Bayes and biases. Questioning the notion of 'confirmation bias'

Marion Vorms

Bayes and biases. Questioning the 'confirmation bias'

Marion VORMS

ABSTRACT. – *'Confirmation bias' refers to people's alleged tendency to select the information that supports what they already believe (or what they want to believe), as well as to interpret novel information as backing their favorite hypotheses. In this paper, I propose a critical appraisal of some uses of this notion. More generally, I criticize some existing attempts at explaining a large number of supposedly irrational social phenomena in terms of 'cognitive biases'. I question the evidential value of some experimental results intended to show the existence of a systematic deviation from accuracy in our representations, and I finally suggest that the Bayesian approach to the psychology of reasoning enables us to account for some of those results in terms that are compatible with the hypothesis according to which our inferences conform to a rational norm.*

Whatever the explanation that is put forward to explain this, it is a commonplace that people usually manage to select information that supports what they believe (or want to believe) and to interpret the information available to them in such a way that it supports their preferred hypotheses. This alleged tendency to confirm what we already believe is central to the repertoire of so-called 'cognitive biases'.

While the concept of bias is part of everyday language, where it denotes a preference, a more or less avowed inclination whereby we lack impartiality or neutrality, it has experienced a remarkable boom in psychological discourse in recent decades. References to myriad 'cognitive biases' that undermine our reasoning extend far beyond the sphere of cognitive and social psychology specialists, and feed into many more or less scientific reflections on the supposed shortcomings of the human mind.

While Amos Tversky and Daniel Kahneman's 'Heuristics and Biases'

programme[1] contributed greatly to this popularity by dramatically demonstrating that most of us commonly make significant errors in our probabilistic estimates and judgments under uncertainty, it is a more polymorphous and less well-defined notion of bias, mostly derived from studies in social psychology, that often takes centre stage. In this context, biases are difficult to distinguish from simple errors of reasoning; moreover, 'biases' sometimes designate the causes of such errors and sometimes their observable effects. Seen in this way, biases have to do with the human mind processing information incorrectly and they are harmful to us – particularly because they threaten the accuracy of our beliefs, causing us to miss our goals by making the wrong decisions. As a result, they are often interpreted as a form of irrationality. The link between these different aspects is rarely questioned, whether we attribute the biases to our cognitive limitations or to emotional motivations.

Cognitive biases – and in particular the so-called 'confirmation bias' – are used to explain a wide range of social phenomena, from the effectiveness of rumours to opinion polarisation[2], the success of 'conspiracy theories'[3] and science denial (e.g. fuelling anti-vaccination movements). But what is the empirical evidence for the existence of such biases? To ascribe them explanatory power, we need to have a broader and more robust empirical basis showing their existence than that provided by the scattered observations for which they are supposed to provide an explanation. In order to avoid explaining what seems irrational by a circular appeal to the idea of a disposition to irrationality, it is necessary to examine the standard(s) of rationality against which such a judgement is made and to note the existence of systematic deviations from these standards.

This article has two complementary objectives. On the one hand, it aims to provide a critical perspective on the notion of cognitive bias (in particular confirmation bias) and certain uses of it, which are often accompanied by the idea that humans have a disposition towards irrationality that leads them to make systematic errors. To do this, I will draw in part on analyses produced by proponents of the 'Bayesian' approach to the psychology of reasoning. Giving the reader an idea of this burgeoning research programme is my second aim.

After briefly tracing the history of the notion of confirmation bias, I will introduce what David Over has called the 'new paradigm' in the psychology of reasoning[4], which is part of the 'Bayesian turn' in

[1] A. TVERSKY, D. KAHNEMAN, Judgment under uncertainty: heuristics and biases, *Science*, 1974, 185 (4157), p. 1124-31.

[2] See S. LEWANDOWSKY, U. ECKER, C. SEIFERT, N. SCHWARZ, J. COOK, Misinformation and its correction: continued influence and successful debiasing, *Psychological Science in the Public Interest*, 2012, 13 (3), pp. 106-31.

