in \Theta$, for $k \in \{1, 2, \ldots, K\}$ –e.g., uniform priors correspond to the case $p_k = 1/K$ for any $\theta_k \in \Theta$.

Eve has observed in each period some realizations of an i.i.d. signal S, taking on a value of either *x* or *y*.[62] Eve does not know the actual rate θ at which this signal is

[62] In the simplest setting, Eve directly observes the realization of the signal –e.g., the weather, performance in some task, the triggering of some emotion. It is also possible however that another agent

generated –i.e., probability (S = x) = θ– but can use the evidence observed to update her priors. Let Eve's history of observation of x be represented by a T-dimensional vector X = (x_1,…, x_T) where $x_t \geq 0$ is an integer representing the number of x-valued realizations of S observed at t –note well that x_t can be larger than 1, if Adam observes several realizations at the same time, e.g., if he is presented some aggregate data. In addition, let Y denote an analogously defined vector so that y_t indicates the number of observations of y at t. The number of times that Eve has observed value x up to period T is denoted as x = $\sum x_t$, that of value y as y = $\sum y_t$, whereas the total number of observations is $\sum x_t + y_t$. Given *data* D = (X, Y), 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, which is the case in our experiment):

$$p_{k|D} = \frac{p_k \cdot \theta_k^x \cdot (1-\theta_k)^y}{\sum_{i=1}^{K} p_i \cdot \theta_i^x \cdot (1-\theta_i)^y} = \frac{\theta_k^x \cdot (1-\theta_k)^y}{\sum_{i=1}^{K} \theta_i^x \cdot (1-\theta_i)^y} \tag{1}$$

Consider now a limited agent (call him Adam) who can omit some data. Adam initially assigns the same prior as Eve to any $\theta_k \in \Theta$. When he observes any evidence, nevertheless, his beliefs over Θ do not change exactly as in expression (1). The intuition here is that Adam displays limited attention and memory and hence may forget or omit some observations of the signal. Formally, let I_t denote the *memory indicator* of any datum observed at time t ($I_t \in \{0, 1\}$ for any t), $\tilde{x} = \sum_{t=1}^{T} x_t \cdot I_t$ denote the recalled number of x-observations, and $\tilde{y} = \sum_{t=1}^{T} y_t \cdot I_t$ the recalled number of y-observations. Intuitively, I_t = 0 means that any observation at time t is not considered. Observe also that Eve's case corresponds to $I_t = 1\ \forall t$; that is, no data omitted. To form his posteriors, we posit that

observes a realization of the signal and communicates it to Eve. Both types of 'observations' are formally equivalent if Eve absolutely trusts the other agent.

Adam applies Bayes' rule, but using the recalled instead of the actual number of *x*- and *y*-observations (the last equality assumes uniform priors):

$$\tilde{p}_{k|D} = \frac{p_k \cdot \theta_k^{\tilde{x}} \cdot (1-\theta_k)^{\tilde{y}}}{\sum_{i=1}^{K} p_k \cdot \theta_i^{\tilde{x}} \cdot (1-\theta_i)^{\tilde{y}}} = \frac{\theta_k^{\tilde{x}} \cdot (1-\theta_k)^{\tilde{y}}}{\sum_{i=1}^{K} \theta_i^{\tilde{x}} \cdot (1-\theta_i)^{\tilde{y}}} \tag{2}$$

Two brief remarks follow. *First*, it is implicit in equation (2) that Adam is not aware that some *specific* observations may have been 'lost'. Otherwise, he would have non-degenerate beliefs about $\tilde{x}$ and $\tilde{y}$, and (2) should be changed accordingly. *Second*, in case Adam has never observed the signal, recalls nothing or paid no attention at all to the signal realizations, his priors are unchanged.

When applying the model to our experiment, we note first that Adam and Eve need not give the same estimate, even if the (objective) data available is the same. Formally, let $\hat{\theta} \in [0, 1]$ denote an agent's point estimation of the actual rate θ, and $(\theta - \hat{\theta})$ denote her/his error. In our experiment, the expected monetary prize is obviously maximized when the agent chooses the mode of the posterior distribution (or any of them if there are several) as the estimate of θ. As we now prove, this means that the *subjective* (empirical) *frequency* of *x*, i.e.,

$$\tilde{f} = \frac{\sum_{t=1}^{T} x_t \cdot I_t}{\sum_{t=1}^{T} (x_t + y_t) \cdot I_t} = \frac{\tilde{x}}{\tilde{x}+\tilde{y}} \tag{3}$$

becomes the optimal estimator for Adam, whereas the actual empirical frequency $f = x/(x + y)$ is optimal for Eve.[63]

[63] A caveat is due when applying this result to the specific realm of our experiment, as $\tilde{f}$ need not equal any of the rates in Θ_m, the space of rates in treatment m. Still one can easily prove that Adam's optimal estimation must be always one of the two available rates closest to $\tilde{f}$. We take care of this point at the start of Section 4.2.

Proposition PE (point estimation): Assume that both f and $\tilde{f}$ belong to Θ. Then Adam's expected payoff is maximized when $\hat{\theta} = \tilde{f}$, and Eve's when $\hat{\theta} = f$.

Proof: Without loss of generality, we suppose Θ = [0, 1] so that (uniform) priors are defined on a continuum and optimization techniques can be applied. While this changes the absolute value of the posteriors given any data D, it does not affect their relative position –in expressions (1) and (2), the denominator is now an integral; but since this integral takes the same value for any $\theta_k \in \Theta$, this change only involves a monotone transformation of the posteriors. For Adam, the maximization of the logarithm of posterior (2) over the domain Θ implies after some standard calculus that the posterior mode equals $\tilde{f}$. Hence $\hat{\theta} = \tilde{f}$ maximizes the probability that the error is nil, in turn maximizing the expected monetary prize. The argument for Eve is analogous. ▪

Although the point is possibly obvious, it is important to stress that the consideration of a subsample of the objective sample leads to biases. In addition, the magnitude of the bias, $|\tilde{f} - f| = |\frac{\tilde{x}}{\tilde{x}+\tilde{y}} - \frac{x}{x+y}|$, depends on the timing of the observations, represented by vectors X and Y. The next corollary, which follows directly from the previous proposition, explores this point a bit further. We distinguish between over-estimation ($f < \tilde{f}$) and under-estimation ($f > \tilde{f}$); observe that over-estimation occurs when $\frac{x}{y} < \frac{\tilde{x}}{\tilde{y}}$ and under-estimation when this inequality is reversed.

Corollary DB (determinants of bias): For a fixed x and y, the magnitude of the bias depends directly on $\tilde{x}$ and inversely on $\tilde{y}$ in case of over-estimation, whereas the opposite relationship holds in case of under-estimation.

Several insights can be gained if attention is restricted to indicators of the type $\mathrm{I_t} = 1$ ($\mathrm{I} = 0$) for any $\mathrm{t} > T - \mathrm{t}^*$ ($\mathrm{t} \leq \mathrm{T} - \mathrm{t}^*$), where t^* is an integer ($0 \leq \mathrm{t}^* \leq \mathrm{T}$). This 'recency' specification has a straightforward intuition: When solving an inference problem on S, Adam considers only the observations of the last t^* periods.[64] In this case, $\tilde{x}$ and $\tilde{y}$ coincide with the number of observations of outcomes *x* and *y* in the last t^* periods, respectively, so that over-estimation occurs when the ratio of *x* and *y*-observations is larger in this subsample than in the whole sample. Further, the magnitude of the bias increases ceteris paribus as this subsample ratio increases. The case of under-estimation is analogous. The example illustrates as well that the average time and last time of perception of outcome *x* affect the magnitude of the bias, even if the actual empirical frequency *f* is unchanged. In our example, more generally, observations perceived far ago are relatively neglected, thus hardly affecting the posteriors –e.g., if a person estimates the rate of divorce in his community by recalling divorces among his acquaintances, recent divorces will substantially affect his estimation; Tversky and Kahneman (1973).

4.2. Data summary and first evidence on biases

Since biases are defined with respect to the Bayesian standard, we organize the exposition here by means of several of its predictions in our experiment. For this, let $\hat{\theta} \in \Theta_m$ denote a subject's point estimation in some phase of Treatment m, given the sample observed in that phase. We have seen that for a Bayesian subject like Eve, the optimal estimate tracks the frequency of red balls drawn from the urn in the phase. For Treatment 1, specifically, one can easily prove by comparing the posteriors of each feasible rate that (1) is maximized when Eve's estimation $\hat{\theta}_{1B}$ is given by:

[64] If the cause of this is that Adam forgets old observations, t^* can be thought of as a proxy for Adam's memory capacity.

$$\hat{\theta}_{1B} = \begin{cases} 0.3 & if \quad f \leq 0.3971 \\ 0.5 & if \quad 0.3971 < f \leq 0.6029 \\ 0.7 & if \quad f > 0.6029 \end{cases} \tag{4}$$

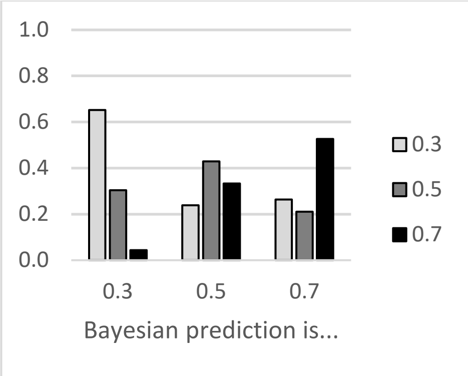
Analogously, Eve's estimation in Treatment 2 is ($\hat{\theta}_{2B} \in \Theta_2$)

$$\hat{\theta}_{2B} = \begin{cases} 0.4 & if \quad f \leq 0.4497 \\ 0.5 & if \quad 0.4497 < f \leq 0.5503 \\ 0.6 & if \quad f > 0.5503 \end{cases} \tag{5}$$

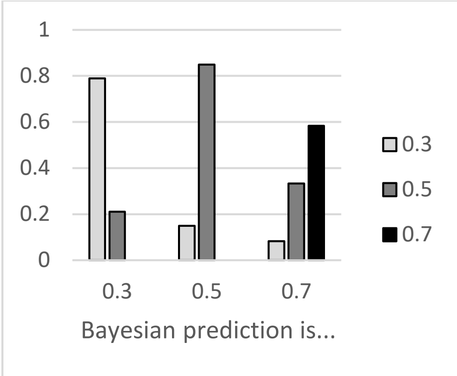
To sum up, therefore, the estimation by a Bayesian subject of the rate of red balls in our experiment *basically* accommodates to the following pattern or rule of thumb:

Hypothesis 1 (Bayesian estimation): In each phase of Treatment m ∈ {1, 2}, any participant chooses the rate in Θ_m that is closest to the empirical frequency of red balls in the sample of 50 extractions observed by her in that phase.

