Journal of Economic Behavior and Organization 195 (2022) 140–157

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Some conditions (not) affecting selection neglect: Evidence from the lab ☆

Raúl López-Pérez[a], Ágnes Pintér[b,*], Rocío Sánchez-Mangas[c]

[a] Institute of Public Goods and Policies (IPP), Spanish National Research Council (CSIC), C/ Albasanz, 26–28, Madrid 28037, Spain
[b] Department of Economic Analysis, Universidad Autónoma de Madrid (UAM), Cantoblanco, Madrid 28049, Spain
[c] Department of Quantitative Economics, Universidad Autónoma de Madrid, Cantoblanco, Madrid 28049, Spain

A R T I C L E I N F O

Article history:
Received 23 March 2020
Revised 23 August 2021
Accepted 29 December 2021

JEL code:
D01
D83
D91

Keywords:
Beliefs
Experiments
Extrapolation
Sampling bias
Selection problem

A B S T R A C T

People often extrapolate from data samples, inferring properties of the population like the rate of some event, class, or group –e.g., the percent of female scientists, the crime rate, the chances to suffer some illness. In many circumstances, though, the sample observed is non-random, i.e., is affected by sampling bias. For instance, news media rarely display (intentionally or not) a balanced view of the state of the world, focusing particularly on dramatic and rare events. In this respect, a recent literature in Economics hints that people often fail to account for sample selection in their inferences. We here offer evidence of this phenomenon at an individual level in a tightly controlled lab setting and explore conditions for its occurrence. We conjecture that people tend to update their beliefs *as if* no selection issues existed, unless they have extremely strong evidence about the data-generating process and the inference problem is simple enough. In this vein, we find no evidence for selection neglect in an experimental treatment in which subjects must infer the frequency of some event given a non-random sample, knowing the *exact* selection rule. In two treatments where the selection rule is ambiguous, in contrast, people extrapolate as if sampling were random. Further, they become more and more confident in the accuracy of their guesses as the experiment proceeds, even when the evidence accumulated patently signals a selection issue and hence warrants some caution in the inferences made. This is also true when the instructions give explicit clues about potential sampling issues.

'Life is the art of drawing sufficient conclusions from insufficient premises.' (Samuel Butler)

1. Introduction

Belief formation often works by extrapolation: people observe a sample and infer properties of the population from it. For instance, a person's estimate of the share of women that are intellectually or academically brilliant may largely depend on

☆ We are grateful to Andrés Barge-Gil, Adrián Caballero, Hubert J. Kiss, Juan Palomino, Jaime J. Sánchez-Gabites, Eli Spiegelman, and particularly to the editors and two anonymous referees for helpful comments and suggestions. López-Pérez and Pintér also gratefully acknowledge financial support from the Spanish Ministry of Economy, Industry and Competitiveness through the research project ECO2017–82,449-P. Sánchez-Mangas wishes to thank financial support from the Spanish Government Projects PID2019–108079GBC22 and ECO2015–70,331-C2–1-R and Comunidad de Madrid Project S2015/HUM-3444.

* Corresponding author.
E-mail address: agnes.pinter@uam.es (Á. Pintér).

https://doi.org/10.1016/j.jebo.2021.12.036

the number of female philosophers and scientists that she knows; internet surveys are often used to track public opinion; a voter's estimate of the outcome of some election can be influenced by her friends' stated intention of vote; our belief on the likelihood of a bank run may depend on the share of depositors that we observe withdrawing their savings; and people's evaluation of the state of the world can be based on recently seen news. As any statistician knows, however, a careful application of this inductive, extrapolative way of thinking requires considering the sample properties; in particular, its representativeness. Yet, some psychological evidence suggests that people often fail to take into account selection problems, i.e., the possibility of sampling biases –e.g., Nisbett and Borgida (1975), Hamill et al. (1980), Fiedler (2000). This idea has found support in a recent but growing literature in Economics as well, to be reviewed later. In this paper we use lab experiments to offer additional evidence on selection neglect, particularly at an individual level, and explore some factors that might attenuate it.

Understanding how people infer in the presence of selection issues is important in first place because many of the data sources that we use (even potentially truthful or credible ones) rarely offer a balanced view of the world, as the following examples can illustrate. **1**: If groups of friends are composed of like-minded people, they are relatively unlikely to exchange arguments challenging the group consensus, hence acting as *echo chambers* that do not offer an accurate picture of how diverse society or the world are –Sunstein, 2001; Mullainathan and Shleifer, 2005; Levy and Razin, 2019. **2**: The news media are more likely to focus on dramatic and rare events. **3**: In Internet, conservative political blogs possibly tend to link with other conservative political blogs, whereas liberal blogs link to other liberal blogs. **4**: Twitter users might tend to participate in political conversations that they find consistent with their ideology (e.g., Barberá, 2014). If people neglect potential selection issues in any of these scenarios, the ensuing beliefs may originate questionable decisions: if the media are more likely to report the occurrence of a crime than its non-occurrence, for instance, voters may form an inflated perception of the crime rate, demanding as a result more resources for policing; in fact, 'too many' resources (from the perspective of a well-informed analyst).[1]

Our experiments seek to explore conditions for selection neglect, a still under-researched topic. In each of our treatments, each subject faces an urn with 100 blue and red balls and is informed that the rate θ of red balls is either 10, 40, or 60%. The actual value of θ for each participant is determined by the computer, with a uniform probability of 1/3, and it is not revealed to the subject till the experiment finishes. From this urn, a sample of five balls is randomly drawn with replacement and (**F1**) the subject is shown a subsample; more precisely, (**F2**) all red extractions are shown, whereas any blue one is observed only with 10% probability. After observing the subsample, the subject guesses θ in an incentive compatible manner and moreover, reports her confidence or probabilistic belief that she is accurate. These latter beliefs were non-incentivized, a point that we discuss in Section 2. Overall, the subject faces the same process, i.e., observation of new sample + guess of θ + report confidence, 10 times. We consider three treatments that differ *only* in whether subjects know features F1 and F2 above, i.e., the fact that there is selection and the precise selection rule, respectively. In Transparent, subjects know F1 and F2. In Fuzzy, in contrast, subjects know F1, but not F2, i.e., they know that they observe *some* of the five drawings in each period, without more details. In Opaque, finally, subjects are not informed either about F1 or F2, i.e., they observe the subsample, but the instructions do not clarify that it is part of a possibly larger sample.

In summary, the *exact* selection rule is known in Transparent. Fuzzy and Opaque, in contrast, are characterized by ambiguity about the data-generating process (as many real-life scenarios where selection issues appear, we surmise). In theory, a Bayesian subject in these two treatments should have extremely detailed priors about that process, and hence assign some (maybe nil) probability to, say, an event such as '9 out of 10 red extractions and 2 out of 10 blue extractions are observed'. In practice, however, most subjects are unlikely to exhibit such degree of sophistication: their priors will be possibly over-simplistic and inaccurate. In this line and based on previous experimental evidence from Economics and Psychology, we conjecture that subjects in Fuzzy and Opaque extrapolate *as if* they believed that there are no selection issues, a heuristic here called WYSIATI, i.e., 'What You See Is All There Is'; Kahneman (2011). Formally speaking, the heuristic amounts to a condition on the agents' priors about the selection rule; otherwise, we posit Bayesian learning. Hence, it can be conceived as a refinement of the Bayesian model, which is largely unconstrained in Fuzzy and Opaque. In Transparent, in turn, the heuristic coincides with the standard model. In short, our account does not conceive selection neglect as a phenomenon incompatible with full Bayesian rationality.

While WYSIATI obviously simplifies inference by omitting any selection issues, it leads to systematically mistaken deductions when there are sampling biases, even if individuals are perfectly rational (but at the same time place very high probabilities on samples being random). In particular, the likelihood of the most frequently observed event, e.g., red in our experiments, is exaggerated/inflated. As expected, this is what we find in Fuzzy and Opaque, where subjects guess 'too often' the highest value of θ, i.e. 0.6; more precisely with a frequency higher than 1/3, that is, the actual frequency. Indeed, the overall frequency of guess $\theta = 0.6$ respectively equals 53.1 and 56.6% in Fuzzy and Opaque, with an *increasing* tendency

[1] Similarly, if an anti-corruption agency becomes more efficient, scandals may appear more frequently in the media, thus leading people to the wrong conclusion that corruption is getting worse. In this line, an editorial in *The Economist* the 4th of June 2016 reckons that perceptions of corruption are often more influenced by exposure than actual activity: "It is a common paradox: the world often becomes aware of corruption when someone is doing something about it." Two additional examples of potential effects of selection neglect follow. I: Parents of poor children may under-estimate the returns to schooling if successful youngsters leave the neighborhood, i.e., the 'sample' observed by those parents is biased; Streufert, 2000. II: If investors evaluate a project based on the profitability of prior *implemented* projects and neglect the potential selection effects, they may become over-optimistic and thus over-invest; Jehiel (2018).

over time. Further, we analyze the individual guesses in detail and observe that the heuristic predicts around 60% of them in Fuzzy and Opaque, whereas the remaining guesses can be partly rationalized if we posit that subjects follow WYSIATI with some error. For instance, deviations from the heuristic are most likely (i) among subjects who are least reflective or (ii) in borderline situations, that is, when the frequency observed in the subsample is close to the threshold that, in theory, predicts a change in the guess. Point (i) is maybe noticeable: most reflective subjects act *more* in line with WYSIATI. As a result, they are more likely to have inflated beliefs in Fuzzy and Opaque. In contrast, inflation does not depend on the subject's major or statistical background. In summary, most subjects seem to use WYSIATI, even the statistically sophisticated, but the least reflective ones commit more 'mistakes' in their application.

We also explore three factors or circumstances that might attenuate selection neglect or at least induce a more cautious attitude among the subjects. Our first conjecture is that selection neglect should be greatly reduced if there is extremely strong evidence about the selection rule (and inference is relatively easy).[2] In this line, we find little inflation, if any, in Transparent, where subjects know the *specific* sampling procedure. Indeed, the overall frequency of guess $\theta = 0.6$ equals 31.3%, close to the actual frequency of 1/3. This suggests that subjects, even those lacking much statistical knowledge, understand sampling biases and how to infer when there are (simple) selection issues. A second conjecture is that subjects in Fuzzy and Opaque might become more circumspect about their guesses as the size of the sample increases and patently signals a selection problem. While such circumspection need not induce them to change their guesses, it seems a prerequisite for such change.[3] In all three treatments, however, average confidence increases as the experiment proceeds and subjects become more experienced. This is perhaps striking, as subjects in Fuzzy and Opaque face a paradox: they often observe samples where most (or all) balls are red, but at the same time know that 60% is the largest possible rate of red balls. As the experiment proceeds, indeed, the mean and median probability of obtaining the subjects' observed sample under the assumption of random sampling tends to zero in any treatment. Despite of this, subjects do not become more circumspect, as we have said. Note as well that average confidence is not significantly different across treatments: subjects are equally certain in the ambiguous treatments as in the treatment with full information about the selection rule. A third conjecture, finally, is that selection neglect occurs when subjects fail to realize that there might be sampling biases. This could be the case in Opaque, except maybe in the last periods, where the samples observed include many red balls. In Fuzzy, however, the instructions clearly indicate the existence of a selection issue and, given the unbalanced samples typically observed by the subjects, they should be aware that the sample could be non-random. Hence, subjects should be less confident in Fuzzy than in Opaque. This is not what we find, though. To prevent selection neglect, therefore, it is apparently not enough to warn people that there could be sampling biases.