[3] See P. HUNEMAN, M. VORMS, Is a unified theory of conspiracy possible?, *Argumenta*, 2018, 3 (2), pp. 49-72.

[4] D. OVER, New paradigm psychology of reasoning, *Thinking and Reasoning*, 2009, 15 (4), p. 249-65.

philosophy, artificial intelligence, statistics, and cognitive science. Finally, drawing from this perspective, I will present some criticisms to the appeal to biases as an explanation of – and evidence for – human irrationality.

CONFIRMATION BIAS: FROM TEST STRATEGY TO MOTIVATED DISTORTION OF INFORMATION

Depending on the context, the term 'confirmation bias' covers a wide range of phenomena[5] corresponding to the different ways in which our beliefs or expectations influence how we search for, select, retain, interpret or evaluate information. After briefly reviewing the history of this concept in cognitive psychology, I will focus on its importance in social psychology. In doing so, I will draw heavily on the work of Ulrike Hahn and Adam J. L. Harris.[6]

Historical overview

Although empirical studies of biases have been conducted in psychology since the beginning of the 20th century[7], it was the work of Peter Wason in the 1960s that first demonstrated experimentally what would come to be known as 'confirmation bias'.

Peter Wason and the 'positive test strategy'

Peter Wason[8] is interested in our hypothesis-testing strategies and, more specifically, the type of information we seek first. In his experiment, he gives participants the task of finding the rule (known only to the experimenter) that governs a set of number triads – the first example they are given is '2-4-6'. The participants have to propose several examples of triads, and the experimenter tells them whether or not they are in accordance with the rule that has to be found. What he finds is that a significant proportion of subjects tend to propose triads that are positive instances of the rule they have in mind (and which they are trying to find out if it is really the one they are looking for) rather than counter-examples.[9]

---

[5] See R. S. NICKERSON, Confirmation bias: a ubiquitous phenomenon in many guises, *Review of General Psychology*, 1998, 2, pp. 175-220.
[6] U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality?, *Psychology of Learning and Motivation*, 2014, 61, pp. 41-102.
[7] See U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, pp. 43-4.
[8] P. C. WASON, On the failure to eliminate hypotheses in a conceptual task, *Quarterly Journal of Experimental Psychology*, 1960, 12 (3), p. 129-40.
[9] A subject who assumes that the rule to be found is 'being a sequence of even numbers in

Wason sees this as a violation of the Popperian canon, which prescribes that we should seek to falsify hypotheses rather than to confirm them.[10] However, it is not clear that this tendency – whether or not we call it 'confirmation bias' – is harmful in most situations, nor that it represents a blatant breach of rationality. In fact, as Joshua Klayman and Young-Won Ha have pointed out[11], this 'positive test strategy' is just as likely to lead to falsification as to confirmation. Moreover, although it is sub-optimal in the context of the task proposed by Wason[12], in the sense that it does not maximise the 'expected value of information'[13], this strategy is advantageous in most real-life situations.[14]

Finally, two points should be stressed. First, the alleged 'confirmation bias' highlighted by Wason concerns the search strategy for the information, not the evaluation of its probative force for or against a hypothesis. Second, this tendency can hardly be explained in terms of motivation or emotional attachment to a hypothesis. As we shall see, 'confirmation bias' will gradually take on a more clearly motivational meaning and will no longer concern only the search for (or selection of) information but also its interpretation – its evaluation and assimilation.

'Conservatism'

'Confirmation bias' has also been used to describe the 'conservatism' in belief revision in the face of new data, as highlighted by experimental work in the 1960s[15] – but here, the notion has a different meaning. Using problem-solving tasks based on elementary probability calculations (typically using urns, where the goal is to guess which contains a majority of tokens of a particular colour based on successive draws), this research shows that subjects revise their beliefs in the right order of magnitude, but in a more moderate way than the calculation of

---

ascending order' will suggest '4-6-8' or '8-12-14' rather than '3-5-7' or '8-6-4'. To make an analogy, in a game of guessing what your partner is thinking about, the player who assumes that his partner is thinking about a horse will ask him if the thing he is thinking about has a mane, rather than if it has a fin.