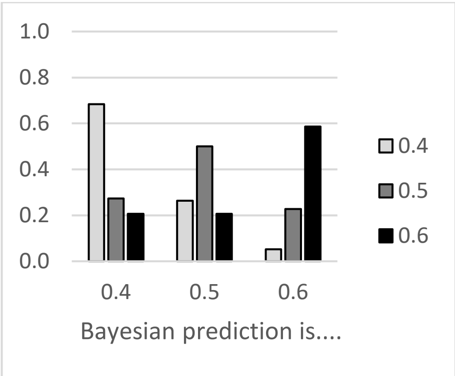
Evidence: Figure 1 shows the distribution of estimations for each phase of each treatment, conditional on the value of f observed by the subject. We distinguish the three intervals for f in either (4) or (5), depending on the treatment considered –see Tables A to E in Appendix II for more detail. For brevity, more precisely, the X-axis of each graph indicates the Bayesian estimate for each interval –e.g., when $f \leq 0.3971$, the estimate is 0.3 in T1; see (4) above. For each interval/Bayesian estimate, furthermore, the three bars above show, from left to right, the proportion of subjects who respectively estimate a rate of 0.3, 0.5 and 0.7 in T1 (0.4, 0.5 and 0.6 in T2). As an illustration, the black, right-hand bar in the graph for T1, phase 2 (right-hand, top) takes a value slightly lower than 0.6 (in fact equal to 0.583; see Table E in the appendix). This means that 58.3% of the subjects made a point estimation of 0.7 when $f > 0.6029$, that is, when the Bayesian estimation happens to be 0.7, as (4) indicates. Consequently, the graph not only provides a summary of the subjects' estimations but also allows us to explore whether they act according to the Bayesian prediction (see Table E for the precise percentage of subjects who make each Bayesian estimation, conditional on the empirical frequency observed).



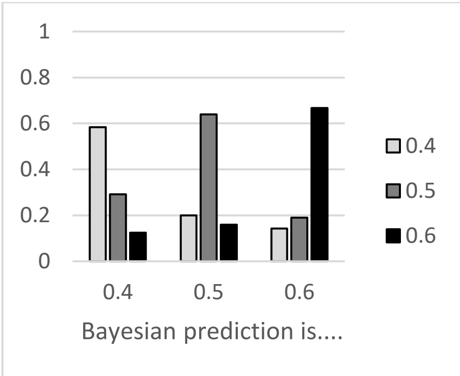
Treatment 1, phase 1



Treatment 1, phase 2



Treatment 2, phase 1



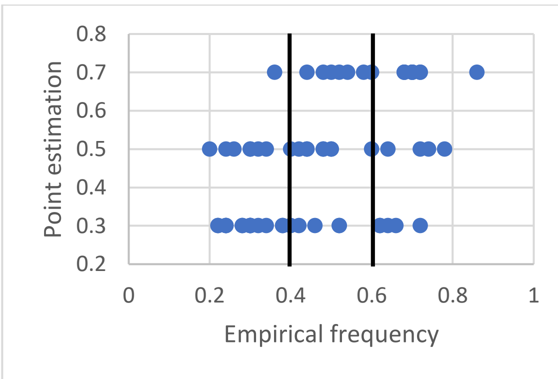
Treatment 2, phase 2

Figure 1: Distribution of participants' estimations conditional on frequency observed (and hence Bayesian prediction)

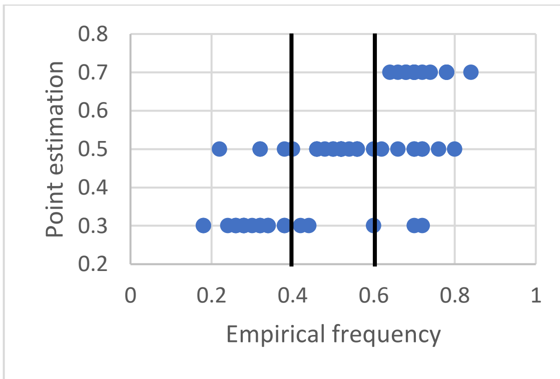
We remark that the frequency of Bayesian estimations happens to be 62.03% across all phases and treatments and equals 63.49 and 60.71% in Treatments 1 and 2, respectively. The mentioned total proportion of Bayesian predictions is significantly larger than we should expect if participants answered randomly (binomial probability test, $\Pr(k \geq 165) < 0.000$). Note also that the overall frequency of Bayesian estimations does not differ significantly across the intervals in (4) and those in (5); χ^2 test, $p = 0.327$ and 0.851 in Treatments 1 and 2, respectively. Similarly, the distribution of Bayesian estimations in each phase is not statistically different across treatments (χ^2 test, $p =$

0.593 for the first phase and $p = 0.211$ for the second). Regarding differences between phases, the probability of a Bayesian prediction was significantly higher in the second phase in T1 (McNemar test, $p = 0.014$) but not in T2 ($p = 0.364$), a point to be explored later.

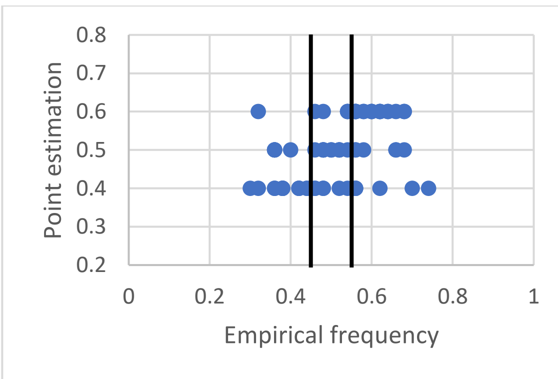
We observe a substantial amount of heterogeneity in the subjects' responses. While Figure 1 already hints this point, the phenomenon is more evident in Figure 2, where each participant is plotted according to her/his estimation and the actual empirical frequency –the two vertical lines in each graph are the bounds of intervals (4) or (5), depending on the treatment. Observe that different participants make different estimations even when the empirical frequency actually observed is basically identical. The next section explores heterogeneity in detail.



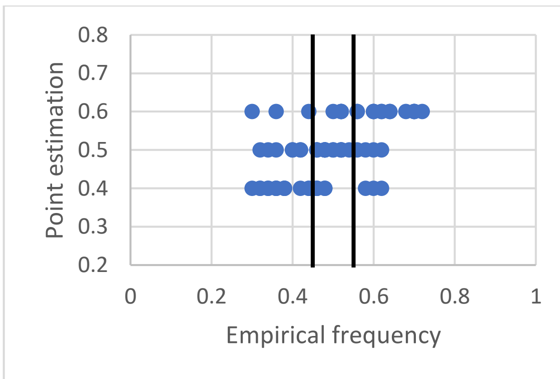
Treatment 1, phase 1



Treatment 1, phase 2



Treatment 2, phase 1



Treatment 2, phase 2

Figure 2: Point estimation vs. empirical frequency

Result 1: Around 62% of the estimations track the empirical frequency actually observed. In a comparison across phases, the rate of Bayesian estimations increases only in Treatment 1. Substantial heterogeneity is observed, as subjects often respond differently to essentially the same evidence.

Although our focus is mostly on participants' estimations, recall that they also indicated how sure they felt about their own estimation, by assigning a probability to the event that this prediction was right. A Bayesian agent would use expression (1) to compute such probability. To check this, we compute (1) for each subject, given the extractions observed, and find that the mean predicted confidence in the first phase is 0.905 in T1 and 0.684 in T2, while in the second phase it equals 0.921 and 0.698, respectively. As expected, these probabilities differ across treatments, which is natural because the inference problem in Treatment 1 is arguably 'easier' than in Treatment 2.

Hypothesis 2 (Bayesian confidence): In each phase of any treatment, participants estimate the probability that their estimation is correct according to expression (1). Given the evidence actually available to the agents, this implies that the mean confidence is higher in each phase of Treatment 1.

Evidence: Contrary to the hypothesis, the mean confidence for the first phase was 0.658 in T1 and 0.677 in T2, and there are not significant differences between treatments in this regard (Mann-Whitney test, $p = 0.462$). In the second phase, the average confidence was 0.681 and 0.670 in Treatments 1 and 2, respectively; again, a non-significant difference($p = 0.935$). In turn, when we compare the predicted confidence (see above) and the actual distribution in each phase and treatment, we find significant differences in T1 (Wilcoxon signed-rank test, $p < 0.000$ in both phases) but not in T2 ($p = 0.988$ and 0.671 in phases 1 and 2, respectively). We additionally remark that the distributions in phases 1 and 2 are not independent but present a positive correlation, both in T1 (Spearman's $\rho = 0.311,\ p = 0.013$) and T2 (Spearman's $\rho = 0.363,\ p = 0.002$).

Result 2: Confidence levels are similar across treatments, although T2 is a substantially more difficult problem.

4.3. Exploring heterogeneous data omission: Classification analysis

The subjects' point estimates, together with the data actually observed, can be used to analyze the extent of data omission, and the differences across subjects in this respect. In effect, note that conditions (4) and (5) above apply for Adam as well, except that they are conditions on $\tilde{f}$, not *f*. Consequently, Adam's estimation in each phase of each treatment must track his corresponding subjective frequency. Conversely, we can infer a subject's $\tilde{f}$ in each phase from his estimate (we allow $\tilde{f}$ to vary across phases, as seems most natural). For this, we use a classification analysis.