Overall, a natural interpretation of our findings is that WYSIATI is a default attitude in the absence of precise information about the selection rule. People seem to use such heuristic because it simplifies their inferences, not necessarily because they do not understand selection problems (or fail to notice them). That is, we do not rule out that subjects in Opaque and particularly in Fuzzy consider the possibility that there are sampling biases, potentially of many different shapes. Since they do not know the specific selection rule, however, they find difficult to mentally operate in such an uncertain scenario, and hence simplify matters, i.e., suppress ambiguity, inferring *as if* there were no selection issues. In contrast, when sampling biases can be easily incorporated into their inferences and there is *precise evidence* of them, as in Transparent, subjects do not omit them. At the same time, subjects in Fuzzy and Opaque seem to be unaware of the extent and consequences of their simplifications, as they happen to be as confident in their guesses as fully informed subjects in Transparent.

The rest of the paper is organized as follows. The next section describes the experimental design. Section 3 presents and discusses a theoretical framework that formalizes the heuristic. Section 4 starts with some summary results and then offers an individual analysis of our data, guided by our theoretical framework. Section 5 surveys previous literature connected with our findings. Finally, Section 6 concludes by discussing some potential ideas for future research.

2. Experimental design and procedures

The experiment consists of three treatments. In Transparent, each subject is assigned a 'virtual urn' with 100 balls, blue and red. The rate of red balls θ is either 0.1, 0.4, or 0.6 –i.e., there can be 10, 40 or 60 red balls in the urn. Each subject is uninformed of the actual value of θ in her urn but knows that it is randomly determined: the probability of having 10, 40, or 60 red balls in the urn is 1/3. The computer then extracts five random draws with replacement from the subject's urn,[4] and *some* of these five draws are shown to the subject (no feedback on order of extraction is provided). Disclosure depends on the color of the draw: all red balls extracted are observed, whereas each blue extraction is observed with a probability of 10%. After the disclosed draws (if any) are presented, the subject must provide a *guess* ϵ {0.1, 0.4, 0.6} of the rate θ in her urn. Additionally, she is asked to report her subjective probability (*confidence* hereafter) that her guess is correct, on a scale

[2] In this paper we focus on the issue of how precise the information about the selection rule is, leaving for further research the interaction between complexity and selection neglect. We add the proviso about complexity, though, as we suspect that it is an important one, particularly taking into account the evidence from Enke (2020); we discuss this point further in Sections 5 and 6.

[3] Think also of a scenario in which people can obtain (at a cost) additional information about the selection procedure. If they have little doubts about the accuracy of their guesses, they are unlikely to search for that information.

[4] Sampling with replacement is most usual in the literature on balls-and-urns experiments. See a review in Benjamin (2019). Regardless, the method of sampling used, i.e., with or without replacement, is immaterial for the test of the WYSIATI heuristic.

from 0 to 100, where 0 means that the guess is surely false and 100 that the guess is surely correct. The above-described process consisting in (1) five extractions with replacement, (2) partial revelation, and (3) joint elicitation of the subject's guess and confidence is repeated 10 periods, always with the same rate θ, so that at the end of the experiment a total of 50 balls are extracted from the urn of each participant. Subjects get always feedback about the previously observed extractions. At the end of the session, one of the 10 periods is randomly chosen for payment. Each subject is paid 10 euros if her guess in the chosen period coincides with θ.[5] All previous information is common knowledge in Transparent.

The Fuzzy treatment is identical to Transparent except that subjects are not informed about the disclosure rules –i.e., that all red draws are shown whereas blue ones are shown with probability 0.1. That is, subjects know that some of the five draws are not observed, but do not know exactly which ones. In turn, the Opaque treatment only differs from Fuzzy in that subjects are not informed that some extractions are possibly not shown, i.e., they are only told that they will observe in each period a variable number of extractions (between 0 and 5). In summary, Transparent differs from the other two treatments in the level of ambiguity regarding the data-generation process. In turn, the key difference between Fuzzy and Opaque lies in how obvious a selection issue is.

The experiment was programmed and conducted with the experimental software z-Tree (Fischbacher, 2007). There was one session per treatment, conducted at LINEEX (University of Valencia) in March and April 2017, with 62 participants per session. Subjects were undergraduate students without previous experience in experiments on statistical inference. Each subject participated in just one treatment; participants are not significantly different across treatments in their socio-demographic characteristics. Upon arrival, subjects received written instructions that described the inference problem (see Appendix I). Subjects could read the instructions at their own pace and their questions were answered privately. Understanding of the rules was checked with a control questionnaire that all subjects had to answer correctly before they could start with the decision problem (Appendix I provides examples of some interfaces).

Once their guesses had been elicited in the 10 periods, subjects completed a brief questionnaire on personal and socio-demographic characteristics (gender, age, major, religiosity, and political ideology), risk attitudes,[6] and cognitive abilities – using an expanded cognitive reflection test (henceforth CRT; Frederick, 2005). We also asked them the number of semesters that they had studied statistics or econometrics at the university. After the CRT, moreover, subjects performed a memory task: a list with 15 randomly generated, four-digit numbers was presented on screen and they were given 60 s to memorize as many as possible. In the next screen they had to answer correctly two simple arithmetic problems: (a) $(14 \times 10) - 25 = ?$ and (b) $(5 \times 8) + 39 = ?$ Only when they did so, a new screen allowed them to introduce the recalled numbers, ordered as they wished. They had 60 s for that and were paid 1 Euro for each correct number. After answering a new set of five 'current affairs' questions and questions regarding the use of the media (to be analyzed in a follow-up paper), each subject had one guess randomly selected for payment, and was paid in private. Each session lasted approximately 75 min, and on average subjects earned 14.3 Euros in Control, 13.7 in Fuzzy, and 12.7 in Opaque, including a show-up fee of 6 Euros.

3. Selection neglect & WYSIATI: an analytical framework

Consider a random sample of size k, extracted from a statistical population with two subsets, Blue (B) and Red (R). This sampling defines a sequence of k i.i.d. realizations of a binary signal S, taking on a value of either B or R. The frequency of red individuals in the population equals $\theta \in [0, 1]$, which implies $Prob(S = R) = \theta$. An agent called Adam knows that any rate $\theta_i \in [0, 1]$ in some set Θ is a potential value of θ –in our experiment, for instance, $\Theta = \{0.1, 0.4, 0.6\}$. Yet Adam does not know the actual θ. Further, there is a selection issue in that Adam does not observe the value of some realizations of S (and possibly does not know k).[7] To update her beliefs about θ, therefore, Adam must make some conjectures about k and the non-observed signals. Formally, let b and $r \in \{0, 1, \ldots, k\}$ respectively denote the number of blue and red elements in the sample of size k, so that $b + r = k$. A *state* is a triple $\omega = (\theta_i, b, r)$; let Ω denote the space of all states. Adam's goal is to estimate a posterior probability distribution (the *beliefs*) over Ω. For this, he first assigns a prior $\tau(\omega) \geq 0$ to each $\omega \in \Omega$ and then uses Bayes' rule to update those priors given the known realizations of S –we later illustrate this procedure with a simple example. The priors satisfy condition $\sum_{\Omega} \tau(\omega) = 1$.

In the standard Bayesian model, Adam's priors would be in principle unconstrained, leaving aside the condition just cited and any further restriction derived from the information available to Adam (see footnote 7). To reduce degrees of freedom, however, it is convenient to introduce some *additional* restrictions on Adam's priors. Thus, our theory amounts to

[5] Subjects received no prize for the accuracy of their confidence levels. While incentivizing beliefs about the co-players' choices in VCM games seems to make the estimations more accurate (e.g. Gächter and Renner 2010), when measuring confidence in one's own performance this effect is ambiguous (e.g. Keren 1991). As we discuss later, moreover, our interest is not on the accuracy of the subjects' stated confidence, but on its tendency along the 10 periods (e.g., increasing or decreasing) and how it compares across treatments. In this regard, we had no reason to believe that our results would be different had subjects been incentivized.

[6] 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.

[7] To prevent any misunderstandings, it is important to stress that not observing some signal does not imply not knowing it. In Transparent, for instance, the total number of extractions k is known and all red signals are observed. Therefore, the total number of blue extractions (even non-observed ones) can be logically inferred. We will assume that if the researcher is able to infer the elements in the sample from the information available to Adam, then Adam is able as well.

a ‘refinement’ of the Bayesian model. An example of a potential refinement is the standard assumption that Adam and the researcher have common priors, i.e., they share τ. Our theory, however, departs from this idea. Intuitively, we posit that Adam has ‘wrong’ subjective priors and assigns a nil prior to any state with an unknown probability, as if he omitted or suppressed such ambiguities. To formalize this idea in a simple but testable manner, we distinguish between two types of states. For the first type, let b_o and r_o respectively denote the number of blue and red balls *actually* observed by Adam. A WYSIATI state is any $\omega = (\theta_i, b, r) \in \Omega$ such that $b = b_o$ and $r = r_o$–i.e., Adam observes all realizations and there is no sampling bias: ‘what you see is all there is’. For the second type of state, think of Adam as making decisions within a context or *frame*, to be defined as all sensorial stimuli that he receives –in short, the frame consists of any evident information. A state ω is said to be *non-ambiguous* when some stimulus signals its objective probability, and ambiguous otherwise. In our experiment, more precisely, a state (θ_i, b, r) is non-ambiguous if its probability can be inferred from the instructions.[8]

Our **WYSIATI principle** formalizes the homonymous heuristic presented in the introduction. It states that Adam assigns a nil prior to any non-WYSIATI state ω, unless ω is non-ambiguous. More formally, $\tau(\omega) = 0$ for any ambiguous $\omega = (\theta_i, b, r) \in \Omega$ such that $b > b_o$ and/or $r > r_o$. Intuitively, Adam constructs a parsimonious representation of the world that is as coherent as possible with the frame. Any ambiguous selection rule is excluded from such representation, probably not because he really believes that the data is representative or reliable, but because that assumption simplifies inference and he believes that it is not ‘too’ inaccurate. To facilitate the test of the principle, we finally assume that Adam’s priors over the remaining states are a re-normalization of the researcher’s priors over the set of all states that are either WYSIATI or non-ambiguous.