[10] K. POPPER, *The logic of scientific discovery*, London, Hutchinson, 1959.

[11] J. KLAYMAN, Y. HA, Confirmation, disconfirmation, and information in hypothesis testing, *Psychological Review*, 1987, 94 (2), pp. 211-28.

[12] Since the rule to be found ('being a sequence of numbers in ascending order') is very broad, the subjects hypothesise stricter rules, naturally suggested by the first example given to them (e.g. 'being a sequence of even numbers in ascending order'), and propose triads which a fortiori fall within the scope of the broader rule. The experimenter's systematically positive response does not allow them to eliminate hypotheses that are nevertheless incorrect.

[13] See U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, p. 74.

[14] See M. OAKSFORD, N. CHATER, A rational analysis of the selection task as optimal data selection, *Psychological Review*, 1994, 101 (4), p. 608-31.

[15] See C. R. PETERSON, L. R. BEACH, Man as an intuitive statistician, *Psychological Bulletin*, 1967, 68 (1), pp. 29-46; P. SLOVIC, S. LICHTENSTEIN, Comparison of Bayesian and regression approaches to the study of information processing in judgment, *Organizational Behavior and Human Performance*, 1971, 6 (6), pp. 649-744.

probabilities would dictate – hence the term 'conservatism'.

Here the normative standard is clear: participants' responses can be accurately assessed in terms of probability calculus. Furthermore, as with Wason's task, the subjects' 'conservatism' cannot be explained in terms of emotional motivation. Finally, the deviation observed is systematic but moderate; it is more a form of epistemic inertia than a serious deviation from what the probability calculus prescribes. The probability calculus thus provides a 'good approximation for a psychological theory of inference'[16], which makes it possible to speak of 'man as an intuitive statistician'.

Kahneman and Tversky's 'Heuristics and Bias' programme

Such optimism, as underlined by Hahn and Harris[17], has been largely tempered by the work of Kahneman and Tversky, who have highlighted significant deviations from the prescriptions of probability and expected utility theory in the decisions we make under uncertainty and in our probability judgment. These deviations are so important that it is not even necessary to use quantitative models to see them. A presentation of their programme would go far beyond the scope of this article, though. Indeed, the biases highlighted by Kahneman and Tversky are only marginally related to belief revision – and consequently to the confirmation bias, although some of Kahneman's own comments, widely reported in the secondary literature, tend to create some confusion.[18] However, it would be wrong to deny the influence that the success of this work had on the development of the notion of bias in social psychology. The question of the links between these research traditions is a complex one.[19]

The study of cognitive biases in social psychology

While the biases highlighted by Kahneman and Tversky are deviations from clearly identified normative standards, the same cannot be said of

[16] C. R. PETERSON, L. R. BEACH, Man as an intuitive statistician, p. 42.

[17] U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, p. 49.

[18] In his book for the general public, Daniel Kahneman (*Thinking, fast and slow*, London, Penguin, 2011) repeatedly suggests a link between confirmation bias and certain phenomena he studies (in particular the heuristic of avaibility, itself linked to the tendency towards credulity, supposedly highlighted by D. T. GILBERT, D. S. KRULL, P. MALONE, Unbelieving the unbelievable. Some problems in the rejection of false information, *Journal of Personal and Social Psychology*, 1990, 59 (4), p. 601-13 – see M. VORMS, A.J.L. HARRIS, S. TOPF, U. HAHN, Plausibility matters: A challenge to Gilbert's 'Spinozan' account of belief formation, *Cognition*, 2022, 220) although the relationship between these concepts is still unclear.