Let (I_1, I_2, …, I_T) denote a 'consideration vector' indicating the value of indicator $I_t \in [0, 1]$ at any time t = 1,..., T, where T = 50 in our experiment. Formally, a *model* is any collection of k consideration vectors (not necessarily different) together with a mapping assigning each participant to one of those vectors. That is, a model is a representation of the subjects' heterogeneity with respect to data omission. Our goal is to find the most empirically relevant models with one (k = 1) and two vectors (k = 2). We focus our analysis for the moment on consideration vectors such that $\mathrm{I_t} = 1$ ($\mathrm{I_t} = 0$) for any $\mathrm{t} > \mathrm{T} - \mathrm{t}^*$ ($\mathrm{t} \leq \mathrm{T} - \mathrm{t}^*$) or $\mathrm{I_t} = 1$ ($\mathrm{I_t} = 0$) for any $\mathrm{t} \leq \mathrm{t}^*$ ($\mathrm{t} > \mathrm{t}^*$), where t^* is an integer ($0 \leq \mathrm{t}^* \leq$ T). In other words, participants consider either some of the 'first' or some of the 'last' observations. If we denote the vector that 'considers' the first (last) t^* observations as t^*_0 (0_t^*), examples of models in this class include [25_0; 0_10], [15_0; 45_0], [0_37; 0_1], and so on. Note that the "Bayesian" vector could be written either as

50_0 or 0_50; in this sense, the "Bayesian model" will be sometimes referred simply as 50_0 (recall that the two vectors in a model can be identical). Our focus on models with two vectors allows us to learn about heterogeneity without compromising the tractability of the analysis. For simplicity, moreover, we are assuming at this stage that subjects either display primacy or recency effects, although we will later allow them to display both simultaneously. Of course, one could also analyze different consideration vectors, but given the evidence available, we find unlikely, say, that people extrapolate from some sample of intermediate observations, excluding therefore some initial and final signals.

To determine the optimal model in each phase, we pool the data across Treatments 1 and 2, as we do not expect differences in the patterns of omission. Then we proceed in four steps for each possible model (note that there are 4950 possible pairs of vectors, including homogeneous models in which both vectors coincide). First, we compute for each subject the estimate of θ predicted by each of the two consideration vectors in the model, given the extractions actually observed by the subject and the respective considered subsamples, as defined by each vector. For each participant, second, we find the difference (in absolute terms) between her actual estimate in the phase and each predicted estimate. Third, the subject is assigned to the vector with the smallest difference or error. Fourth, we compute the model error as the average individual error of the vectors assigned to the participants. The model error hence provides a measure of the goodness of fit of the model. For example, an error of 0.05 means that, after mapping each subject to the vector that fits best her estimations, there is still an average absolute difference of 0.05 between the vector prediction and the participant's actual prediction. Incidentally, this can be compared with the expected error of an individual making random estimates, which is 0.178 and 0.089 in Treatments 1 and 2, respectively, and hence equal to 0.1335 overall.

Figure 3 below conveys graphically a lot of information about our results in the first and the second phase of the experiment. Each square is a matrix with 99 rows and columns, each one corresponding to a different consideration vector within the class of vectors considered. Starting from the origin, the vectors are ordered as 1_0, 2_0,…, 49_0, 50_0, 0_49,…, 0_2, 0_1 in each axis. A cell in the square represents therefore a model, e.g., the fourth row and the sixth column, starting from the origin, define model [4_0; 6_0]. In this vein, the diagonal or 45º line corresponds to the homogeneous models; note also that each square is symmetric along that line. To compare all the potential models in terms of their error, we use lighter tones for the models with lower error.

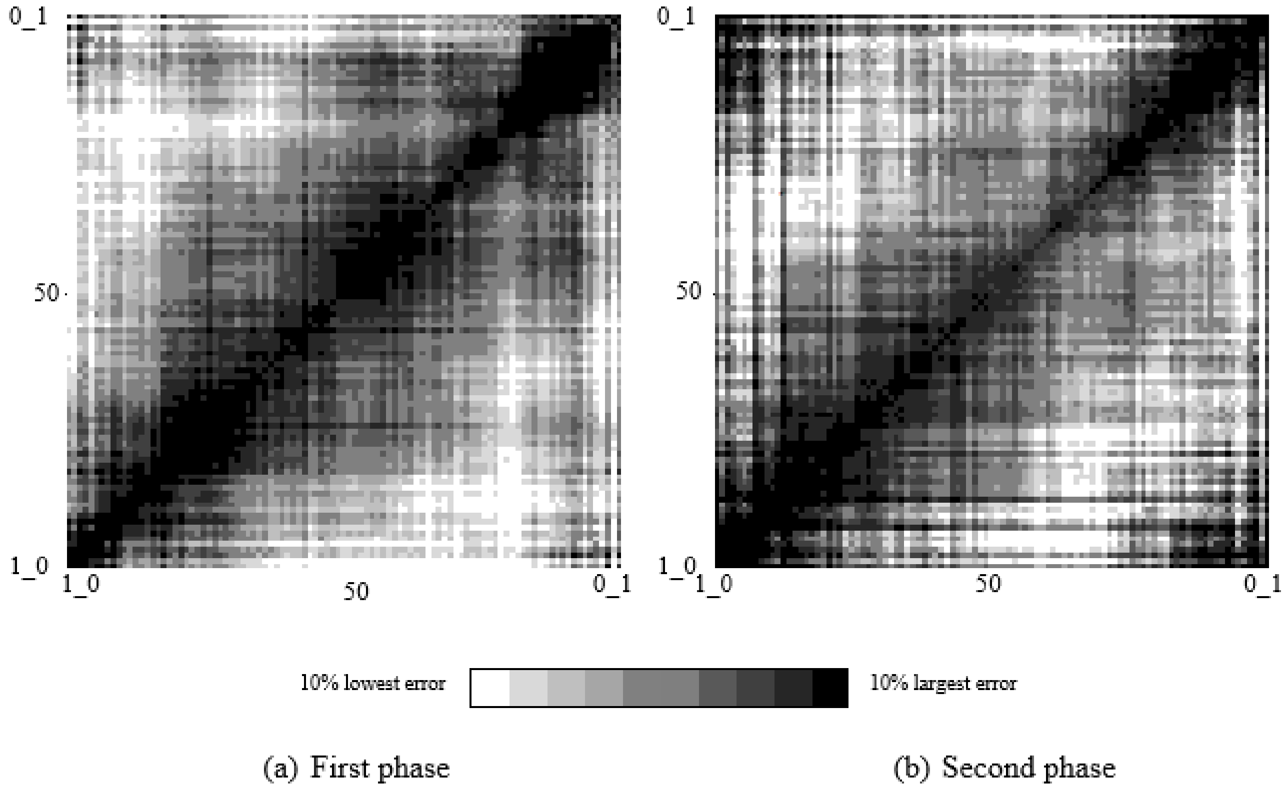


Figure 3: Errors of all models

There are a number of remarkable things in these figures. First, the cells in the diagonal appear in dark colors, so that homogenous models perform relatively badly. Second, not every heterogeneous model performs equally well. For instance, models in which *all* subjects consider only a few observations (either the first or the last ones; see the upper, left-hand corner in each square) appear to fail considerably, especially in the second phase. In contrast, models in which (i) some subjects consider many observations and (ii) the remaining subjects consider relatively few observations work much better; in fact, these models often are among the top 5%. Third, the differences between the first and the second phase of the experiment are not apparent, but homogeneous models seem to improve, particularly those close to the Bayesian model.

For further comparison of the different models it is convenient to resort to some aggregated evidence. For this, we divide each square of Figure 3 into different regions, and compute the average model error of all models in a region. In that manner we can compare the relative performance of models in which, say, one vector considers a few initial observations while the other considers many of the last ones, etc. Specifically, we distinguish four homogeneous regions and six main types of heterogeneity. Table 1 shows the average model error of each of these regions, distinguishing between phases. The upper number in each cell is the average error in the first phase of all models in which the first consideration vector includes some or all of the observations in the corresponding row, whereas the second vector considers observations in the range of the corresponding column. As an illustration, the number 0.052 in bold and underlined is the average error in the first phase of all models such that the first vector considers between 17 and 34 of the last observations, while the second vector considers between 1 and 16 of the first observations. Plainly speaking, this is the average error of those models that assume that some subjects tend to focus on a 'fairly high' number of the last observations, whereas others focus on the first few observations. The number below in each cell corresponds to the second phase.

Second consideration vector		First … observations			Last … observations		
First consideration vector		**1-16**	**17-34**	**35-50**	**35-50**	**17-34**	**1-16**
Last … observations	**1-16**	0.058 0.051	0.059 0.043	0.058 0.042	0.064 0.041	0.067 0.046	0.082 0.061
	17-34	**0.052** 0.042	0.058 **0.040**	0.062 0.044	0.068 0.045	0.073 0.051	
	35-50	**0.054** **0.040**	0.065 0.043	0.070 0.047	0.077 0.049		
First … observations	**35-50**	0.058 0.044	0.070 0.049	0.074 0.053			
	17-34	0.068 0.050	0.078 0.055				
	1-16	0.076 0.063					

Note: In each cell, the number above (below) corresponds to the first (second) phase. N = 133 in both phases. The numbers underlined and in bold correspond to the two groups of models in each phase with lowest average error, i.e., the two best families of models. The table is symmetric; for that reason, some cells are left empty.

Table 1: Average error of each family of models (first and second phase)

Although perhaps not apparent at first sight, our results indicate that the models that best fit our data are characterized by two consideration vectors that differ in the *location* of the considered sample. Specifically, models in which some individuals consider some share of the last observations while others focus on a number of the first observations seem to fit particularly well the data. That is, people are better characterized as a mix, with some exhibiting recency effects and others primacy effects, rather than all of them exhibiting a single effect. For first evidence in this line, we consider the three non-empty quarters of Table 1 and compute the average error of all models within each one. We find that the average error of all models where both vectors consider some number of the last observations, i.e., the upper, right-hand quarter, equals 0.070 and 0.047 in the first and second phase, respectively. In turn, the average error when only primacy effects are possible, i.e., the lower, left hand-quarter, is 0.069 and 0.051 in the first and second phase, respectively. Finally, the average error of the 'mixed' models, i.e., those in the upper, left-hand quarter, is 0.059 and 0.043 In Phase 1 and 2, respectively. If we moreover focus our attention on these mixed models, it seems that the best ones are those where one type exhibits primacy effects over a relatively small sample, whereas the other type displays recency effects in a relatively large sample. That is, people exhibit either primacy or recency effects, but to a different extent. To clarify this point, we present in Table 1 the average errors of the best two groups of models in each phase underlined and in bold.