3.1. Predictions

We first apply the principle to Transparent. Note well that here there is no uncertainty about b and r (see footnote 7 above). Given b and r, moreover, any of the three states $\omega = (\theta_i, b, r)$, for $\theta_i \in \{0.1, 0.4, 0.6\}$, is non-ambiguous: Adam knows that the prior $\tau(\omega)$ of any such state, which can be denoted as $\tau(\theta_i)$, equals 1/3. Since any state that is a priori (objectively) possible is non-ambiguous, we are back in the standard Bayesian model. Given that the signal follows a binomial distribution, further, posterior beliefs for rate θ_i given *history* H = (*k, r*) are:

$$p(\theta_i|\mathrm{H}) = \frac{\tau(\theta_i)\cdot\binom{k}{r}\theta_i^r(1-\theta_i)^{k-r}}{\sum_{\theta_i\in\Theta}\tau(\theta_i)\cdot\binom{k}{r}\theta_i^r(1-\theta_i)^{k-r}} = \frac{\binom{k}{r}\theta_i^r(1-\theta_i)^{k-r}}{\sum_{\theta_i\in\Theta}\binom{k}{r}\theta_i^r(1-\theta_i)^{k-r}}, \tag{1}$$

where the last equality follows from the fact that the priors are uniform. Expression (1) has implications with respect to the optimal guess of θ at any period, which must be the mode of the posterior distribution over $\Theta = \{0.1, 0.4, 0.6\}$, as this estimate maximizes the probability that the 10 euros prize is won if the period is finally selected for payment. Specifically, let $f = r/k$ denote the (actual) empirical frequency of red balls given a sample with a total of k extractions. If we compare expression (1) for the three possible values of rate $\theta_i \in \{0.1, 0.4, 0.6\}$ and do some algebra, we have that the optimal guess is $\theta_i = 0.1$ if $f \le 0.226$, $\theta_i = 0.4$ if $0.226 < f \le 0.5$, and $\theta_i = 0.6$ if $0.5 \le f$. In effect, it suffices to check that

$$p(\theta_i = 0.6|\mathrm{H}) \ge p(\theta_i = 0.4|\mathrm{H}) \leftrightarrow 0.6^r\cdot 0.4^{k-r} \ \ge\ 0.4^r\cdot 0.6^{k-r} \ \Rightarrow f \ge 0.5$$

and

$$p(\theta_i = 0.1|\mathrm{H}) \ge p(\theta_i = 0.4|\mathrm{H}) \leftrightarrow 0.1^r\cdot 0.9^{k-r} \ \ge\ 0.4^r\cdot 0.6^{k-r} \ \Rightarrow f \le \frac{\ln\frac{2}{3}}{\ln\frac{1}{4}+\ln\frac{2}{3}} \ \approx 0.22629.$$

As a rule of thumb, the predicted guess in any period is the rate in Θ closest to frequency f, except when $0.226 < f \le 0.25$, in which case $\theta_i = 0.4$ is optimal. Further, there are two modes when $f = 0.5$.[9] Fig. 1 summarizes the discussion.

We turn now to individual predictions in Fuzzy and Opaque, where the exact sampling procedure is ambiguous. By the WYSIATI principle, Adam infers as if $b = b_o$ and $r = r_o$ was certain.[10] Although $r = r_o$ happens to be true in our design,

[8] A subtler, but possibly less falsifiable, idea is that a state ω is non-ambiguous when (a) Adam believes that his prior $\tau(\omega)$ is “well-grounded”, i.e., backed by “abundant” experience, and (b) $\tau(\omega)$ is cued by the frame. On point (b), suppose that Adam’s memory has associated, through repeated co-observation, some sensorial input I with the value of $\tau(\omega)$. The observation of I could then bring such value to his mind. For instance, if an advertisement for an investment company mentions the performance of two successful funds but Adam learns that the company also operates 28 additional funds, he is likely to conclude (particularly is he is an expert) that the company has selected for such ad its best funds, not the mediocre or bad ones –Koehler and Mercer (2009). A different limitation of our theory is that it assumes that people can incorporate sampling issues when the uncertainty in this respect can be precisely quantified, i.e., in non-ambiguous settings. We suspect though that people will also resort to WYSIATI even if there is knowledge of the relevant objective probabilities, provided that the inference problem is complex ‘enough’; see our posterior discussion of Enke (2020). Regardless, we leave these subtleties for further research.

[9] Properly speaking, there are also two modes in the hypothetical case that $f = \frac{\ln\frac{2}{3}}{\ln\frac{1}{4}+\ln\frac{2}{3}}$. We omit this possibility because the probability that the frequency of red takes exactly such value in a subject’s sample is basically nil. Note also that the threshold of 0.226 that we consider in Fig. 1 is slightly lower than this theoretical one. If $f = 0.226$, therefore, the subject should unequivocally guess a rate of 0.1.

[10] Although they may not observe the whole sample, subjects in Fuzzy know that five balls are drawn in every period. It follows that the WYSIATI principle cannot be taken literally there; it is just a very simple way to formalize the idea that Adam neglects sampling issues. One can model this idea in

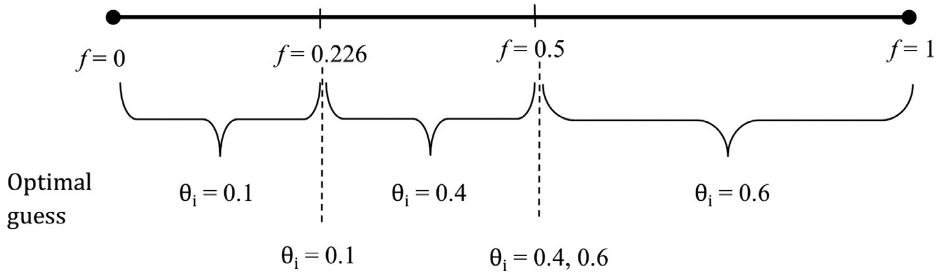

Fig. 1. Optimal guesses conditional on empirical frequency of red.

note that $b_o \leq b$, as some blue balls might not be selected out of the sample. Given the *observed history* $\mathrm{H_o} = (b_o, r_o)$, therefore, Adam will compute his posterior beliefs as follows (as before, the last equality follows from the fact that priors are uniform):

$$\mathrm{p}(\theta_\mathrm{i}|\mathrm{H_o}) = \frac{\tau(\theta_\mathrm{i}) \cdot \binom{b_o + r_o}{r_o} \theta_i^{r_o} (1-\theta_\mathrm{i})^{b_o}}{\sum_{\theta_\mathrm{i} \in \Theta} \tau(\theta_\mathrm{i}) \cdot \binom{b_o + r_o}{r_o} \theta_i^{r_o} (1-\theta_\mathrm{i})^{b_o}} = \frac{\binom{b_o + r_o}{r_o} \theta_i^{r_o} (1-\theta_\mathrm{i})^{b_o}}{\sum_{\theta_\mathrm{i} \in \Theta} \binom{b_o + r_o}{r_o} \theta_i^{r_o} (1-\theta_\mathrm{i})^{b_o}}$$

Again, her optimal guess coincides with the mode of this posterior distribution. Let $f_o = r_o/(b_o + r_o)$ denote the *observed* empirical frequency of red balls given history $\mathrm{H_o}$, that is, the percentage of red in the observed subsample. Following an analogous argument as in Transparent, the mode of the posterior distribution is $\theta_\mathrm{i} = 0.1$ if $f_o \leq 0.226$, $\theta_\mathrm{i} = 0.4$ if $0.226 < f_o \leq 0.5$, and $\theta_\mathrm{i} = 0.6$ if $0.5 \leq f_o$. Further, there are also two modes ($\theta_\mathrm{i} = 0.4, 0.6$) when $f_o = 0.5$, and we add a case that did not appear before: when a subject has observed no ball in period t and the precedent ones, the original, uniform priors are unchanged and hence any rate is an optimal guess ($\theta_\mathrm{i} = 0.1, 0.4, 0.6$).[11] Leaving aside the few exceptions noted, in any case, the rule of thumb is that the rate chosen should be closest to f_o.

4. Data analysis

This section is organized as follows. In 4.1, we report some aggregated data on subjects' guesses and degree of confidence. Consistent with our theory, the data hints that subjects do not consider the possibility of sampling biases when there is ambiguity in this respect, inflating their guesses as a result. Then we check the empirical relevance of the WYSIATI heuristic/principle at an individual level in Section 4.2. In Section 4.3 we use a regression analysis to provide more details on the factors that affect the subjects' guesses and level of confidence, as well as on the probability of deviation from what the principle predicts. In 4.4 we analyze how WYSIATI performs relative to some alternative heuristics that subjects might use in Fuzzy and Opaque and explore the performance of an alternative Bayesian model as well.

4.1. Summary of results

Before we present our findings, it can be worth to summarize our predictions. Our three treatments correspond to three different situations. In Transparent, the exact selection rule is known and subjects are expected to act as standard Bayesians. Specifically, their guesses should track the actual empirical frequency f of red balls. In Fuzzy and Opaque, subjects do not know the rule, and we conjecture them to infer as if there was no selection issue, even when it should be obvious in Fuzzy that the sample might not be representative. Consequentially, as we proved before, subjects in Fuzzy and Opaque use the frequency of observation of the red balls f_o to guess rate θ. Since most of these subjects observe a high frequency of red, their predictions become 'inflated' with respect to those in Transparent. In summary, we have the following hypothesis.

Hypothesis 1. (H1): *In comparison with Transparent, subjects in Fuzzy and Opaque tend to guess more frequently the highest rate,* i.e.*, 0.6. Hence, the average elicited guess is biased upwards in these two treatments. Further, the distribution of guesses in these treatments are not different.*

Fig. 2 below plots the distribution of guesses in the three treatments ($N = 620$ in each treatment; Tables A–C in Appendix II offer more disaggregated data). Consistent with H1, the distribution is similar in Fuzzy and Opaque but quite different in Transparent. Indeed, considering individual-averaged guesses, a Mann-Whitney test of the equality of distributions in

a more elegant manner, although at a significant expositional cost. We abstain from this, not only because of the higher cost, but also because we suspect that the WYSIATI principle reflects rather well the kind of reasoning that subjects tend to use in Fuzzy.

[11] This case did not appear in Transparent because, when the subjects observe no balls in a period there, they are assumed to know that the five extracted balls of that period are actually blue.

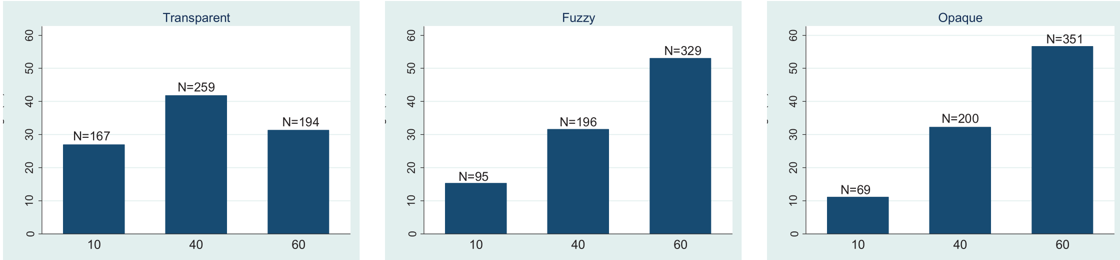

Fig. 2. Distribution of guesses of the number of red balls in each treatment.

Table 1
Average guess across treatments and periods.