[19] See J. I. KRUEGER, D. C. FUNDER, Towards a balanced social psychology: causes, consequences, and cures for the problem-seeking approach to social behavior and cognition, *Behavioral and Brain Sciences*, 2004, 27 (3), p. 313-27 [316].

the cognitive biases studied in social psychology. Given the long, heterogeneous and sometimes incoherent[20] list of 'biases' that can be drawn up, it is difficult to come up with a precise definition. While all biases have to do with deficiencies in reasoning that supposedly taint the process of belief formation and condemn us to erroneous or inaccurate representations of reality, an explicit characterisation of what constitutes correct reasoning is lacking. From optimism bias to false consensus bias, not forgetting – among many others – retrospective bias, attribution bias, confirmation bias, and the various variations of each, a pessimistic view of the human mind emerges, to say the least. This pessimistic view is echoed and reinforced by a number of popularisers[21] who try to draw up an exhaustive list[22] of the pitfalls we should be aware of – whether in an attempt to avoid them or to bemoan our spectacular maladjustment.

Without any pretension to exhaustiveness, I will only be looking at confirmation bias here. More specifically, my focus is on one aspect of it, the 'biased assimilation' of information.[23]

A version of 'motivated' confirmation bias: biased assimilation of information

Our alleged tendency to confirm our beliefs with new evidence is likely to occur in at least two stages: when we *select* information and when we *interpret* it. The selection of confirmatory information seems to be more difficult to uncover experimentally than its interpretation (leaving aside considerations of testing strategies such as those studied by Wason, for example). In fact, as Hahn and Harris point out[24], the content of information is only partially correlated with its source; even if I only consult sources that I know will generally provide information in line with my beliefs, this cannot guarantee that I will not get contradictory information.[25] If, on the other hand, we use the term 'information selection' to mean the selection of the information itself rather than its

[20] See J. I. Krueger, D. C. Funder, Towards a balanced social psychology: causes, consequences, and cures for the problem-seeking approach to social behavior and cognition, table 1, p. 317.

[21] E.g. the proponents of so-called 'zetetics' who, in their laudable effort to combat errors in reasoning and the spread of false and socially harmful representations, often dogmatically present the results of 'the science' of cognitive biases.

[22] See the 'cognitive bias codex', a collaborative categorization of biases : https://de.m.wikipedia.org/wiki/Datei:Cognitive_bias_codex_en.svg.

[23] For a summary table of the main types of phenomena that fall into the category of 'confirmation bias', see U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, table 2.1, p. 46.

[24] *Ibid*, p. 71.

[25] For a meta-analysis of studies on 'selective exposure' to information, which concludes that the empirical evidence in favour of such a bias is weak (due in particular to the existence of results pointing in the direction of an opposite bias), see W. HART, D. ALBARRACÍN, A. H. EAGLY, I. BRECHAN, M. J. LINDBERG, L. MERRILL, Feeling validated versus being correct: a meta-analysis of selective exposure to information, *Psychological Bulletin*, 2009, 135 (4), p. 555-88.

source, then we have to admit that a process of evaluation or monitoring has already taken place: I cannot select information that 'suits' me without having at least a cursory acquaintance with its content.

It is, therefore, the evidence of systematic distortion at the level of *judgment* – of the *interpretation* of information – that I will focus on here. In one of its most common meanings, confirmation bias refers to our alleged tendency to minimise the weight of information that contradicts our beliefs and, conversely, to exaggerate the weight of information that confirms them. Most of the time, this bias is assumed to be 'motivated' – that is, we look for reasons to believe what we want to believe. This alleged bias is used to explain many phenomena, such as the persistence of beliefs despite, and sometimes because of, the presentation of disconfirming information – making debunking strategies counterproductive, for example.