A more fine-grained analysis involves the comparison of individual models. In the first phase, the model that best fits our data is the heterogeneous model [2_0; 0_35], with an error of 0.041, while the best-fitting homogeneous model, [0_35; 0_35], has an error of 0.077 and the homogeneous Bayesian model [50_0; 50_0] has an error of 0.080. We note that the median individual error is significantly lower in the optimal model than in the best homogeneous model and the Bayesian model (one-sided sign test, $p < 0.000$). On the other hand, there are no significant differences between the median individual error in the best homogeneous model and the Bayesian model (one-sided sign test, $p = 0.434$). The optimal model in the first phase has a possibly natural interpretation: Some subjects pay scarce attention to the observations, mostly on the first ones (maybe they get bored soon), while other subjects pay more attention and as the extractions proceed they possibly realize that it is a good idea to count the number of red and blue drawings. In this vein, the heterogeneous model [2_0; 50_0], which assumes that part of the individuals are purely Bayesian, also fits well the data, with an error of 0.042. In fact, there are not significant differences between the median individual error of models [2_0; 0_35] and [2_0; 50_0] (one-sided sign test, $p = 0.500$). Considering model [2_0; 0_35], finally, we note that vectors (2_0) and (0_35) explain strictly better the estimates in the first phase of 21.8% and 40.6% of the subjects, respectively, whereas both vectors fit equally well the estimates of 37.6% of the participants.

In the second phase, the best-fitting model is [12_0; 0_33], with an error of 0.031, while the homogeneous Bayesian model has an error of 0.053 and a significantly greater median individual error ($p < 0.000$). It must be noticed though that the comparison between the heterogeneous models [12_0; 0_33] and [2_0; 0_35] –the latter one, recall, is the optimal model in the first phase– does not show significant differences in the second phase between their median individual errors (two-sided sign test, $p = 0.5034$). Since many of the best models exhibit no significant differences in their performance, the regression analysis in the next section focuses conservatively on model [2_0; 50_0], the best performer under the restriction that some subjects are Bayesian.[65] We note that this model perfectly replicates 74.44% of the subjects' estimates across both phases, whereas model [2_0; 0_35] replicates 74.81 % of them. Even if we consider minimal and unvaried heterogeneity (the same model with two vectors for both phases), therefore, a heterogeneous model *significantly* improves the Bayesian one, which replicates 62.03% of the estimations. This is remarkable because, as we stressed in the introduction, it suggests that data omission is relevant even in very stylized scenarios as this one.[66] Let us add that, when subjects are classified within model [2_0; 50_0], we have that the estimates of 40.6 % (36.1 %) of them are strictly better explained by the Bayesian vector (50_0) in the first (second phase). Further, both vectors explain equally well the estimates

[65] Since many of the best models exhibit non-significant differences, our focus on a single model can be misleading (although it facilitates further comparison with the Bayesian model). We stress therefore that, from our point of view, the main take-home message from our classification analysis is the aggregate data in Table 1, clearly suggesting that subjects do not always weight equally all observations in their inferences. Incidentally, we have abstained from conducting a classification analysis with more than two types because, from our point of view, the key implications from the analysis can be conveyed by focusing on two types, i.e., heterogeneous data omission is a significant phenomenon, even in simple settings. Further, any additional insights like a more accurate picture of heterogeneity are likely to be too specific, that is, difficult to extrapolate to other scenarios.

[66] A brief clarification is in order, though: given our procedure, it seems pretty clear that any model that allows for two types of individuals must perform –in terms of average error– better or at least no worse than a homogeneous model that considers exclusively one of those types. That said, this does not guarantee at all that the different performance is quantitatively relevant or, more important, statistically significant, as is our finding.

of 36.8 % and 54.1 % of the subjects in Phase 1 and 2, respectively. The remaining subjects are unequivocally assigned to vector (2_0). Overall, therefore, the most conservative evaluation of the extent of data omission indicates that the percentages of Bayesian participants are 77.44 and 90.23 in the first and second phase, respectively. The next result summarizes our main findings so far:

Result 3: A model that assumes that agents omit data and are heterogeneous in this regard (two types of agents) replicates 74.81% of the subjects' estimations and significantly outperforms the Bayesian model. In the best heterogeneous models (in both phases), one type of agents estimates based on a limited number of the initial observations, while the second type uses a larger subsample, apparently of the last observations.

As a final remark, our classification analysis has considered as well a much wider family of consideration vectors such that $\mathrm{I_t} = 1$ for any $\mathrm{t} \leq a$ and $\mathrm{t} > T - b$ (and $\mathrm{I_t} = 0$ otherwise), where a and b are integers ($0 \leq a, b \leq \mathrm{T}$, $0 \leq a + b \leq \mathrm{T}$). In other words, we have explored as well the possibility that Adam considers both the first a and the last b balls (observe that this family of vectors includes the vectors that we analyzed before; i.e., when Adam considers only the first *or* only the last balls). For $\mathrm{T} = 50$, there are 1,275 different vectors and 813,450 different models. In spite of the potential benefits of considering a much wider range of models, the results obtained do not significantly improve our previous analysis. In the first phase, for example, the model that best fits our data is [2_3; 44_0], with an error of 0.039, while the optimal model in our prior analysis, i.e., [2_0; 0_35] has an error of 0.041; the median individual error is not significantly lower in the former model (one-sided sign test, $p = 0.500$). For the second phase, the optimal model is [7_3; 0_33], with an error of 0.027. In turn, the optimal model in our previous analysis, i.e., [12_0; 0_33], has an error of 0.031 and, again, we cannot conclude that the median individual error is significantly lower in the first than in the latter (one-sided sign test, $p = 0.304$). Across both phases, models [2_3; 44_0] and [7_3; 0_33] replicate 76.32% and 75.56% of the subjects' estimations, respectively.

4.4. Determinants of the individual's error, estimation, type, and confidence

This section explores a number of questions. The first one studies what factors influence a subject's estimation in T1 and T2, and complements our prior classification analysis, analyzing further the empirical relevance of model [2_0; 50_0]. Specifically, Table 2 below reports the results from two regressions, one for each treatment. The dependent variable is a subject's estimations in the two phases of the corresponding treatment. In turn, the right-hand side includes (i) phase, which takes value 0 for the first phase and 1 for the second one; (ii) the actual frequency of red balls in the corresponding phase; and (iii) the frequency of red balls in the first two extractions. We also interact variables (ii) and (iii) with the subject's type in the phase, which is a binary variable taking value 1 (Bayesian) if, given the observed sample in the considered phase, vector 50_0 predicts the subject's estimate with a smaller (or equal) error than vector 2_0. In other words, this dummy indicates how the subject is classified in a phase according to model [2_0; 50_0]. Additionally, we consider controls like the subject's gender, income level, major, religiosity, ideology, knowledge on Statistics, index of risk aversion, number of correctly recalled numbers in the memory task, and CRT results.

The results from regressions (A) and (B) suggest that model [2_0; 50_0] performs comparatively well in explaining point estimations. In both treatments, in effect, the probability of a large estimation increases significantly with the frequency of the participant's considered subsample. Specifically, for the non-Bayesian participants, only the frequency of the small subsample has a significant effect on their point estimations. Conversely, for the Bayesian individuals, the frequency of the whole sample has a positive and significant effect on their point estimations. Hence, the regression analysis confirms our prior classification analysis. More than this, the model explains a considerably larger proportion of the variance ($R^2 = 0.50$ and 0.31 for Treatments 1 and 2, respectively) than a model that considers only the frequency of the whole sample and the control variables ($R^2 = 0.24$ and 0.17 for Treatments 1 and 2, respectively).

Model	Predictors							
	Freq. 50	**Freq. 2_0**	**Type x Freq. 50**	**Type x Freq. 2_0**	**Phase**	**CRT**	**Memory**	**Sex**
(A)	-3.16 (1.92)	17.32*** (2.93)	20.69*** (3.23)	-17.84*** (2.95)	-0.76 (0.49)	0.48 (0.51)	-0.83* (0.50)	-0.20 (0.59)
(B)	-4.52 (3.45)	20.19*** (4.96)	23.03*** (5.18)	-22.46*** (5.05)	0.15 (0.37)	0.43 (0.40)	-0.25 (0.46)	-0.25 (0.40)

Note: Ordered logit regressions are used in both (A) and (B). The dependent variable in model A (B) is the subject's estimations in each phase of T1 (T2); N = 126, 140, respectively. The pseudo-R^2 in A and B equals 0.5 and 0.31, respectively. Robust standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2: Regression analysis: Determinants of the subjects' estimations

In turn, model (C) in Table 3 refers to the determinants of a subject's deviation from model [2_0; 50_0], with the data pooled across treatments. The regression is a probit where the dependent variable is an indicator taking value 1 when our prediction and the subject's actual estimation coincide. Among the independent variables, we consider (i) treatment (T2 = 1), (ii) phase, and a binary variable labeled (iii) critical region, which takes value 1 when the empirical frequency *f* is near the Bayesian decision thresholds. Specifically: if *f* is in the interval [0.35, 0.45] or [0.55, 0.65] in Treatment 1 (and in interval [0.4-0.6] in Treatment 2), the variable takes value 1, and 0 otherwise. Arguably, inference is more complex *for a Bayesian type* when the frequency falls within these intervals than out of them –these subtleties seem in contrast irrelevant for a non-Bayesian type, given the tiny subsample considered. We also add the same controls as in regressions (A) and (B) above.

Predictors	(C)	(D)	(E)	(F)
	Probit	OLS	Probit	OLS
	Estimation = optimal model prediction	Subject's error	Type	Confidence
Phase	-0.71 (0.65)	0.05 (0.05)	0.58*** (0.20)	0.99 (2.41)
CRT	0.52*** (0.18)	-0.03*** (0.01)	0.61*** (0.22)	0.89 (2.15)
Memory	0.01 (0.21)	0.01 (0.01)	-0.03 (0.23)	2.45 (2.71)
Sex	0.13 (0.20)	-0.01 (0.01)	-0.42* (0.22)	8.97*** (2.44)
Critical region (CR)	0.04 (0.63)	0.02 (0.04)		-0.33 (2.43
Treatment	-0.12 (0.25)	-0.02 (0.01)		0.02 (2.58)
Type	-0.61 (0.52)	-0.13*** (0.04)		-1.80 (3.66)
Type x CR	0.69 (0.66)	-0.03 (0.04)		
Phase x CR	0.79 (0.97)	-0.08 (0.05)		
Phase x Type	1.25* (0.70)	-0.08* (0.05)		
Phase x Type x CR	-1.94* (1.05)	0.14** (0.06)		
Constant	0.06 (0.65)	0.25*** (0.04)	1.04** (0.43)	58.19*** (7.42)
(Pseudo-)R^2	0.10	0.46	0.11	0.07
N	266	266	266	266

Table 3: Regression analysis; determinants of subjects' errors, type and confidence

Our results indicate that overall subjects conform more closely to model [2_0; 50_0] if their type is Bayesian (in the second phase) and the higher they score in the CRT. Observe in this regard that the results of the CRT appear to be a good measure of how willing someone is to meditate on a problem (e.g., Frederick, 2005). Since we also find a significant, negative correlation between the probability of conforming with the model and the critical region (for the Bayesian types in the second phase), we infer reassuringly that a significant share of the deviations from model [2_0; 50_0] are not due to a different logic of inference, but basically to mistakes, which tend to diminish if a subject is attentive, particularly if she has some experience and the frequency that she observes is not very close to the Bayesian theoretical thresholds (4) or (5).