Period	Treatment			TOTAL
	Transparent	Fuzzy	Opaque	
1	42.1	40.6	45.6	42.8
2	39.4	41.9	44.4	41.9
3	35.5	45.0	47.3	42.6
4	39.7	43.9	52.9	45.5
5	36.9	45.6	47.7	43.4
6	37.6	48.9	48.1	44.8
7	39.2	46.0	48.5	44.6
8	37.9	47.9	47.9	44.6
9	36.9	49.4	47.6	44.6
10	36.6	51.0	49.8	45.8
TOTAL	38.2	46.0	48.0	44.1

Fuzzy and Opaque does not reject the null hypothesis (p-value $= 0.438$). In contrast, the p-values for the comparison of the distributions of guesses in Transparent vs. Fuzzy and in Transparent vs. Opaque are both lower than 0.001. If in addition we compare the frequency of choosing 0.6 under the three treatments, we find that the null hypothesis of equal frequencies is not rejected when we compare Fuzzy and Opaque (p-value $= 0.514$), but it is rejected at any significance level if we compare Transparent and Fuzzy or Transparent and Opaque (p-value < 0.001 in both cases). Note also that the frequency of choice of the highest rate, i.e., 0.6, is significantly larger than 1/3 in Fuzzy and Opaque, but not in Transparent, evidence again that beliefs are 'inflated' in the former treatments.[12]

The 'inflation' of the beliefs is not only persistent; in fact, it becomes more acute in the latter periods. As our regression analysis confirms in Section 4.3, in effect, the rate of choice of 0.6 in treatments Fuzzy and Opaque increases as the experiment proceeds,[13] a phenomenon that is not observed in Transparent. Because of these dynamics, there is an *increasing* difference between the average guesses in Fuzzy and Opaque and those in Transparent; see Table 1 below, which indicates the average guess in any period and treatment. In fact, the average guess in Transparent tends to decrease, and coincidentally reaches in the last period a value that coincides with the expected number of red balls in the urn, that is, $(10+40+60)/3 \approx 36.6$. Since inflation makes less likely that a guess coincides with the actual target, note also that the share of subjects who get the estimation prize of 10 Euros is relatively larger in Transparent. Indeed, this share respectively equals 54.8% in Transparent, 46.8% in Fuzzy and 38.7% in Opaque. Thus, the inflated guesses in Fuzzy and Opaque are not innocuous but lead to worse outcomes.

Result 1. (inflation): *Subjects report inflated guesses in Fuzzy and Opaque and inflation increases as the experiment proceeds. The distributions of guesses in these two treatments are not significantly different. There is no such inflation in Transparent.*

To further analyze the dynamics in our treatments, we calculated for each subject the distance between the correct guess, i.e., the actual value of θ in her urn, and her actual guess in each treatment and period.[14] Fig. 3 plots the average distance over time in the three treatments. Since the distance is defined as the number of red balls in the urn minus the subject's actual guess, note that it can take negative values (in fact, the average distance almost always does so, in any treatment).

[12] In a one-sided test where the null hypothesis states that the frequency of the guess 0.6 equals 1/3, whereas the alternative assumes it to be larger than 1/3, the p-value is 0.716 in Transparent, but it is lower than 0.001 in Fuzzy and Opaque. All these tests have been performed with individual-averaged data.

[13] In this respect, we speculate that the inflation in Fuzzy and Opaque would have grown bigger if subjects had seen, say, 60 extractions instead of 50.

[14] We thank an anonymous referee for this idea.

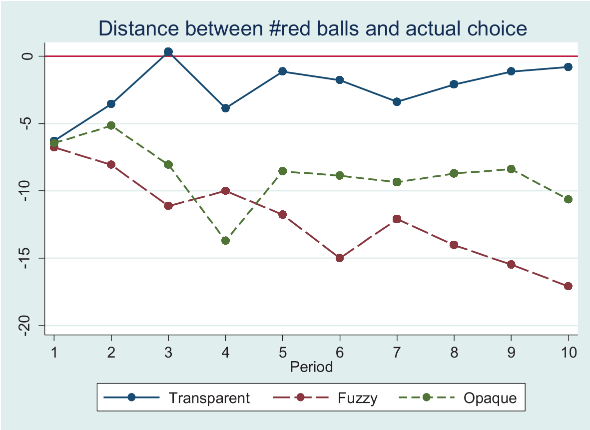

Fig. 3. Average distance between the correct and actual guesses, per treatment and period.

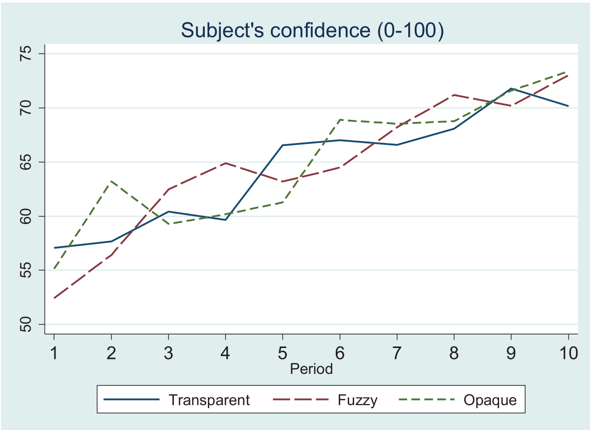

Fig. 4. Average confidence per treatment and period.

The average distance is relatively small in Transparent, but significantly bigger in Opaque and Fuzzy.[15] In Fuzzy and Opaque, further, it is much more likely that subjects guess 60, which explains why the average distance is typically far below zero in these treatments. Moreover, as the experiment evolves, the observed frequency makes subjects more and more likely to choose 60, thus committing a larger average error. Hence, the distance in these two treatments is growing over period, contrary to what we observe in Transparent.

We turn now our attention to the level of confidence that subjects report after each of the 10 guesses, that is, the probability with which they believe their guess is correct. Fig. 4 below depicts the average level of confidence in each period of each treatment (see Table D in Appendix II for the precise figures). We make three remarks. The first two are closely related to the WYSIATI principle: apparently, subjects in Fuzzy and Opaque infer *as if* they assigned nil (or very low probability) to the possibility that the sample is biased, i.e., as if they assumed that the data is reliable. Specifically, *first*, we observe very minor differences across treatments; something to be later confirmed by our regression analysis in Section 4.3. This is consistent with the principle: if subjects in Fuzzy and Opaque strongly suspected that the data was unreliable, their average levels of confidence should be significantly lower than the average level in Transparent, where subjects have all the relevant information. Clearly, this is not what Fig. 4 shows.

For the *second* remark, consider (contrary to WYSIATI) a subject in Fuzzy or Opaque who believes a priori that sampling biases could occur with some probability; more precisely, biases favoring the observation of red balls. If she observes period after period that most extractions are red, one might expect her priors that sampling is non-random to be reinforced (see Section 4.4 for some Bayesian qualifications to this statement, though). As a consequence, the subject might state decreasing levels of confidence –particularly in the last periods, where the sample size is larger. This should not take place in Trans-

[15] We have tested in each period whether the average distance equals zero. At 5% significance level, there is no evidence to reject the null hypothesis in 9 out 10 periods in Transparent (the *p*-value is 0.02 in the first period but then ranges between 0.2 and 0.7 in periods 2 to 10). In contrast, in Fuzzy the hypothesis is rejected in all periods and in Opaque in 9 out of 10 periods, with p-values below 0.01 in most of them.

Table 2
Frequency of choices consistent with the WYSIATI principle.

	Treatment			
Period	**Transparent**	**Fuzzy**	**Opaque**	**TOTAL**
1	58.1	48.4	66.1	57.5
2	59.7	51.6	46.8	52.7
3	58.1	53.2	50.0	53.8
4	50.0	54.8	74.2	59.7
5	64.5	58.1	56.5	59.7
6	51.6	61.3	61.3	58.1
7	53.2	59.7	66.1	59.7
8	51.6	62.9	67.7	60.8
9	62.9	69.4	62.9	65.1
10	59.7	69.4	67.7	65.6
TOTAL	57.0	58.9	62.0	59.2

parent. Thus, we should observe an increasing difference between Transparent and the other treatments as the experiment progresses. Yet Fig. 4 visibly indicates this not to be the case. In other words: in a scenario characterized by an ambiguous selection problem, people do not become more circumspect about their inferences when the evidence strongly suggests such a problem.

In any treatment, *third*, average confidence increases as the experiment progresses; this is a significant result, as our analysis in Section 4.3 confirms. This increase in confidence has at least two possible interpretations: (i) As the sample size grows, subjects gain confidence, and (ii) as subjects play more periods, they understand better the statistical problem and hence consider their guesses more accurate. Yet the evidence seems relatively more consistent with account (ii). In effect, recall that subjects 'observe' all extractions in Transparent. This means that the sample size tends to be substantially larger there, particularly in the last periods. Consequently, subjects in Transparent should be more confident, especially in the last period. As we said before, this is not what we see in Fig. 4. We explore in more detail this question in Section 4.3 but note that the relative insensitivity to sample sizes that subjects seem to exhibit in their inferences, is something in line with abundant prior evidence; e.g., Tversky and Kahneman (1971).

Result 2. (confidence): *The average subject is similarly confident in the accuracy of her guess at any given period in all treatments,* i.e., *irrespectively of the information she has about the reliability of the data. Average confidence increases as the experiment progresses.*

4.2. Exploring individual guesses: A test of the WYSIATI principle

In Section 3, we applied the model to each of our treatments and derived a series of predictions. They can be used to explore the fit of the WYSIATI principle at an individual level. Hypothesis 2 summarizes our predictions for the different treatments:

Hypothesis 2. (H2): *In Transparent, guesses of rate θ depend on the actual frequency of red extractions, while in Fuzzy and Opaque they track the observed frequency of red balls.*

4.2.1. Evidence

Table 2 below indicates the percentage of guesses consistent with H2 in each period of each treatment.[16] As we can see, H2 overall predicts 57, 58.9, and 62% of the guesses in Transparent, Fuzzy, and Opaque, respectively. We note as well that the fit of the model does not improve as the experiment proceeds in Transparent, but there is some positive trend in Fuzzy and Opaque; the regression analysis below provides more details on this.

Result 3. *In Transparent, 57% of the guesses are predicted by the heuristic, with no significant improvement along time. In Fuzzy and Opaque, around 60% of the subjects guess in accordance with the heuristic, with a positive trend over time.*

Note that the principle is equally successful across treatments: leaving aside some small differences, around 60% of the subjects follow the treatment-relevant frequency (see H2) in the three treatments. This is somehow reassuring, as we do not predict a different rate of success in any treatment. Since the explanatory power of the principle is not extremely high, however, one might wonder whether the choices left unexplained follow a very different logic than the one suggested by the principle.

[16] For more detailed data, Table A in Appendix II displays the number of subjects choosing each rate θ_i in Transparent, for any history –determined by period and empirical frequency of red, presented in intervals [0, 0.226], (0.226, 0.5], (0.5, 1]. In addition, Tables B and C in Appendix II respectively display the number of subjects in Fuzzy and Opaque choosing each rate θ_i, conditional on the observed history.

Table 3
Frequency of guesses consistent with the two-types model.

	Frequency (%)			
Period	**Transparent**	**Fuzzy**	**Opaque**	**TOTAL**
1	58.1	48.4	66.1	57.5
2	62.9	58.1	59.7	60.2
3	66.1	64.5	61.3	64.0
4	59.7	64.5	74.2	66.1
5	72.6	75.8	66.1	71.5
6	56.5	77.4	72.6	68.8
7	69.4	75.8	69.4	71.5
8	66.1	72.6	75.8	71.5
9	64.5	75.8	71.0	70.4
10	72.6	75.8	74.2	74.2
TOTAL	64.8	68.9	69.0	67.6

For some reasons, we think that this is not the case: many deviations from the principle can be rationalized as simple mistakes. To start, we will show in Section 4.3 that deviations are more likely when the inference problem is arguably more difficult, that is, when the relevant frequency is close to the thresholds indicated in Fig. 1. Further, people scoring lower on the CRT, who possibly pay less attention to the information in the screens, instructions, etc., are significantly more likely to deviate.[17] Finally, we have evidence that some subjects apply the logic of the principle in a (cognitively) 'myopic' manner. In other words, their guesses in period t are based only on the treatment-relevant frequency *in that period*. This group contrasts with the possibly more attentive, 'non-myopic' subjects, who consider the frequency across t and all prior periods (if any). In short, myopic subjects follow the WYSIATI principle imperfectly, focusing on recent evidence, whereas non-myopic ones estimate in absolute accordance with our predictions. This extension of our theory is called henceforth the 'two-types' model.