Biased assimilation and belief polarisation

In a famous 1979 study, Charles G. Lord, Lee Ross, and Mark Lepper[26] claimed to demonstrate that subjects were 'biased' in their 'assimilation' of information, as evidenced in particular by the polarisation of their beliefs and opinions in the face of 'mixed evidence'. Subjects divided into two groups, one made up of supporters of the death penalty who believed in its deterrent effect and the other made up of opponents who did not believe in such an effect[27], were successively presented (in controlled and counterbalanced order) with extracts from alleged scientific studies (in reality written by the experimenters) supporting or denying the existence of a deterrent effect of the death penalty. While the effects of the arguments presented to the subjects should, according to the experimenters, cancel each other out – neutralise each other – the subjects' responses show a tendency to take the arguments in favour of their initial opinion at face value and to easily detect flaws in the opposing arguments. In addition to these differences in the evaluation of the quality of the arguments presented to them (an evaluation they were explicitly asked to make), the results showed a polarisation of beliefs and opinions: proponents of the death penalty declared themselves more convinced of its deterrent effect at the end of the experiment, and vice versa.

According to the authors of this study, while it may be reasonable to give more or less credence to information depending on whether it supports or contradicts what we believe to be true (p. 2106), such an

[26] C. G. LORD, L. ROSS, M. R. LEPPER, Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence, *Journal of Personality and Social Psychology*, 1979, 37 (11), p. 2098-109.

[27] Although these two aspects (being for or against the death penalty / believing in its deterrent effect or not) are potentially independent, the two groups here are homogeneous with regard to these two parameters, so I will not distinguish between them.

evaluation should not, in turn, allow the information to reinforce the original belief (against which its credibility was assessed). In other words, the credence we give to information because it supports a belief we already hold and the corresponding discrediting of information that contradicts it should not, in their view, justify reinforcing that belief – at the risk of making it unfalsifiable (p. 2107).

Whatever the epistemological basis of such considerations – to which I will return in the third part of this article – let me highlight an important aspect of Lord et al.'s approach that is characteristic of many studies of biases in social psychology. In a case such as this, there can be no question of assessing any distortion or deviation from accuracy, since there is no such thing as a 'right answer'. It is the difference in responses between subjects that in itself reveals a flaw in the processing of information – the polarisation of opinion highlights the pathological nature of the underlying inferences.

Implicit and intuitive norms

In this study, and more generally in studies of 'motivated'[28] reasoning, the biases seem to reveal a kind of irrationality that goes deeper than in the research mentioned above, in the sense that the inaccurate representation they produce is a goal rather than a side effect. Subjects fail to optimise the processing of the information they receive not because of cognitive limitations that force them to use shortcuts (or 'heuristics'), but because of emotional motivations – because they want to continue to believe what they like.

However, the norms that subjects are expected to follow are rarely made explicit. Hahn and Harris[29] point out that social psychologists are generally sceptical about the use of normative models and deliberately adopt a more strictly descriptive approach. More often than not, the normative standard, or the criterion of accuracy, is not pre-existing in the experiment, but is, so to speak, inherent in the experimental design itself, allowing biases to emerge and, in some cases, to be measured. Typically, an experiment designed to show the undue influence of a particular type of information on participants' judgments will provide them with that information while making it clear that it is irrelevant – while 'neutralising'[30] it. Lord et al. simply state, without any further analysis, that "[l]ogically, one might expect mixed evidence to produce some moderation in the views expressed by opposing factions" or "[a]t worst, one might expect such inconclusive to be ignored" (p. 2099).

---

[28] Z. KUNDA, The case for motivated reasoning, *Psychological Bulletin*, 1990, 108 (3), p. 480-98.

[29] U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, p. 58.

[30] A typical example is the classic study of the fundamental error of attribution by E. E. JONES and V. A. HARRIS, The attribution of attitudes, *Journal of Experimental Social Psychology*, 1967, 3 (1), p. 1-24.

This normative statement, like the alleged neutrality of the evidence presented to the participants, is based entirely on the intuitions of the experimenters. We will see later that these can be misleading and deserve careful scrutiny.