Column (D) explores the determinants of a subject's bias. Specifically, the dependent variable in (D) is the difference in absolute terms between the subject's actual estimation and the Bayesian one. We pool again the data across treatments. We find that the so-called Bayesian types commit a relatively lower error, particularly in the second phase and when the problem is easy (variable CR). This is straightforward because these types tend to consider the whole sample. In contrast, the non-Bayesian types operate with a reduced subsample, often dissimilar to the full sample, and hence their error tends to be larger in average.[67]

The model in column (E) explores the determinants of a subject's type using a probit model. Our analysis suggests that both the phase and the CRT play an important role. Since CRT is a measure of the willingness to exert cognitive effort, it is not surprising that individuals with a higher score tend to consider also larger samples, all other things being equal. As we discuss later, further, we expected the share of Bayesian types to be higher in the second phase, since only in that phase did subjects know from the outset that (i) they would be asked to estimate the number of red balls in the urn and that (ii) they would obtain a payment of 10 euros in case that their estimation was correct. We stress also that memory, as measured by the score obtained in the memory tasks, is

[67] Note also that there seems to be no treatment effect in what regards a subject's error. We have received the comment, however, that there is an issue with the interpretation of the corresponding coefficient. In the second treatment, that is, the subjects' potential answers are closer together than in the first. The coefficient on the Treatment variable therefore bundles together the mechanical effect of pulling the answers closer together with the effect of making the task harder, and these effects should go in opposite directions, masking the actual treatment effect. However, we have also conducted a probit regression where the dependent variable is an indicator taking value 1 when the subject's estimation equals the Bayesian one, and there is again no significant treatment effect (results available upon request). Once we control for the effect of the CR, therefore, the treatment seems to have no additional effect on the deviations from Bayes' rule.

not correlated with the subject's type. It seems hence that in our experiment the key cognitive variable is the degree of attention, and not the subject's memory capacity. For further evidence in this respect, recall that the extractions from the urn were interrupted during 90 seconds between the 42nd and the 43rd draw in both phases. If memory was influential in determining the considered subsample, one might expect subjects to systematically over-weight the last 8 extractions in their inferences; this does not seem to be the case, given the results from our classification analysis.[68]

Overall, the analysis in columns (C) to (E) allows us to answer an important question: Do subjects improve across phases? More generally, do they act more in accordance with the Bayesian model when they have some experience with the estimation task and the incentives are made unambiguous? To explore this point, we compare the variance of the estimation errors across Phases 1 and 2 in each treatment. As we explained in the introduction, a potential rationale for a reduction in the variance is that subjects are certain in Phase 2 that paying attention to the extractions increases the likelihood of a reward, which should increase the accuracy of their estimations.

One signal in favor of our conjecture is that, in regression (E), we find a positive and very significant correlation between type and phase: Subjects act more as a Bayesian, i.e., considering the whole sample, in Phase 2. As a result, they commit smaller errors, as we see in regression (D). Interestingly, there is no significant overall effect of the variable phase in regression (D), whereas the variable type is indeed highly significant. This hints

[68] As we noted before, additional control variables were included in all the regressions, including variables related to familiarity with Statistics. Specifically, we included a dummy variable that takes value 1 for subjects who attended a course on Statistics during the last year and another one that takes value 1 if the participant's study program has a substantial mathematical background. None of the controls seems to play a significant role in any of the regressions conducted, with the notable exception of gender in regressions (E) and (F).

that the key change across phases refers to the considered sample: Less attentive subjects in the first phase tend to pay more attention to the extractions in the second one.

In summary, the evidence suggests that unambiguous incentives plus some experience with the estimation task alleviate subjects' errors, a key reason being that subjects pay more attention to the sample.[69] We stress though that this statement refers to the average effect across treatments. As we noted in Result 1, Bayesian estimations only increase significantly in T1, and a disaggregated analysis of regression (D) indeed points out differences across treatments (results available upon request). It seems that T2 is a substantially more complex problem than T1. One natural reason is that, by construction, the critical region (CR) in T2 includes the available rates, i.e., the elements of set $\Theta_2 = \{0.4, 0.5, 0.6\}$. This means that in a large sample (50 extractions) the frequency is relatively more likely to fall into the CR in T2 than in T1. Indeed, if we run regression (D) for each treatment, we find that the effect of the CR variable and its interactions is more accused in T2. This possibly explains why clear, unambiguous incentives are not so effective in T2.

Finally, column (F) reports results from an OLS regression where the dependent variable is the subject's confidence (in each phase). As noted in Section 4.2, there are not significant differences across treatments and phases. If subjects were purely Bayesian, the only coefficient significantly different from zero should be the one associated to the critical region. In effect, the level of confidence should be lower when the observed frequency of red balls is near the values of the decision thresholds stated in (4) and (5). For example, in Treatment 1, a Bayesian individual would be more confident of her estimation when the observed sample contains a 20 percent of red balls than a 38 percent:

[69] Note well, however that we cannot disentangle whether this effect is entirely due to learning, entirely due to the non-ambiguous incentives, or to both factors.

while in the first case it is rather clear that the amount of red balls contained in the urn is 30 with very high probability, in the second case chances are that there may be both 30 or 50 red balls in the urn with relatively high probability (it is just slightly more likely that there are 30 red balls in the urn). We could also conjecture that individuals should be more confident in the second phase if there are learning effects. Additionally, one might expect that, since non-Bayesian subjects focus on a sample of small size, their confidence should be relatively lower. Yet none of these hypotheses was supported by our analysis. The only relevant explicative variable found was the participants' gender. All else unchanged, male participants tended to state confidence levels 8.97 percentage points more than female ones.

5. Conclusion

We contribute here to the issue of how people infer frequencies and probabilities based on experience, a question that seems important for at least three reasons. First, accurate models of inference should improve our static predictions of decision-making under uncertainty. Second, beliefs can interact with choice in dynamic settings so that a theory of how expectations evolve may be crucial to account for phenomena like bank runs, market bubbles, and crashes. Third, if human inference is subject to biases which we deem undesirable for whatever reasons –e.g., because they lead to individual choices that, after reflection, most decision makers find suboptimal–, a model may suggest ways to reduce or prevent them.

Our first main contribution is developing a quasi-Bayesian framework for the analysis of inference with data omission. It aims to incorporate some of the insights associated to the availability heuristic, particularly the fact that fluency affects inference. Further, it helps to formally study several issues related to this heuristic, e.g., people can be heterogeneous in what they omit. The framework also allows a more precise analysis of the conditions under which biases appear, and hence their determinants. Indeed, it has several implications that contrast with the standard Bayesian model. *First*, posteriors depend not only on the frequency of observation of each outcome, but also on the timing of the observations. If an agent tends to focus on recent observations due to decay, for example, we observe recency effects –Hogarth and Einhorn, 1992; Hertwig et al., 2004. In this line, an agent's posteriors can be different depending on whether the realizations are observed sequentially or simultaneously, even if the information presented in both cases is identical. *Second*, beliefs may change even if no data is received, as the mere passing of time will affect the timing of the observations. In general, unstable memories make beliefs more volatile than in the Bayesian paradigm. *Third*, the magnitude of the bias depends on how different the recalled subsample is from the objective one, which in turn should depend on the attention paid by the agent (or his good memory).

The second key contribution of the paper is offering experimental evidence to explore a series of questions like (i) how heterogeneous people are regarding data omission, (ii) cognitive correlates of data omission, and (iii) the role of complexity in understanding the effect of learning and incentives on accuracy. Regarding heterogeneity, our main finding is that, although the majority of the participants behave in accordance to the Bayesian model, a non-negligible number of individuals seemed to rely on very small subsamples –especially in the first phase, in which the reward for solving the problem was not known in advance by the subjects. This is remarkable because the problem faced by the participants was arguably easier than many of the inference problems that common people face in their daily life. Further, our analysis suggests that the scope and the distribution of data omission among individuals is likely due to differences in the attention devoted to the problem, more than to differences in individuals' memory constraints. Finally, a traditional view on biases is that they can be corrected by experience and incentives. Our evidence suggests that these two factors together can alleviate data omission in our setting, although the improvement depends on how complex the inference problem is. Thus, the improvement induced by the introduction of a non-ambiguous monetary incentive seems much clearer in Treatment 1, which was easier, than in Treatment 2.

In future research, we plan to consider additional treatments with higher prizes or with feedback about mistakes (informing that the estimation was inaccurate, or indicating the magnitude of the error): Which are more effective? Other treatments might explore inference with data omission in complex environments with many signals and outcomes, non-stationarity, and non-uniform priors, or when inference stretches over days, weeks or years. Also, we intend to extend our analysis of inference with data omission to environments with ambiguity, where priors are not fixed. This is motivated by Moreno and Rosokha (2016), who show that individuals update their beliefs differently under ambiguity than under risk. Another important issue is how data omission is influenced by the specific valence and intensity of the emotions triggered by each outcome; in our design this is arguably irrelevant, but in other settings the trace a memory leaves could depend on the specific feelings attached to such memory. Further, our framework allows for several straightforward extensions: (i) the signal has more than two outcomes, (ii) two or more correlated signals, so that subjects can measure conditional frequencies, (iii) the signal is non-stationary, etc. Additionally, availability has some facets that are absent in our framework. On one hand, the 'construction' of the 'subjective state space', i.e., the process by which an agent considers some contingencies/states and not others, may depend on the fluency of those signal outcomes associated with a contingency. For instance, there are often various ways in which a politician may lose support, but voters will possibly fail to consider those contingencies that are infrequent. This in turn should affect the voters' estimation that the politician will lose an election (Tversky and Kahneman, 1973, p. 208).[70] On the other hand, Schwarz et al. (1991) has shown that the *ease* with which instances come to mind also affects the evaluation of frequencies; and not only the fluency.[71] Last, but not least, carefully designed field studies should prove that our framework has relevant implications out of the lab –as always, the proof is in the pudding.