To explore the empirical accuracy of this extension, we first conduct a classification analysis of the subjects in each treatment. We compute, for each subject, the number of guesses consistent with the WYSIATI principle according to the treatment-relevant frequency in (i) the most recent period and (ii) across all periods. A subject is classified as 'myopic' if the number of such guesses is higher in case (i) than in (ii). The percentage of myopic subjects is 38.8, 38.7 and 35.5% in Transparent, Fuzzy and Opaque, respectively. Using this classification, we then compute the share of guesses that are explained by the two-types model in any treatment and period. This data can be found in Table 3, which is to be contrasted with Table 2 above, showing the empirical relevance of our 'single-type' theory presented in Section 3.

In percentage points, the two-types model explains 7.8, 10, and 7 p.p. more than the single-type model in Transparent, Fuzzy, and Opaque, respectively. Overall, the improvement in all treatments equals 8.4 percentage points. Although these figures are not very large, they provide further evidence that a significant part of the guesses unexplained by the WYSIATI principle can be rationalized by introducing small adjustments, in particular, allowing for mistakes. Note also that the two-types model predicts that myopic subjects switch guesses more frequently: these types operate with a smaller sample, which is obviously more likely to present substantial variations in the frequency of red balls from period to period. This prediction is confirmed by the data, as myopic subjects switch guesses in average in 5.1 periods (out of 10), whereas non-myopic subjects switch in 3.1 periods. This phenomenon is observed in any treatment: while myopic subjects switch in average in 5.1, 4.8, and 5.5 periods in Transparent, Fuzzy, and Control, the corresponding figures for the non-myopic subjects are 3.7, 3.1, and 2.6, respectively.

We have also investigated the possibility that some subjects in Transparent behave as those in Fuzzy or Opaque, i.e., without considering the non-observed blue extractions, so that the condition stated in Footnote 7 does not hold. For each subject and period, more precisely, we computed whether the subject's guess is more consistent with the predictions from the theory taking the actual (Transparent) or the observed (Fuzzy and Opaque) empirical frequency of red balls. We found that 16.3% of the subjects' guesses in Transparent are better explained by assuming that they track the observed, instead of the actual, frequency.

Result 4. (deviations): *A significant share of the deviations from the WYSIATI principle seems to be caused by subjects' mistakes, including considering only the current period sample (not the whole one).*

4.3. Further tests of the WYSIATI principle: Regression analysis

In this section, we perform some regression analyses to explain the subjects' guesses, the probability of deviation from the WYSIATI principle, and the subjects' degree of confidence in terms of characteristics of the experimental design, as well

[17] Psychologists often contend that the CRT score is a good measure of how willing someone is to meditate on a problem (Kahneman, 2011). In each CRT problem, there is one specious answer, salient and therefore tempting, but wrong. For instance, in the question "A bat and a ball together cost $1.10, and the bat costs $1.00 more than the ball: how much does the ball cost?" the trap answer is $0.10. Participants must therefore think carefully enough to avoid the trap.

Table 4
Determinants of subjects' guesses; estimation results.

TREATMENT	Transparent	Fuzzy	Opaque
Period (1–10)	−0.004 (0.004)	0.029 (0.007)***	0.013 (0.007)**
Treatment-relevant frequency (current period)			
Overall effect	0.679 (0.092)***	0.079 (0.053)	0.095 (0.066)
For myopic subjects	1.131 (0.114)***	0.239 (0.081)***	0.339 (0.108)***
For non-myopic subjects	0.374 (0.135)***	−0.020 (0.074)	−0.031 (0.086)
Treatment-relevant frequency (all prior periods)			
Overall effect	0.141 (0.141)	0.119 (0.102)	0.249 (0.090)**
For myopic subjects	−0.400 (0.155)***	−0.118 (0.124)	−0.081 (0.136)
For non-myopic subjects	0.533 (0.202)***	0.278 (0.118)**	0.429 (0.094)***
Observations	620	620	620
Pseudo R^2	0.218	0.120	0.085

Note: Standard errors clustered at the individual level. Average partial effects on Pr(choice=60|X) reported; standard errors of the average partial effects computed through the Delta Method. ***, **, *: significant at 1%, 5% and 10% significance level, respectively.

Table 5
Determinants of a deviation from the WYSIATI principle; estimation results.

TREATMENT		Transparent	Fuzzy	Opaque
Period (1–10)		0.001 (0.007)	−0.030 (0.007)***	−0.018 (0.006)***
Frequency interval	Easy (rest of the unit interval)	0.349 (0.067)***	0.344 (0.046)***	0.274 (0.049)***
	D1: 0.226±0.05	0.556 (0.095)***	–-	–-
	D2: 0.5 ± 0.05, excluding 0.5	0.459 (0.099)***	–-	–-
Myopic subject		0.199 (0.055)***	0.215 (0.062)***	0.262 (0.065)***
Number of correct answers in CRT		−0.108 (0.024)***	−0.101 (0.030)***	−0.040 (0.031)
Observations		620	619	617
Pseudo R^2		0.135	0.165	0.101

Note: Standard errors clustered at the individual level. Average partial effects reported; standard errors of the average partial effects computed through the Delta Method. In Fuzzy and Opaque, some observations in D1 are lost due to lack of variability of the dependent variable; further, there are no observations in D2. The reference interval is the 'very easy' one, that is, freq = 0.5 and 'no balls seen'. ***, **, *: significant at 1%, 5% and 10% significance level, respectively.

as individual characteristics. There are variables that are common in all models, such as the dynamics of the experiment, and specific variables that are related to the treatment-relevant frequency of red balls (i.e., actual empirical frequency in Transparent and observed frequency in Fuzzy and Opaque). In Table 4 the dependent variable is the subject's guess about the number of red balls, with values {10, 40, 60}. We use an ordered probit model to estimate the probability of choosing each of these alternatives. We have included, as explanatory variables, the period, the treatment-relevant frequency of red balls in the current period and in all prior periods, the interaction of these frequencies with the variable that indicates the subject's type (myopic vs. non-myopic) and some controls related to individual characteristics. To simplify, we only report the average partial effects of each variable on the probability of a guess of 60.[18]

Our results confirm prior findings. Regarding the evolution of the experiment, captured by the variable 'period', we find no significant effect in Transparent, but a positive and significant effect in Fuzzy and Opaque, in line with Result 1 above. In these two treatments, it is more likely a guess of 60 red balls as the experiment evolves. Confirming the results from our two-types model analysis, in addition, the likelihood of a guess of 60 co-moves in all treatments with the (treatment-relevant) frequency of red balls in the current period for the myopic subjects, and with the overall frequency for the non-myopic subjects.

Result 5. (determinants of subjects' guesses): *Other things equal, the probability of choice of the maximal guess,* i.e.*, 60, increases as the experiment evolves in Fuzzy and Opaque, but not in Transparent. In line with the WYSIATI principle, non-myopic subjects' guesses increase with the actual empirical frequency in Transparent, and with the observed one in Fuzzy and Opaque. For myopic subjects we find analogous results for the treatment-relevant frequencies of the current period.*

In Table 5 we estimate probit models on the probability of making a guess that deviates from the WYSIATI principle, in each treatment. Again, we report the average partial effects of each variable on this probability. Consistent with Result 3 above, the dynamics of the experiment has no effect in Transparent. Yet it has a significantly negative one in Fuzzy and Opaque, i.e., the period is associated with a lower probability of deviation, with quite similar quantitative effects in both treatments.

[18] Since we estimate an ordered probit model, the sign of the partial effects of each variable on the probability of a guess of 10 is, by construction, the opposite of the sign of the partial effect on the probability of a guess of 60. Results on the average partial effects on the probability of every outcome, as well as the estimated coefficients from the ordered probit model (instead of average partial effects) are available from the authors upon request.

On a different topic, deviations from the principle might be more likely in borderline, difficult cases, i.e., when the frequency is close to the thresholds in Fig. 1. To check this, we consider different intervals for the relevant frequency in each treatment (when all the history of extractions is considered). The reference is a 'very easy' case that includes frequency 0.5 in the three treatments (with this frequency there are two optimal guesses, 40 and 60 red balls, and hence, the probability of deviation lowers) and those cases in which no balls are observed in Fuzzy and Opaque (in these cases, any guess is optimal and therefore, there is no deviation). The interval D1 includes 0.05 points above and below the threshold 0.226 for which the optimal guess goes from 10 to 40 red balls. Analogously, the interval D2 includes 0.05 points above and below the threshold 0.5 (without including 0.5), for which the optimal guess goes from 40 to 60 red balls. These two intervals D1 and D2 around the threshold values can be arguably considered as difficult ones. The rest of the unit interval is considered as an 'easy' case, intermediate between the 'very easy' one and the frequencies around the thresholds (without 0.5, that is included in the 'very easy' case). Focusing on the Transparent treatment, it is clear that the probability of deviation is higher when the frequency is around the thresholds, even in the easy interval. When the frequency is in this interval (with respect to the reference category) the probability of deviation increases around 35 p.p. The effect is much higher when the frequency is around the thresholds, 56 p.p. in D1 and 46 p.p. in D2.[19] In Fuzzy and Opaque there is not enough variability to estimate all these effects, more specifically those of D1 and D2. However, we can compare what happens in the 'very easy' and the 'easy' case. We find that the probability of deviation increases around 30 p.p. in the latter (compared to the former one).

We also report how the subject's type is associated with the probability of deviation from WYSIATI. As expected by construction, for myopic subjects this probability is higher (around 20 p.p.) than for the non-myopic ones, with some slight differences across treatments. Concerning the CRT, a good performance is associated with a lower probability of deviation. This is found in all three treatments, with a slightly higher effect in Transparent and Fuzzy; the effect in Opaque is non-significant at the usual levels, but the p-value is not too high, 0.198.

Result 6. (determinants of a deviation): *Deviations from the principle are more likely (i) among the least reflective subjects, (ii) among the myopic subjects, (iii) and when the inference problem is more complex. This suggests again that many of the deviations are the result of applying the principle with some error, not the result of a substantially different logic of choice. The probability of deviation in Fuzzy and Opaque (but not in Transparent) is significantly reduced as more and more periods are played.*

On a different issue, we have used linear models to explore the factors that might influence the subjects' confidence in their guesses in each treatment. For brevity, here we summarize our main findings –the precise estimation results are shown in Table E in Appendix II. First, confirming Result 2 above and now controlling for other factors, as more periods are played, the subjects' confidence increases. The quantitative effect is around 2 points per each additional period, a bit lower for Fuzzy (this is the only treatment where the effect is not significant at standard levels, as the p-value equals 0.109). Being myopic seems not to be related with the subject's confidence in Transparent, second, but has a negative effect in Fuzzy and Opaque, ranging from 10 to 5 p.p. although is only significant in Fuzzy. Third, the performance in the CRT is associated with more confidence only in Transparent, but we find no effect in Fuzzy and Opaque. Fourth, it seems that the overall number of actually observed balls does not play a role in Fuzzy or Opaque: a larger 'observed' sample does not increase confidence. We do not check for this 'sample size' effect in Transparent because subjects should know the number of actual extractions (including observed and unobserved ones). Since this number is given by $5 \cdot t$, where t is the period, there is perfect collinearity between the period and the size of the sample.