Finally, although some of the experimental results are spectacular and seem to be in line with sociological observations (e.g. regarding opinion polarisation and the success of misinformation), it is regrettable that, as Joachim Krueger and David Funder[31] point out, no theory of inference is proposed to explain and integrate the various experimental results. As we shall see, the Bayesian approach to the psychology of reasoning proposes such a theory.

THE BAYESIAN APPROACH TO REASONING

Since the end of the 1990s, a research programme in the psychology of reasoning has been developing, particularly in the UK, with an inseparable normative and descriptive aim, embodying the psychological version of the 'Bayesian turn' that has taken place in many fields related to the analysis of reasoning, particularly in epistemology and philosophy of science. Whereas the study of human reasoning has long been dominated by a logicist paradigm, according to which rational thought is governed by logic, the Bayesian approach places probabilities, interpreted as the degrees of belief of a rational agent, at the centre of the analysis.

Bayesian preliminaries

The term 'Bayesian' derives from the name of the Reverend Thomas Bayes, the 18th-century English clergyman and mathematician who devised a theorem in probability – Bayes' theorem – but its many uses today, not only in statistics but also in the philosophy of science and cognitive science, rest on theoretical pillars that are irreducible to mathematical theory and of which there is no trace in Bayes' work.

Bayesian epistemology and Bayesian belief revision theory are based on a subjective – or 'Bayesian' – interpretation of probabilities, according to which probabilities measure degrees of belief held by agents, rather than objective properties of the world. Agents' mental states are conceived as graded beliefs – rather than categorical ones – which must obey the axioms of probability calculus. In this view, we do not believe that it is raining or that it is not raining 'outright'; we rather believe that it is raining with a degree of 0.7 (for example) and therefore (on pain of

[31] J. I. KRUEGER, D. C. FUNDER, Towards a balanced social psychology: causes, consequences, and cures for the problem-seeking approach to social behavior and cognition, p. 2.

inconsistency) that it is not raining with a degree of 0.3. So it is both a thesis about the nature of mental states and a thesis about rationality.

As a normative theory of reasoning, Bayesian epistemology is also dynamic: it states how beliefs change in the face of new evidence. This is where Bayes' theorem comes in. Its epistemological interpretation is that an agent's degree of belief in a hypothesis *H* in the face of evidence *E*, i.e. the (subjective) probability of *H* given *E*, denoted $P(H/E)$, depends on his prior belief in *H* (his degree of belief in *H* before he became aware of *E*), denoted $P(H)$, the conditional probability of *E* given *H* (the probability of *E* if *H* is true), denoted $P(E/H)$, and the probability of *E* regardless of whether *H* is true or not, denoted $P(E)$. Bayes' theorem applied to belief revision is expressed as follows:

$$P(H \mid E) = P(H) \text{ x } P(E \mid H) / P(E)$$

Insofar as $P(E)$ equals $P(H) \text{ x } P(E/H) + P(\neg H) \text{ x } P(E/\neg H)$, the ratio between the probability of the evidence if the hypothesis is true $P(E/H)$ and the probability of the evidence if the hypothesis is false $P(E/\neg H)$, known as its 'likelihood ratio', provides a measure of the confirmatory nature of the evidence for the hypothesis in question – the effect it should have on our degree of belief in that hypothesis. In short, the confirmatory value of a piece of evidence for a given hypothesis is the higher the more likely the evidence is if the hypothesis is true *and* the less likely the evidence is if the hypothesis is false. For example, a symptom will be all the more diagnostic of a disease if it is common in that disease, but also if it is rare in other cases (this is why the loss of taste and smell, although less common than fever in Covid-19 patients, is still more symptomatic).