[70] A further complication is that subjective representations may depend as well on the task at hand, again by association. For instance, the set of contingencies that an agent considers when discussing politics in an online forum may differ from the one she considers when she is in a voting booth.

[71] Our favorite explanation of this phenomenon is that people tend to associate fluency and ease, i.e., when many instances of an event come to mind, people often feel that they come easily, and vice versa. When people find difficult to recall instances, therefore, they conclude that the rate of occurrence must be low.

References

Abaluck, J., & Adams, A. (2017). What do consumers consider before they choose? Identification from asymmetric demand responses. NBER Working Paper No. 23566.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of financial economics*, *49*(3), 307-343.

Barkan, R., D. Zohar, and I. Erev (1998). Accidents and Decision Making under Uncertainty: A Comparison of Four Models, *Organizational Behavior and Human Decision Processes*, 74(2), 118-144.

Barron, G., and I. Erev (2003). Small Feedback-Based Decisions and Their Limited Correspondence to Description-Based Decisions, *Journal of Behavioral Decision Making*, 16(3), 215-233.

Barron, G., Leider, S., & Stack, J. (2008). The effect of safe experience on a warnings' impact: Sex, drugs, and rock-n-roll, *Organizational Behavior and Human Decision Processes*, 106(2), 125-142.

Barron, G., and G. Ursino (2013). Underweighting Rare Events in Experience Based Decisions: Beyond Sample Error, *Journal of Economic Psychology*, 39, 278-286.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2017). Memory, attention, and choice. NBER Working Paper No. 23256.

Caplin, A., and Dean, M. (2015). "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review*, 105 (7): 2183-2203.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4), 1145-77.

Davis, J. H. (1984). Order in the Courtroom. In D. J. Miller, D. G. Blackman and A. J. Chapman (Eds.), *Perspectives in Psychology and Law*. New York: Wiley.

DellaVigna, S., and Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*, 64, 709-749.

Ebbinghaus, Hermann (1913). *On memory: A contribution to experimental psychology*. New York: Teachers' College.

Erev, I. (1998). Signal Detection by Human Observers: A Cutoff Reinforcement Learning Model of Categorization Decisions under Uncertainty, *Psychological Review*, 105(2), 280-298.

Erev, I. (2007). On the weighting of rare events and the economics of small decisions. In *Developments on Experimental Economics* (pp. 59-73). Springer, Berlin, Heidelberg.

Eysenck, M.W., Eysenck, M.C. (1980). Effects of monetary incentives on rehearsal and on cued recall. *Bulletin of the Psychonomic Society*. 15, 245–247.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2), 171-178.

Fox, C. R., & Hadar, L. (2006). "Decisions from Experience" = Sampling Error + Prospect Theory: Reconsidering Hertwig, Barron, Weber and Erev (2004). *Judgement and Decision Making*, 1(2), 159-161.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4), 25-42.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4), 1661-1710.

Gabaix, X. (2019). Behavioral Inattention, in *Handbook of Behavioral Economics*, Vol. 2, eds. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier.

Gennaioli, N. and A. Shleifer (2010). What Comes to Mind, *The Quarterly Journal of Economics*, 125 (4), pp. 1399-1433.

Hertwig, R., G. Barron, E. U. Weber, and I. Erev (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice, *Psychological Science*, 15, 534–539.

Hilbig, B. E., and A. Glöckner (2011). Yes, they can! Appropriate weighting of small probabilities as a function of information acquisition, *Acta Psychologica*, 138(3), 390-396.

Hogarth, R. M., and H. J. Einhorn (1992). Order Effects in Belief Updating: The Belief-Adjustment Model, *Cognitive psychology*, 24(1), 1-55.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, 47, 263–291.

McFadden, D. (1999). Rationality for Economists? *Journal of Risk and Uncertainty*, 19, 73-105.

Moreno, O., & Rosokha, M. (2016). Learning under compound risk vs. learning under ambiguity – an experiment. *Journal of Risk and Uncertainty, 53*(2), 137-162.

Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*, 117(3), 735-774.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental Psychology*, 64(5), 482.

Oliveira, H., Denti, T., Mihm, M. and Ozbek, K. (2017). Rationally inattentive preferences and hidden information costs. Theoretical Economics, 12: 621-654.

Pennington, N. and Hastie, R. (1986). Evidence evaluation in complex decision making, *Journal of Personality and Social Psychology*, 51(2), 242-258.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117(3), 775-816.

Rabin, M., and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The quarterly journal of economics*, 114(1), 37-82.

Rakow, T. and Miler, K. (2009). Doomed to repeat the successes of the past: History is best forgotten for repeated choices with nonstationary payoffs. *Memory & cognition*, *37*(7), 985-1000

Samuelson, L. (2004). Modeling Knowledge in Economic Analysis, *Journal of Economic Literature*, 42(2), pp. 367-403.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, 12(6), 1423-1452.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: another look at the availability heuristic. *Journal of Personality and Social psychology*, *61*(2), 195-202.

Tversky, A. and D. Kahneman (1973). Availability: A Heuristic for Judging Frequency and Probability, *Cognitive Psychology*, 5, pp. 207-232.

Tversky, A. and D. Kahneman (1974). Judgement under Uncertainty: Heuristics and Biases, *Science*, 185 (4157), pp. 1124-1131.

Weber, E. U., S. Shafir, and A. R. Blais (2004). Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation, *Psychological Review*, 111(2), 430-445.

Yechiam, E., Barron, G., & Erev, I. (2005). The role of personal experience in contributing to different patterns of response to rare terrorist attacks, *Journal of Conflict Resolution*, 49(3), 430-439.

Appendix I: Instructions for treatment 1[72]

Thank you for participating in this experiment. You will be paid some money at its end; the precise amount will depend on chance and your decisions. All decisions are anonymous, that is, the other participants will not get any information about your decisions, nor do you get any information about the decisions of the others.

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. There are no tricky questions, choose simply as you prefer. 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 first phase of the experiment

This experiment has two independent phases or parts. In the first one there is a 'virtual urn' with 100 balls. Each ball is either blue or red. Let us call R the number of red balls in the urn. You do not know either R or the number of blue balls (that is, 100 – R). You only know that R can be equal to 30, 50 or 70 and has been randomly selected by the computer: The probability of having 30, 50 or 70 red balls in the urn is 1/3. To understand this with a merely illustrative example, it is as if R had been determined by

[72] The instructions for treatment 2 were identical except that the set of rates was accordingly adjusted.

rolling a six-sided die, so that if number 1 or 2 had been rolled there would be 30 red balls (R = 30), if number 3 or 4 had been rolled then R = 50, and if number 5 or 6 had been rolled then R = 70. Important: The value of R will not change throughout this phase; the urn is hence always the same one during this phase.

During the phase, 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 each of the extractions, one by one. Once all extractions have been made, you will receive instructions to perform a task that will have a potential associated payoff, to be explained then as well. Posteriorly, you will receive instructions for the second phase.

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

Description of the second phase of the experiment

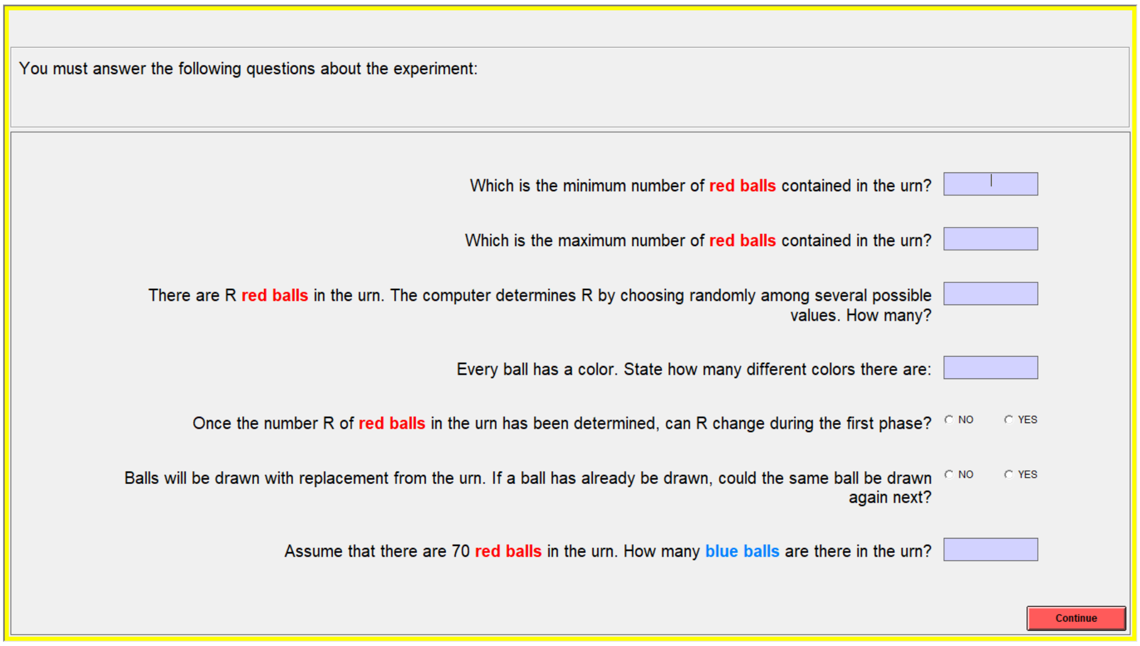
This second phase is very similar to the first one. There is again a virtual urn with 100 balls, blue and red. The number R of red balls is unknown, and has been randomly determined by the computer, choosing one of the options 30, 50 or 70, each one with probability 1/3. Important: Since the value of R is chosen at random, it need not be the same as that in the first phase. In any case, once R is selected, the content of the urn will be always the same during this second phase.

During the phase, the computer will perform several extractions of balls 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 each of the extractions, one by one.