Result 7. (confidence): *As the number of periods increases, the confidence of the subjects regarding the accuracy of their guesses increases in all treatments, although not always significantly. This effect does not seem to be due to an increase in the sample size.*

To finish, we note that we have controlled in all previous estimations for individual characteristics, such as gender, age, major, knowledge of statistics, risk aversion, and a measure of the subject's memory (how many 4-digit numbers shown during 60 s they are able to remember). The most interesting result regarding individual characteristics is that, other things equal, being a male increases the confidence on the guess made, around 6 and 7 p.p. respectively in Transparent (significant at 10%) and Fuzzy (significant at 12%). This result is in line with a vast literature that confirms that men in general are more (over)confident than women (e.g., Fellner and Maciejovsky, 2007; Niederle and Vesterlund, 2007), and even more when it is about some mathematics-related task (e.g. Dahlbom et al., 2010, Jacobsson et al., 2013). Regarding other variables, we stress that the number of semesters that a subject has studied statistics or econometrics at the university and her/his major has no systematic effect. Studying Engineering or Sciences, for instance, does not make a subject less likely to have inflated beliefs in Fuzzy and Opaque (i.e., more likely to deviate from our model's prediction).

4.4. A comparison of WYSIATI with other potential explanations

It is a natural question whether some alternative hypothesis could perform better or at least as well as WYSIATI in our experiment. For instance, subjects might be Bayesian but have different priors than WYSIATI assumes. To think about this,

[19] We have tested, from the estimated probit model, whether there is a statistical difference between the coefficients of the frequency intervals D1 and D2 in Transparent. We have not found evidence of such difference (p-value: 0.325).

let p_R and p_B respectively denote the probability of *disclosure* of a red and a blue ball. As we know, $p_R = 1$ and $p_B = 0.1$ in our experiment, but subjects are uncertain about those figures in Fuzzy and Opaque –participants in Opaque do not know the sample size *k* either; for parsimony, we abstract from this issue in what follows. In principle, a Bayesian might assign strictly positive probabilities to multiple states, e.g., a probability of 0.8 that $p_R = p_B = 1/2$, and a probability of 0.2 that $p_R = 7/8$ and $p_B = 1/3$. Then one could predict what such a Bayesian would guess in our treatments, covering many other different priors as well. While an exhaustive exploration along these lines appears to be extremely costly and possibly highly artificial, we present here two more tractable approaches. The first one simplifies matters by assuming that subjects assign all the probability mass to just one state, i.e., they are certain about p_R and p_B. Note that the WYSIATI principle is an example along these lines: it literally says that subjects act as if they believed for sure that $p_R = p_B = 1$ –a slightly broader, but similar principle is $p_R = p_B \leq 1$, which means that people do not expect sampling biases. The first alternative approach to WYSIATI that we consider here assumes that people believe for sure that there are selection issues, i.e., that p_R is different than p_B. Two extreme options possibly stand out here. On the one hand, people might be sure that p_R is much higher than p_B –yet subjects in this case should act in a similar way as in Transparent, and we know that this is not the case. On the other hand, subjects might believe that sampling strongly favors the blue balls, i.e., p_B is much higher than p_R. Given such priors, inflation should get even more acute than with WYSIATI. Between these two extremes there are other options, say, a relatively small bias favoring the red balls, e.g., $p_R = 0.6$, $p_B = 0.5$.

While a full comparison of all those options with the idea of selection neglect, i.e., $p_R = p_B$, appears to be again too costly, we here consider six heuristics alternative to WYSIATI which, in essence, represent different degrees of belief about the sampling bias in Fuzzy and Opaque. In a nutshell, each heuristic takes the subject's observed sample in each period and corrects for the bias by simply adding balls to that sample; then the subject's guess is derived by extrapolation, i.e., she guesses the rate closest to the empirical frequency of red in the whole enlarged sample (recall Fig. 1). The six considered heuristics are either adding 1, 2 or 3 red balls or 1, 2 or 3 blue balls to the observed sample. For clarity, consider the '+ 2 blue balls' heuristic and a subject who observes just 1 red ball in the first and second periods. If the subject applied WYSIATI, she should guess $\theta = 0.6$ in the second period (and in the first one too). If she added 2 blue balls in each period, in contrast, the frequency of red balls in the expanded sample would be 1/3 and hence the subject should guess $\theta = 0.4$ in both periods. As said, the different heuristics represent different beliefs about the sampling bias: in one extreme, '+3 red balls' indicates a belief that the blue balls are strongly over-selected, hence the need to 'compensate' with many red balls; in the other extreme, '+3 blue balls' signals that the subject expects strong over-selection of the red balls. The other heuristics represent less extreme beliefs about selection, whereas WYSIATI entails adding zero balls in every period.

To compare the performance of WYSIATI with the other heuristics, we perform six simulations, one for each heuristic. In each simulation, we derive each subject's guess in every period, assuming that all subjects follow the same heuristic in each period. We take into account that the maximum number of balls in each period is 5. For example, in the simulation that adds 2 blue balls, only one is actually added if the number of observed balls was 4. For each simulation, we compute subjects' guesses according to the empirical frequency of red balls in the enlarged sample, as explained above. We stress that we consider in all simulations the frequency of red balls in the enlarged sample, taking the given period and all prior ones –i.e., we do not assume subjects to be myopic as in the two-types model. Figs. 5 and 6 show the results for Fuzzy and Opaque, respectively.

As we can see, when adding red balls, the percentage of actual guesses replicated by the corresponding heuristics is only slightly below those replicated by WYSIATI in both treatments. Moreover, adding 1,2 or 3 red balls does not make any difference in most of the cases. This is most likely to happen because in our design the observed samples tend to contain many red balls and adding even more hence changes little in the predictions in comparison with WYSIATI. The simulations adding blue balls are more interesting. In Fuzzy, the average percentage (across periods) of actual guesses replicated by WYSIATI is 59%, and those replicated by the simulations are 51, 44 and 39%, respectively, when 1, 2 or 3 blue balls are added. In Opaque, the figures are: 62% of actual guesses explained by WYSIATI, and 57, 51 and 44% for the three simulations, respectively. This signals that WYSIATI is more empirically relevant than the other heuristics and suggests again that subjects do not tend to anticipate sampling biases in their inferences. It is also worth noting one pattern found in the two treatments: the higher the number of blue balls added, the higher the distance with WYSIATI. We believe that this occurs because, when just one blue ball is added, the empirical frequency of red balls is often still high, and thus, the prediction of this heuristic tends to coincide with WYSIATI (which happens to explain better our results). However, when we add two or three blue balls, this is no longer true.

To control for potential overlaps, we have computed the percentage of actual guesses that are replicated by each heuristic but not by WYSIATI. In a sense, this measures the marginal explanatory power of each heuristic (recall in any case that each heuristic explains less guesses overall than WYSIATI). In the scenarios that add red balls, this figure is on average, i.e., across all periods, a mere 2% in Fuzzy and 1% in Opaque. As said above, therefore, these heuristics seem to make almost always the same predictions as WYSIATI. In the simulations adding blue balls, the results are quite different. In Fuzzy, the average percentages are 9.7, 16.1 and 17.4% when 1, 2 and 3 blue balls, respectively, are added. In Opaque, the figures are 9.5, 13.1 and 14.7%, respectively. These values are arguably not very high, and in any case, we cannot perfectly discriminate whether they reflect the capacity of a few subjects to incorporate selection issues, or simply noise. Given the evidence from our regression analysis (see Table 5), however, we find more plausible the second possibility. Overall, therefore, we believe that Figs. 5 and 6 provide further evidence that individuals generally do not consider the potential existence of a selection bias when they make their guesses, in line with the WYSIATI principle.

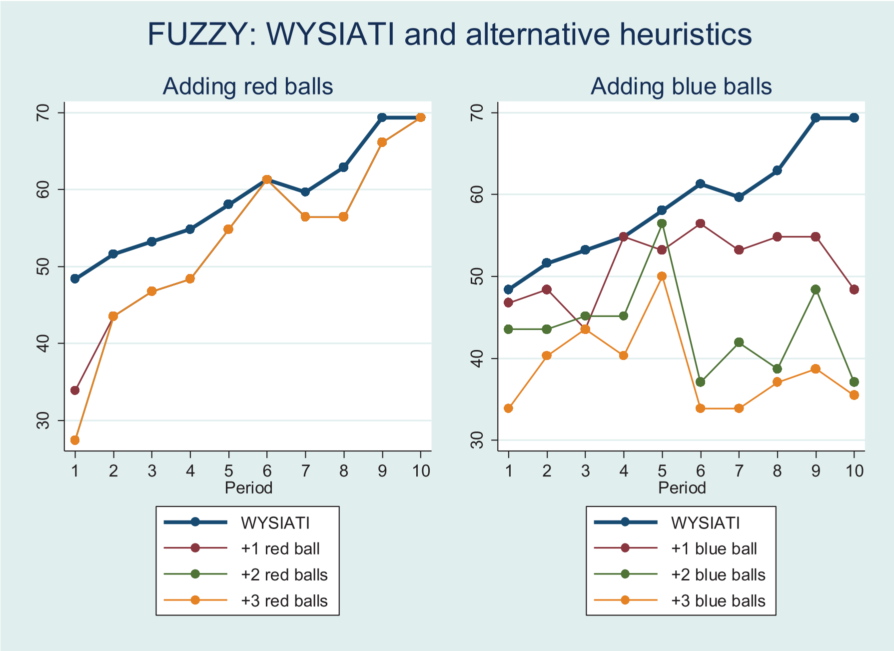

Fig. 5. Percentage of actual guesses in Fuzzy replicated per period by WYSIATI and alternative heuristics.

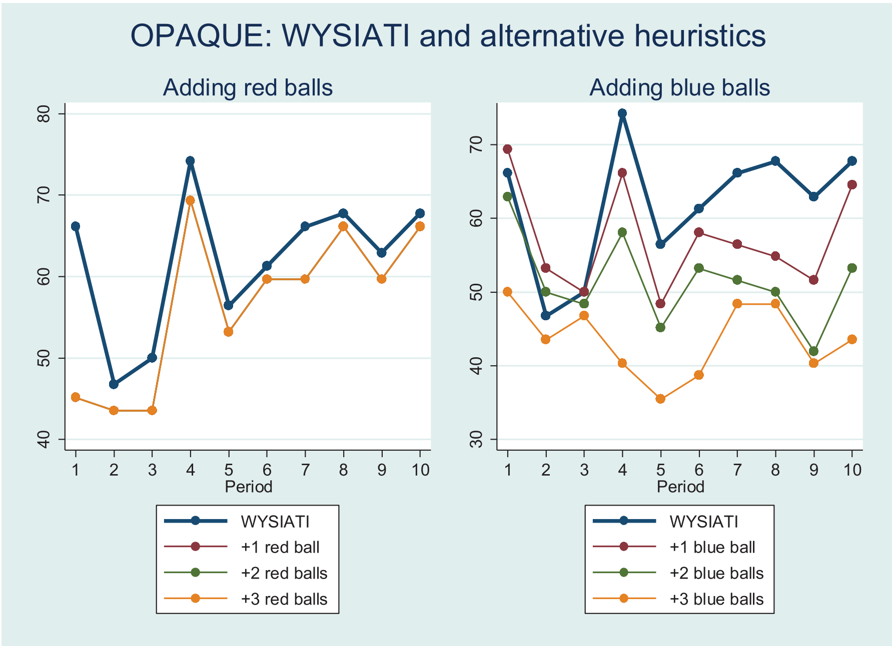

Fig. 6. Percentage of actual guesses in Opaque replicated per period by WYSIATI and alternative heuristics.

To further check (and somehow qualify) this point, however, the second alternative approach explored here assumes that subjects are Bayesian and consider two states of the world regarding the data generating process (DGP), i.e., either WYSIATI (undefined number of extractions in each period, $p_R = p_B = 1$) or the actual process (five extractions per period, $p_R = 1$, $p_B = 0.1$). This seems to us the most interesting possibility if we allow subjects to have non-degenerate priors about the DGP. Let $\Pr(W) > 0$ denote the prior that WYSIATI is true. Our question here refers to the sustainability of such a belief: given the data that subjects in Fuzzy and Opaque actually observe, when should their posteriors about WYSIATI converge towards one? If $\Pr(W) < 1$, our answer, *grosso modo*, is that sustainability requires that the subject sufficiently *overestimates* the actual number of red balls in her urn. We add here a few clarifying remarks to this (perhaps puzzling) statement and invite the reader to consult the web appendix for more detail. First, convergence does not depend on the initial value of Pr(W), at least within the range considered in our analysis. In contrast, second, sustainability is conditional on the actual number of red balls in the subject's urn. To start, those subjects with 60 balls in their urns will never overestimate that number and their beliefs will be in general unsustainable. Consequently, those subjects should guess as in Transparent (at least in the last periods), which incidentally is not what we observe.[20] When a subject has 10 or 40 red balls in her urn (particularly in the first case), on the opposite, many beliefs about the content of her urn lead to estimates of the average rate that are larger than the actual rate –e.g., a subject with uniform priors estimates an average rate of (1/3)*(10 + 40 + 60), which is larger than 10. These subjects should hence become increasingly convinced throughout the experiment that there is no selection issue, as the WYSIATI principle says. This second approach, therefore, hints that the findings in this paper can be partly explained even if we assume that subjects are (initially) uncertain about the existence of sampling issues. To grasp some of the intuition here, consider a subject who observes in some period no balls at all. While this outcome is rather unlikely in the state of the world corresponding to the actual DGP (it requires 5 blue balls to be extracted in a row), it is perfectly consistent with WYSIATI (if no balls are extracted, WYSIATI conceives a single outcome, with probability 1). If a subject observes no draws in many periods, therefore, she will become increasingly convinced that WYSIATI is the true DGP. Similarly, if the subject observes a small sample with just 1 ball, there are just two possible outcomes (red or blue), and the actually most likely observed one (red) has a relatively large prior if the subject believes with enough probability that there are many (say, 60) red balls in her urn. Hence, such evidence should also reinforce WYSIATI. In a sense, WYSIATI has an advantage in that tiny observed samples are more consistent with it (when the disclosed samples are larger, in contrast, they tend to have too many red balls, and that operates against WYSIATI).[21] Since subjects with 10 or 40 red balls are more likely to observe tiny samples which are not very unbalanced (provided that they put a high prior on the state '60 red balls in the urn'), our prediction about sustainability cited above follows. The moral is that even statistically sophisticated agents who allow for the existence of different DGP may end up inferring as the WYSIATI principle indicates, even if they initially assign a small (but incorrect) prior to that possibility. In slightly different words, since WYSIATI implies by its very nature smaller samples than the actual DGP and hence less outcomes among which to share the probability mass, Bayesian agents facing tiny samples can (wrongly) conclude that there are no sampling issues at all. While it is perhaps debatable whether actual agents show the degree of sophistication assumed, the analysis hence suggests that the WYSIATI principle might be the outcome of people updating their beliefs about the DGP.

5. Related literature

Our paper is closely related to a psychological literature claiming that people often neglect selection issues. For instance, subjects in Nisbett and Borgida (1975) equated some observed, extreme behavior to the modal one for the population, both when the sampling procedure was unspecified, i.e., ambiguous, or explicitly described as random. Hamill et al. (1980) similarly report that subjects' attitudes toward welfare recipients –i.e., their maturity, capacity for hard-working, etc.– were equally influenced by exposure to a single case, both when subjects were told that the case was highly (i) typical or (ii) atypical. Several factors could explain selection neglect in these studies. For instance, Hamill et al., 587) contend that "the vivid sample information probably serves to call up information of a similar nature from memory [...]. These memories then would be disproportionately available when judgments are later made about the population". In our experiment, in contrast, memory biases arguably play little (if any) role, e.g., subjects are always given feedback about prior extractions, they have the instructions available, and it is unlikely that any of the extractions observed are cognitively more 'available', as they only differ in their color. Relatedly, we control for the subjects' priors and the information that they use in their inferences so that we can provide evidence on the extent to which individuals follow the WYSIATI principle.

This study is also related to an emerging empirical literature in Economics studying belief updating when people do not observe some signal realizations. In general terms, our contribution to this literature is to provide detailed evidence at an individual level in line with the WYSIATI principle, documenting as well how ambiguity affects selection neglect, even when both the evidence accumulated by the subjects and contextual cues strongly suggest the existence of selection biases. We also explore the reasons why people sometimes deviate from the principle, finding that they can be partly attributed to mistakes. Among the most closely related papers, Koehler and Mercer (2009) offer evidence that companies selectively (and possibly strategically) advertise their better-performing stock mutual funds. Nonetheless, investors (both experts and novices) respond to mutual fund advertisements as if such data were unselected, in line with the WYSIATI principle. The authors' preferred explanation is that people do not automatically think about sampling matters unless the exact data selection process is made transparent or cued. Our findings add further evidence in this line from a controlled experimental settings, indicating moreover that the mere accumulation of evidence signaling a selection issue is not enough to make people doubt about their inferences. In turn, the experiment in Enke (2020) has a random variable which is uniformly distributed over the set {50, 70, 90, 110, 130, 150}. Out of six realizations, subjects first observe one of them and then also

[20] We have considered the specifications in Table 5 for the probability of making a guess that deviates from the WYSIATI principle, but now adding a dummy variable with value 1 for subjects with 60 red balls in the urn and 0 otherwise. We have found no significant results for this variable.

[21] A further advantage of WYSIATI is that it is a less precise or more ambiguous state of the world than the actual DGP. That is, while the actual DGP says that there will be exactly 5 extractions in any period, WYSIATI is undefined a priori on this issue (it just says that the number of actual draws could vary).

those above (below) 100, depending on whether the first observed outcome is above (below) 100 as well. Based on such evidence, subjects must infer the average outcome. Akin to our Transparent treatment, therefore, there is a sampling bias, but subjects know the selection rule. Even in this case, interestingly, a substantial share of the subjects infers as if they fully neglected the non-observed outcomes, i.e., in line with the WYSIATI principle, thus suggesting that the scope of the problem is even broader than what our parsimonious framework concedes. Similarly, but in the context of an investment experiment, Barron et al. (2019) find evidence for selection neglect in a treatment in which subjects know the data generating process. The results from these two latter studies, therefore, contrast with the evidence from Transparent, where subjects fully account for the sampling bias. Hence, we complement this perspective by showing that, keeping complexity constant, the ambiguity of the specific selection rule is a factor behind selection neglect; we expand on this issue in the conclusion.

We also contribute to a literature on quasi-Bayesian inference, where agents misread the world due to cognitive limitations, but are then assumed to operate as Bayesians given this misreading. A strand of this literature, including Barberis et al. (1998), Rabin and Schrag (1999), Mullainathan (2002), Rabin (2002), and Benjamin et al. (2016), explore inference when agents misread or misremember the signals they observe. In another strand, more closely related to our paper, agents do not misinterpret the signal, but have a simplified representation of the state space. For instance, Gennaioli and Shleifer (2010) and Bordalo et al. (2016) consider a world defined by several dimensions. Some of them are fixed by the available data and hypothesis that the agent wants to evaluate, but others are uncertain. Agents simplify the world by focusing on the 'stereotypical' values of these residual dimensions, that is, those values that are frequently observed in a world in which the target hypothesis is true (as well as the data), but uncommon when the complementary hypothesis is true. In our framework, for instance, the residual dimensions of state $\omega = (\theta_i, b, r)$ are possibly the pair (b, r), while any hypothesis refers to the rate θ. A difference with Gennaioli and Shleifer (2010) is that the WYSIATI principle does not mean that people focus on the stereotypical values of b and r, but on those that are either non-ambiguous (given the frame) or simplify most inference (i.e., assuming that 'what you see is all there is'). The idea that people have reduced representations of the world also relates our paper to the behavioral literature on inattention –see Gabaix (2019) for a nice review. Another related issue is correlation neglect –DeMarzo et al., 2003, Glaeser and Sunstein, 2009; Levy and Razin, 2015; Ortoleva and Snowberg, 2015. This is the idea that, when people exchange ideas with others, they tend not to realize how often others have observed the same evidence as them –e.g., the same news. Since WYSIATI implies a sampling-bias neglect (unless the selection rule is non-ambiguous), it implies as well that agents will discount the possibility that they mostly interact with people who have similar evidence.

There is a growing literature that studies how others' private information affects players (and their decisions) in strategic settings, mainly in auctions and voting. In strategic interactions, (naïve) players may underestimate selection effects or more precisely the correlation between other players' actions and the information they have. Kagel and Levin (1986), Holt and Sherman (1994), Eyster and Rabin (2005), and more recently Araujo et al. (2021) have, among others, studied implications of such correlation neglect in markets affected by adverse selection. In addition, Esponda and Vespa (2014) find in a simple voting situation that the majority of the subjects fail to be strategic, but this mistake is primarily driven by the fact that these subjects are unable to extract information from hypothetical events. This failure turns out to be fairly robust to experience, feedback, and hints about pivotality. Similarly, Esponda and Vespa (2018) find that subjects are unable to account optimally for endogenous sample selection (caused by private information of other players). The authors report that subjects respond to observed data, but at the same time pay no attention to the possibility that this data may be biased. Jin et al. (2021) study an information disclosure game. Besides finding a high overall disclosure rate among senders, they observe that receivers are not sufficiently skeptical about undisclosed information, as they (mistakenly) think that non-disclosed states are higher than they are in reality. While these articles consider selection neglect (and its consequences in strategic settings), we explore several conditions that might affect the prevalence of this phenomenon, including ambiguity, the evidence accumulated, and explicit contextual cues.

6. Conclusion

Our claim is that, when faced with some evidence, people often update their beliefs as if they faced a representative sample of the population. The main reason is that data-generating procedures are often ambiguous or uncertain. Under these conditions, people simplify matters and infer *as if* there was no sampling bias; a heuristic we call the WYSIATI principle. As Kahneman (2011, p. 114) states, "System 1 is not prone to doubt". As a result, people tend to equate the frequency of observation and occurrence of the target event when there are ambiguous selection problems.