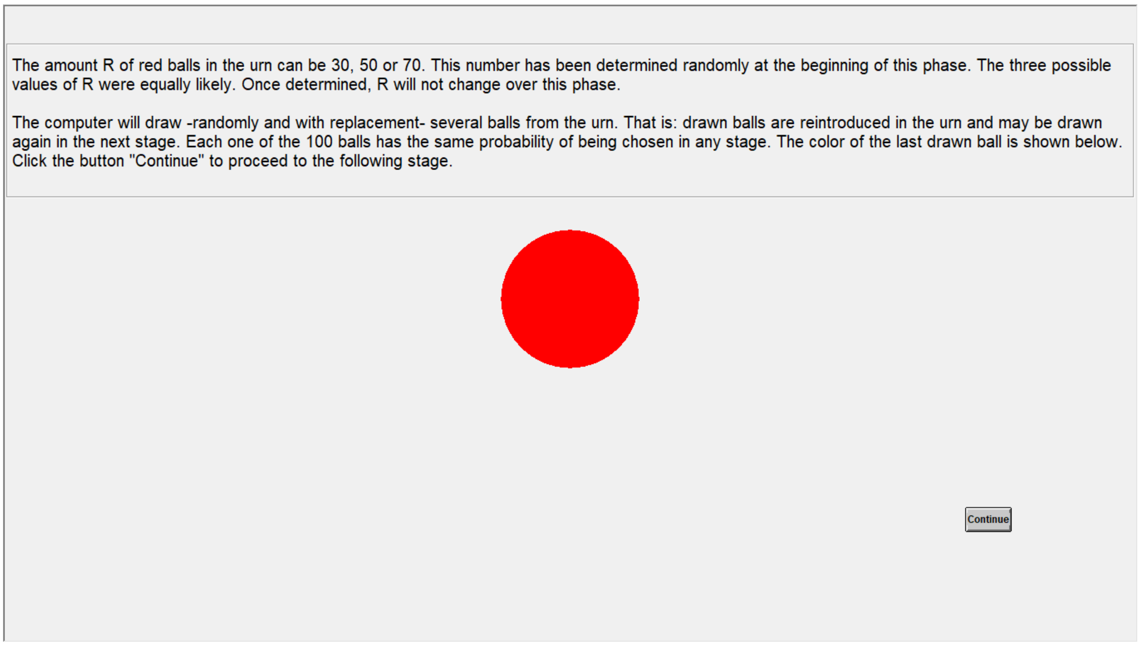
Once all extractions have been made, you will have to estimate the number R of red balls that you believe there are in the urn in this second phase. Analogously as in the first phase, you will get a payoff of 10 Euros if your estimation coincides with the actual value of R during this phase (you get nothing otherwise). Once you have estimated R, you will complete an anonymous questionnaire and afterwards will be paid in private and in cash. In short, your payment will include three components. First, **5** Euros for participating in the experiment. Second, **10** Euros for each estimate that exactly matches the value of R in the corresponding phase. Finally, one item in the questionnaire has an associated prize; the on-screen instructions will explain it later. This prize is completely independent of your estimates of R in the first and second phase.

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

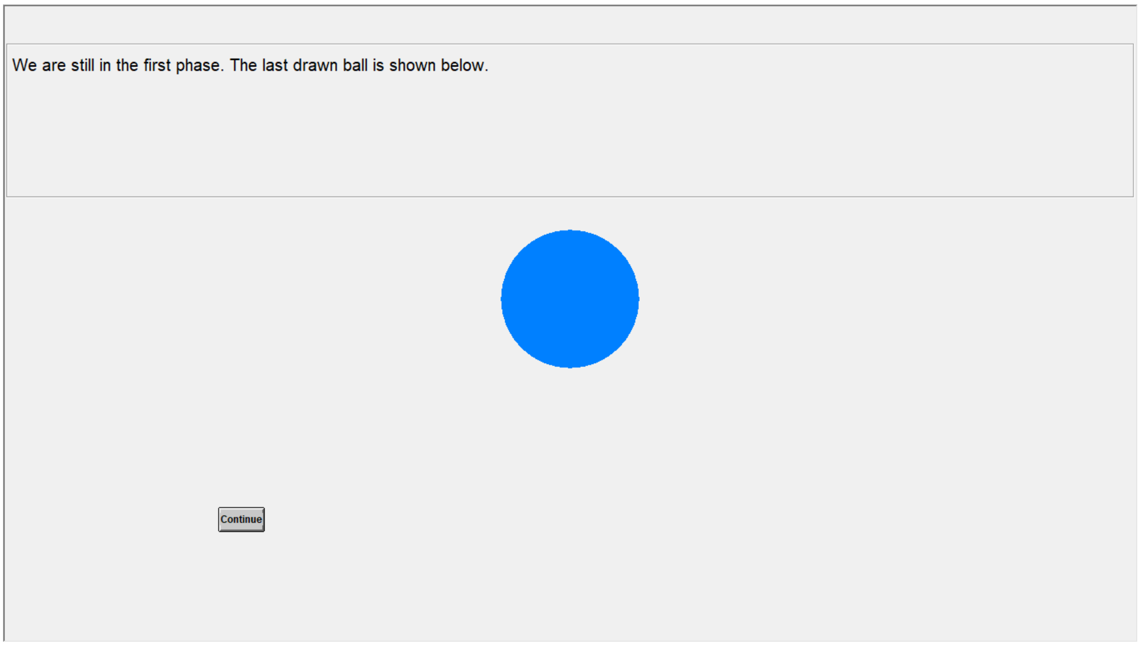
Appendix II: Screenshots



Screenshot 1: control questionnaire



Screenshot 2: Example of a draw (for the two first stages)



Screenshot 3: Example of a draw (for stages 3 to 50)

Appendix III: Additional data

Empirical frequency	Frequency of each estimate			Estimate		Confidence	
	0.3	0.5	0.7	Mean	SD	Mean	SD
0.1-0.2	0	1	0	0.500	0.000	50.00	0.00
0.2-0.3	11	4	0	0.353	0.008	64.53	16.52
0.3-0.4	5	3	1	0.411	0.019	67.33	12.92
0.4-0.5	2	7	3	0.517	0.016	62.17	23.61
0.5-0.6	2	1	4	0.557	0.031	67.14	20.50
0.6-0.7	4	1	7	0.550	0.034	73.75	15.16
0.7-0.8	1	3	2	0.533	0.019	58.00	20.64
0.8-0.9	0	0	1	0.700	0.000	70.00	0.00

Note: No subject observed an empirical frequency lower than 0.1 or larger than 0.9 in this part of T1.

TABLE A. Results of the first phase of the experiment (Treatment 1)

Empirical frequency	Frequency of each estimate			Estimate		Confidence	
	0.3	0.5	0.7	Mean	SD	Mean	SD
0.1-0.2	**2**	**0**	**0**	**0.300**	**0.000**	**70.00**	**20.00**
0.2-0.3	**7**	**1**	**0**	**0.325**	**0.066**	**66.00**	**17.91**
0.3-0.4	**6**	**4**	**0**	**0.380**	**0.098**	**64.50**	**11.93**
0.4-0.5	**2**	**6**	**0**	**0.450**	**0.087**	**73.75**	**18.50**
0.5-0.6	**1**	**10**	**0**	**0.482**	**0.057**	**69.91**	**12.58**
0.6-0.7	**1**	**4**	**8**	**0.608**	**0.127**	**70.77**	**16.51**
0.7-0.8	**1**	**4**	**5**	**0.580**	**0.133**	**65.00**	**20.44**
0.8-0.9	**0**	**0**	**1**	**0.700**	**0.000**	**50.00**	**0.00**

Note: No subject observed an empirical frequency lower than 0.1 or larger than 0.9 in this part of T1.

TABLE B. Results of the second phase of the experiment (Treatment 1)

Empirical frequency	Frequency of each estimate			Estimate		Confidence	
	0.4	0.5	0.6	Mean	SD	Mean	SD
0.2-0.3	1	0	0	0.400	0.000	82.00	0.00
0.3-0.4	7	5	1	0.454	0.004	72.38	15.78
0.4-0.5	9	7	2	0.461	0.005	57.11	21.93
0.5-0.6	5	7	12	0.529	0.006	72.04	15.73
0.6-0.7	2	3	8	0.546	0.006	73.85	11.95
0.7-0.8	1	0	0	0.400	0.000	0.00	0.00

Note: No subject observed an empirical frequency lower than 0.2 or larger than 0.8 in this part of T2.

TABLE C. Results of the first phase of the experiment (Treatment 2)

Empirical frequency	Frequency of each estimate			Estimate		Confidence	
	0.4	0.5	0.6	Mean	SD	Mean	SD
0.2-0.3	1	0	1	0.500	0.100	70.00	30.00
0.3-0.4	10	5	1	0.444	0.061	70.38	19.74
0.4-0.5	8	13	2	0.474	0.061	66.09	16.61
0.5-0.6	2	8	8	0.533	0.067	66.50	19.38
0.6-0.7	1	1	8	0.570	0.064	66.00	31.42
0.7-0.8	0	0	1	0.600	0.000	50.00	0.00

Note: No subject observed an empirical frequency lower than 0.2 or larger than 0.8 in this part of T2.

TABLE D. Results of the second phase of the experiment (Treatment 2)

Treatment	Phase	Bayesian prediction in T1/T2		
		0.3/0.4	**0.5/0.5**	**0.7/0.6**
T1	1	65.2	42.9	52.6
	2	78.9	85	58.3
T2	1	68.4	50.0	58.6
	2	58.3	64.0	66.7
Pooled data		67.1	60.2	59.1

Note: N = 63, 70 in T1 and T2, respectively. We consider three intervals for the empirical frequency, identified by the corresponding Bayesian prediction, as (4) and (5) indicate.

Table E: Proportion of Bayesian estimations, conditional on the empirical frequency observed

Overall conclusions

The aim of this dissertation is to shed light on the process of belief formation, with special focus on the determinants and implications of biased beliefs. In this section, we summarize our main contributions on this matter, and we outline some ideas for future research.