The problem of course appears if the two frequencies are actually different. Consider for instance journalism. If rare or shocking events ('man bites dog') receive disproportionate coverage while ordinary, 'nice' events are not considered newsworthy, readers may conclude that extreme and terrible events are more common than they are in reality. The press can in that manner feed pessimism, leading people (even experts) to inflate the probability of a massive stock-market crash, the danger of terrorist attacks or airplane crashes, the crime rate, or the prevalence of genocide and war in the world. In this line, a survey conducted by the end of 2015 revealed that majorities in 15 countries around the world believed that the world is getting worse –in the US, for instance, 65% thought so, and only 6% responded that the world is getting better.[22]

Partly motivated by these phenomena, we explore the empirical relevance of the WYSIATI principle, and find that it replicates a substantial share of the subject's individual guesses. Further, many of the guesses not explained by the heuristic seem to be caused by simple mistakes, not by a different logic of inference. In this line, the likelihood of a deviation increases as subjects are less reflective, apparently put more emphasis on recent observations (rather than taking into account the whole sample), or have borderline evidence. The results across treatments suggest moreover that ambiguity strongly influences the prevalence of selection neglect, even when people become experienced in the inference problem at hand and the context hints that the evidence observed might be selective. The alternative theory that subjects cannot make correct inferences is at odds with the data from Transparent, where inflation is absent and the average guess converges over time to the Bayesian prediction. The idea that subjects are unaware of potential sampling biases seems in turn incoherent with the fact that subjects in Fuzzy are not more cautious or conservative than those in Opaque: the inflation rate and the average confidence are basically the same in these two treatments. Finally, subjects do not *seem* to operate with complex priors on the data-generating process, because their guesses and confidence levels in Fuzzy and Opaque increase as the experiment proceeds (although our analysis in 4.4 suggests some nuances in this respect).

Our results indicate that inflation can be a lasting phenomenon, but also one that can be alleviated if the exact selection rule is non-ambiguous (and sufficiently simple, we presume). Alternatively, the objective probability of the target event could be given. Following with our prior example, the media might reduce the inflation generated by their emphasis on tragedy and calamity if, together with the news on some uncommon event, they reported empirical frequencies of that type of events. In our setting at least, people can adjust their beliefs in a Bayesian manner (although perhaps not perfectly) if they get sufficiently detailed evidence.

There are a number of issues that we leave for future research, from which we mention three. One refers to group polarization –e.g., Sunstein, 2001. This phenomenon is indeed a multifaceted one, but we conjecture that part of the problem is that groups do not realize that the information they receive is not sufficiently 'varied', i.e., depicting different points of view and a representative sample of data. While polarization is possibly compounded by the fact that groups do not trust those sources of information that contradict their worldview, it seems to us that it would be alleviated if agents were aware of the potential biases in the samples they observe. Another issue is that selection neglect possibly implies a neglect as well of the statistical causes of the bias. For instance, do people tend to under-estimate the prevalence of (a) selection effects and (b) biased reporting of evidence by others, either intentional or unintentional? To illustrate point (a), suppose that, for whatever reasons, female scientists appear less frequently than their male peers in textbooks, magazines, and TV shows for children. If children ignore the potential selection effects causing female scientists to be less numerous than their male peers, they will think as if 'what they see is all there is', hence affecting their stereotypical perceptions of who is (and possibly who can be) a scientist.[23]

A final topic is sampling-bias neglect in complex statistical problems: multiple signals, some of them time-varying, absence of feedback, etc. Compared to the *relative* simplicity of Fuzzy and Opaque, our impression is that, other things equal, inflation should be even more severe and lasting in complex environments. This point is *maybe* related to the evidence from Enke (2020) and Barron et al. (2019), who find selection neglect even when the selection rule is known. In contrast, we observe no sampling bias neglect in Transparent. While explaining these differences across the studies is out of the scope of this paper, one could argue that inference in Enke (2020) and Barron et al. (2019) is cognitively more demanding than in Transparent. In Enke (2020), for instance, the signal can take six numerical values, i.e., 50,70,90,110,130, or 150. In our design, the signal is either red or blue-colored. In the experiment in Enke (2020), further, subjects have to estimate an average of six of these numbers, some observed, but others unobserved. To estimate θ in some period t, in contrast, participants in our study must simply divide the number of red extractions by 5•t. Perhaps people (or at least some types) also follow the WYSIATI principle in non-ambiguous but computationally complex scenarios, as their limited attention is focused on the ensuing computations. More research is indeed warranted.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.12.036.

[22] See https://yougov.co.uk/news/2016/01/05/chinese-people-are-most-optimistic-world/. Indonesia and China were the only two countries surveyed where majorities did not respond that the world outlook was worsening.

[23] Children's drawings of scientists may reflect stereotypical perceptions. The meta-analysis of 5 decades of Draw-A-Scientist studies in Miller et al. (2018) finds that American children draw female scientists more often in later decades, maybe due to an increasingly more balanced depiction in the mass media. In the 1960s and 1970s, 99.4% of the kids did not draw female researchers; today around one in three does it.

References

Araujo, F.A., Wang, S.W., Wilson, A.J., 2021. The times they are a-changing: experimenting with dynamic adverse selection. Am. Econ. J. Microecon 13 (4), 1–22. doi:10.1257/mic.20190088.
Barberá, P., 2014. Birds of the same feather tweet together: bayesian ideal point estimation using twitter data. Political Anal. 23, 76–91.
Barberis, N., Shleifer, A., Vishny, R.W., 1998. A model of investor sentiment. J. Financ. Econ. 49, 307–343.
Barron, K., Huck, S., Jehiel, P., 2019. Everyday Econometricians: Selection Neglect and Overoptimism When Learning from Others. Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin WZB Discussion Paper No. SP II 2019-301.
Benjamin, D.J., Rabin, M., Raymond, C., 2016. A model of non-belief in the law of large numbers. J. Eur. Econ. Assoc. 14 (2), 515–544.
Benjamin, D.J., Bernheim, D., DellaVigna, S., Laibson, D., 2019. Errors in probabilistic reasoning and judgment biases. Handbook of Behavioral Economics. Elsevier Press.
Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2016. Stereotypes. Q. J. Econ. 131 (4), 1753–1794.
Dahlbom, L., Jakobsson, A., Jakobsson, N., Kotsadam, A., 2010. Gender and overconfidence: are girls really overconfident? Appl. Econ. Lett. 18, 325–327.
DeMarzo, P., Vayanos, D., D, Zwiebel, J., 2003. Persuasion bias, social influence and uni-dimensional opinions. Q. J. Econ. 118 (3), 909–968.
Enke, B., 2020. What you see is all there Is. Q. J. Econ. 135 (3), 1363–1398.
Esponda, I., Vespa, E., 2014. Hypothetical thinking and information extraction in the laboratory. Am. Econ. J. Microecon. 6 (4), 180–202.
Esponda, I., Vespa, E., 2018. Endogenous sample selection: a laboratory study. Quant. Econ. 9, 183–216.
Eyster, E., Rabin, M., 2005. Cursed equilibrium. Econometrica 73 (5), 1623–1672.
Fiedler, K., 2000. Beware of samples! a cognitive-ecological sampling approach to judgment biases. Psychol. Rev. 107, 659–676.
Fellner, G., Maciejovsky, B., 2007. Risk attitude and market behavior: evidence from experimental asset markets. J. Econ. Psychol. 28, 338–350.
Fischbacher, U., 2007. Z-tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Frederick, S., 2005. Cognitive reflection and decision making. J. Econ. Perspect. 19 (4), 25–42.
Gabaix, X., Bernheim, D., DellaVigna, S., Laibson, D., 2019. Behavioral inattention. Handbook of Behavioral Economics, 2. Elsevier.
Gächter, S., Renner, E., 2010. The effects of (incentivized) belief elicitation in public goods experiments. Exp. Econ. 13, 364–377.
Gennaioli, N., Shleifer, A., 2010. What comes to mind. Q. J. Econ. 125 (4), 1399–1433.
Glaeser, E.L., Sunstein, C.R., 2009. Extremism and social learning. J. Leg. Anal. 1 (1), 263–324.
Hamill, R., Wilson, T.D., Nisbett, R.E., 1980. Insensitivity to sample bias: generalizing from atypical cases. J. Pers. Soc. Psychol. 39 (4), 578–589.
Holt, C.A., Sherman, R., 1994. The loser's curse. Am. Econ. Rev. 84, 642–652.
Jacobsson, N., Levin, M., Kotsadam, A., 2013. Gender and overconfidence: effects of context, gendered stereotypes, and peer group. Adv. Appl. Sociol. 3, 137–141.
Jehiel, P., 2018. Investment strategy and selection bias: an equilibrium perspective on overoptimism. Am. Econ. Rev. 108 (6), 1582–1597.
Jin, G.Z., Luca, M., Martin, D., 2021. Is no news (perceived as) bad news? An experimental investigation of information disclosure. Am. Econ. J. Microecon. 13 (2), 141–173. doi:10.1257/mic.20180217.
Kagel, J.H., Levin, D., 1986. The winner's curse and public information in common value auctions. Am. Econ. Rev. 76, 894–920.
Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus & Giroux.
Keren, G., 1991. Calibration and probability judgements: conceptual and methodological issues. Acta Psychol. 77, 217–273 (Amst).
Koehler, J.J., Mercer, M., 2009. Selection neglect in mutual fund advertisements. Manag. Sci. 55 (7), 1107–1121.
Levy, G., Razin, R., 2015. Correlation neglect, voting behavior, and information aggregation. Am. Econ. Rev. 105 (4), 1634–1645.
Levy, G., Razin, R., 2019. Echo chambers and their effects on economic and political outcomes. Annu. Rev. Econ. 11, 303–328.
Miller, D.I., Nolla, K.M., Eagly, A.H., Uttal, D.H., 2018. The development of children's gender-science stereotypes: a meta-analysis of 5 decades of U.S. draw-a–scientist studies. Child Dev. 89 (6), 1943–1955.
Mullainathan, S., 2002. A memory-based model of bounded rationality. Q. J. Econ. 117 (3), 735–774.
Mullainathan, S., Shleifer, A., 2005. The market for news. Am. Econ. Rev. 95 (4), 1031–1053.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Q. J. Econ. 122, 1067–1101.
Nisbett, R.E., Borgida, E., 1975. Attribution and the psychology of prediction. J. Pers. Soc. Psychol. 32, 932–943.
Ortoleva, P., Snowberg, E., 2015. Overconfidence in political behavior. Am. Econ. Rev. 105 (2), 504–535.
Rabin, M., 2002. Inference by believers in the law of small numbers. Q. J. Econ. 117, 775–816.
Rabin, M., Schrag, J.L., 1999. First impressions matter: a model of confirmatory Bias. Q. J. Econ. 114 (1), 37–82.
Streufert, P., 2000. The Effect of underclass social isolation on schooling choice. J. Pub. Econ. Theory 2, 461–482.
Sunstein, C.A., 2001. Echo Chambers: Bush V. Gore, Impeachment, and Beyond. Princeton University Press.
Tversky, A., Kahneman, D., 1971. Belief in the law of small numbers. Psychol. Bull. 76, 105–110.