In Chapter 1, we propose a model of motivated beliefs in an economy in which consumers are uncertain about the consequences of their actions. From this model, some valuable insights can be drawn. For instance, we show that people who experience anticipatory feelings are definitely more prone to act in the present to increase their future utility, only when self-deception is not an option. Otherwise, anticipatory utility may turn into self-deception rather than action, particularly when individual actions have little effect on the economy or when these consequences are realized far in the future. This has important implications. Think, for example, of problems like climate change or public health. Should policymakers design campaigns that appeal to people's fear or guilt? Besides ethical considerations, our model suggests that these campaigns may backfire if reality denial is relatively easy to sustain. Interestingly, we show also that self-deception does not necessarily lead to a larger aggregate consumption of risky products, relative to the benchmark model in which belief distortion is not possible.

In Chapters 2 and 3, we conduct an experiment to test the specific predictions derived from the main economic models of optimism, as well as the link between self-serving recall and optimism. In contrast to the field literature on this matter, we propose an experimental design that guarantees a large control both over prior probabilities and the informative signals provided. Also, different from other lab experiments, the informative signals are provided sequentially, distracting tasks are placed between them, and the sample size is relatively large. This design allows a more refined analysis of the link between memory and optimism. Overall, we find scarce evidence supporting the predictions from the main economic models of optimism, although models like Rabin (1994) fit our data relatively better. Further, while we find evidence of biased recall -our participants are more likely to recall 'good news'-, it does not translate into optimistic estimations. We find this result relevant from a methodological perspective, since it suggests that recall and estimation tasks -frequently used in experimental research- might elicit samples with different properties. Hence, the choice of one type of task or another is not neutral, which in turn could partially explain the amalgam of results yielded by the empirical literature so far.

The last chapter is devoted to data omission in environments in which there is no preference over the different states of nature. We contribute to the existing literature with a new quasi-Bayesian framework and an experiment to explore some important issues; namely, people's differences regarding data omission, its cognitive correlates, as well as the role of complexity, incentives and experience. Regarding these questions, we find that, while most of our participants behave in accordance to the Bayesian model, a noteworthy portion of them seem to rely on very small subsamples, omitting almost all the information available. Further, this heterogeneity appears to be due to differences in the level of attention devoted to the problem rather than different memory constraints. Finally, while experience and incentives altogether can alleviate data omission according to our evidence, their effect is modulated by the complexity of the problem.

Beyond our contributions, several considerations must be made about our work and many questions remain open for further research. Our tentative agenda aims to deal with some of them.

Regarding Chapter 1, the considered market structure can be kind of specific. Although alternative structures are explored in the Appendix, we aim to extend our model to more common settings, including oligopolies and monopolistic competition. Also, following Akerlof and Dickens (1982) and Brunnermeier and Parker (2005), our model assumes that consumers simply choose their preferred beliefs. A natural extension should consider the models by, for example, Bénabou (2013) or Bénabou and Tirole (2002). These models put the emphasis on how individuals internalize new information. To incorporate this approach would allow us to explore, for example, the strategic use of information by firms, governments or NGOs in contexts where consumers may interpret this information in a self-serving way.

The results from the experiment referred in Chapters 2 and 3 show scarce support for the economic models of optimism. Yet, it might be argued that the estimation task was too easy and the evidence too apparent as to sustain inflated beliefs. This is in line with the idea of cognitive cost of belief distortion included in models like Rabin (1994). On this matter, our agenda includes the development of an experiment that would consider these costs as a control variable, most likely by varying the complexity or the informativeness of the signals.

Finally, another line of research should focus on the valence and the intensity of the emotions triggered by the informative signals. This variable has not been considered properly in any of the experiments presented in this Thesis. Still, people may pay more attention to signals that trigger more intense -positive or negative- feelings, relative to emotionally neutral evidence (Slovic et al., 2007). This, in turn, may contribute to explain the heterogeneous results found by the experimental literature on optimism.

Conclusiones generales (in Spanish)

El objetivo de esta disertación es arrojar luz sobre el proceso de formación de creencias, con especial énfasis en los determinantes e implicaciones de las creencias sesgadas. En esta última sección, se resumen nuestras principales contribuciones en esta materia, así como algunas ideas para la investigación futura.

En el Capítulo 1, proponemos un modelo de creencias motivadas en una economía en la que los consumidores no están seguros de las consecuencias de sus acciones. A partir de este modelo, pueden extraerse varios resultados valiosos. Así, por ejemplo, mostramos cómo aquellas personas que experimentan sentimientos anticipatorios son definitivamente más propensas a actuar en el presente para incrementar su utilidad futura, solo si el autoengaño no es una opción. En caso contrario, la utilidad anticipada podría llevar al autoengaño, en lugar de a un cambio en las acciones, especialmente cuando las acciones individuales tienen poco impacto en la economía o cuando dicho impacto solo se materializa en un futuro lejano. Esto tiene importantes implicaciones. Podemos pensar, por ejemplo, en problemas como el cambio climático o la salud pública. ¿Sería sensato por parte de los *policymakers* diseñar campañas que apelen a los miedos o a la culpa de la gente? Más allá de consideraciones éticas, nuestro modelo sugiere que estas campañas podrían resultar contraproducentes si los individuos pueden negar fácilmente la realidad. Curiosamente, nuestro modelo también muestra que el autoengaño no necesariamente lleva a un mayor consumo agregado de productos potencialmente dañinos, sino que esto depende de las características específicas de la economía.

En los Capítulos 2 y 3, llevamos a cabo un experimento para comprobar algunas predicciones específicas derivadas de algunos de los principales modelos económicos de optimismo, así como la conexión entre el recuerdo interesado y el optimismo. A diferencia de la literatura de campo que existe sobre esta materia, nuestro diseño experimental garantiza un mayor control tanto de las probabilidades a priori, como de las señales informativas observadas por los individuos. Por otra parte, a diferencia de otros experimentos de laboratorio, las señales informativas son suministradas secuencialmente, incluyendo tareas distractoras, y el número de señales es relativamente grande. Este diseño permite un análisis más refinado de la conexión entre memoria y optimismo. En general, encontramos poca evidencia a favor de las predicciones de los principales modelos de optimismo, aunque algunos modelos como el propuesto por Rabin (1994) se ajusta relativamente mejor a nuestros datos. Más aun, aunque encontramos evidencia de recuerdo sesgado -los participantes de nuestro experimento recuerdan las 'buenas noticias' relativamente mejor-, esto no es traduce en estimaciones optimistas. Este resultado es relevante desde una perspectiva metodológica, en tanto que sugiere que las tareas de estimación y de memoria -frecuentemente utilizadas en la investigación experimental- podrían generar muestras subjetivas con diferentes propiedades. En este sentido, la elección de un tipo de tarea u otro no sería neutral, lo que podría explicar parcialmente la amalgama de resultados que la literatura empírica ha arrojado hasta ahora.

El último capítulo se centra en la omisión de datos en situaciones en las que no hay una preferencia sobre los diferentes estados de la naturaleza. Contribuimos a la literatura existente con un marco teórico cuasi-bayesiano y un diseño experimental con los que explorar algunas cuestiones importantes como, por ejemplo, la heterogeneidad entre individuos con respecto a la omisión de datos, las variables cognitivas relacionadas con la omisión de datos, así como el papel de la complejidad, los incentivos y la experiencia. En relación a estas cuestiones, nuestros resultados sugieren que, si bien la mayoría de los participantes se comporta de manera acorde al modelo bayesiano, una parte no despreciable de ellos parece omitir casi toda la información disponible, utilizando en su lugar submuestras de tamaño muy reducido. Por otra parte, esta heterogeneidad parece deberse a diferencias en el nivel de atención prestado más que a diferencias en la capacidad memorística. Por último, encontramos que la experiencia y los incentivos pueden reducir la omisión de datos, aunque su efecto es modulado por la complejidad del problema.

Más allá de nuestras contribuciones, cabe realizar algunas consideraciones sobre nuestro trabajo, así como sobre las cuestiones que quedan abiertas para una futura investigación. Nuestra agenda provisional busca responder a algunas de estas preguntas.

Con respecto al Capítulo 1, la estructura de mercado considerada puede ser algo específica. Aunque en el Apéndice correspondiente se exploran algunas estructuras alternativas, es nuestra intención extender nuestro modelo a escenarios más comunes, incluyendo el caso de los oligopolios y la competencia monopolística. Por otra parte, y siguiendo a autores como Akerlof y Dickens (1982) y Brunnermeier y Parker (2005), nuestro modelo asume que los consumidores pueden elegir libremente sus creencias preferidas. Una extensión natural de nuestro modelo debería considerar enfoques alternativos, como el propuesto por Bénabou (2013) y Bénabou y Tirole (2002). Estos modelos hacen especial énfasis en cómo los individuos internalizan la nueva información disponible. Incorporar este enfoque nos permitiría explorar, por ejemplo, el uso estratégico de la información por parte de empresas, gobiernos o ONGs en situaciones donde los consumidores podrían interpretar esta información de manera interesada.

Los resultados del experimento referido en los Capítulos 2 y 3 muestran un escaso respaldo de los modelos económicos de optimismo. Con todo, puede argumentarse que la tarea de estimación era demasiado simple y que la evidencia era demasiado contundente como para poder sostener unas creencias sesgadas. Esto iría en la línea de los costes cognitivos asociados a la distorsión de creencias incluida en modelos como el de Rabin

(1994). A este respecto, nuestra agenda investigadora incluye el desarrollo de un experimento que considere estos costes como una variable de control, posiblemente mediante distintos grados de complejidad o fiabilidad de las señales.

Por último, otra línea de investigación estaría enfocada en la valencia e intensidad de las emociones desencadenadas por las señales informativas, variables que no han sido consideradas de manera específica en ninguno de los experimentos presentados en esta Tesis. En este sentido, las personas podrían prestar más atención a aquellas señales que despiertan unas emociones más intensas -ya sean positivas o negativas-, en comparación con aquella evidencia emocionalmente neutral (Slovic y otros, 2007). Consideramos que esto, a su vez, podría contribuir a explicar la heterogeneidad de los resultados hallados por la literatura experimental sobre optimismo.

References

Akerlof, G. A., & Dickens, W. T. (1982). The Economic Consequences of Cognitive Dissonance. *The American Economic Review*, *72*(3), 307–319.

Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets. *Review of Economic Studies*, *80*(2), 429-462.

Bénabou, R., & Tirole, J. (2002). Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, *117*(3), 871–915.

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, *95*(4), 1092–1118.

Rabin, M. (1994). Cognitive dissonance and social change. *Journal of Economic Behavior & Organization*, *23*(2), 177–194.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research*, *177*(3), 1333-1352.