While this model constrains the way in which we need to revise our beliefs, it also takes into account the agent-relativity of the probative value of the same information. As a result, it tells us nothing about the probative value of any piece of information 'in the absolute', and can in no way indicate a good answer that is valid for everyone. However, the Bayesian model is related to the accuracy of representations; indeed, it can be shown that following the Bayesian rules maximises the chances of an agent arriving at an accurate representation, whatever her initial beliefs.[32]

The Bayesian approach to inference has had a considerable influence in the philosophy of science, where it provides a very powerful tool for analysing the testing and confirmation of scientific hypotheses.[33] More generally, Bayesian theory makes it possible to account for many normative intuitions about belief revision, allowing, for example, a precise analysis of the way in which conclusions can be drawn on the basis of contradictory or partially

[32] See B. DE FINETTI, *Theory of probability*, New York, Wiley, 1974; H. LEITGEB, R. PETTIGREW, An objective justification of Bayesianism II: the consequences of minimizing inaccuracy, *Philosophy of Science*, 2010, 77 (2), p. 201-35.

[33] See C. HOWSON, P. URBACH, *Scientific reasoning: the Bayesian approach*, Chicago, Open Court, 1996.

incompatible information from different sources.[34] As we shall see, the probabilistic turn in the psychology of reasoning is based on the inseparable adoption of a Bayesian approach to epistemology and a psychological theory accompanied by an experimental research programme.

The probabilistic turn in the psychology of reasoning

The Bayesian programme in the psychology of reasoning rests on two inseparable major hypotheses: first, that the correct standard of both inductive and deductive reasoning is Bayesian, and second, that humans are qualitatively Bayesian. This provides a new prism through which many experimental results and observations can be reinterpreted and new ones produced.

The theory

The Bayesian programme in the psychology of reasoning conceives of rationality as the ability to reason about uncertainty, rather than as to implement deductive rules marked by certainty. In this view, human cognition is therefore more akin to the solution of inferential probabilistic problems than of logical ones.

Not only does human reason appear to be largely flawed when examined in the light of logical canons – which in itself should be taken as an incentive to rethink these canons in order to make them more appropriate to the reality of human cognition – but these canons are also largely unsuited to the nature of most of the problems we have to solve in everyday life. Replacing the logicist standard within a probabilistic approach thus allows one to reconsider many observations that have led to a pessimistic conclusion about human rationality, and to reassess the performance of agents in a much more charitable light. While, as we shall see, the Bayesian approach allows one to reinterpret certain experimental results in the social psychological literature on biases, it should be noted that the work of Mike Oaksford and Nick Chater[35] applies not only to everyday reasoning, which is essentially probabilistic due to the uncertainty of worldly events, but first and foremost to the prima facie 'logical' tasks typically proposed by psychologists of reasoning to test the deductive skills of humans. By reinterpreting conditional statements such as 'If A, then B' as probabilistic statements ('P(B | A) is high') rather than logical implications, they show that many

[34] See L. BOVENS and S. HARTMANN, *Bayesian epistemology*, Oxford, Oxford University Press, 2004.

[35] M. OAKSFORD, N. CHATER, *Bayesian rationality: the probabilistic approach to human reasoning*, Oxford, Oxford University Press, 2007; New paradigms in the psychology of reasoning, *Annual Review of Psychology*, 2020, 71, p. 305-30.

of the violations of logic demonstrated in the laboratory actually correspond to the implementation of correct probabilistic reasoning schemes that may be inappropriate to the task at hand, but are widely operational in the real world.

This reinterpretation of syllogisms in the light of conditional probabilities makes it possible to account for many normative intuitions and to evaluate the rationality of our reasoning practices. Ulrike Hahn and her colleagues have developed a normative Bayesian theory of argumentation[36], according to which the strength of an argument depends essentially on its content. This theory makes it possible to understand that different arguments in the same form can have very different strengths. Consider for example 'there is no evidence of life outside the Earth, therefore there is no life outside the Earth' on the one hand, and 'no toxic effect of this drug has been demonstrated in clinical trials, therefore this drug is safe (not toxic)' on the other. Interpreted as deductions, these two arguments are equally invalid; they are instances of the so-called 'argument from ignorance', traditionally categorised as a fallacy. However, while the first seems largely unacceptable, the second seems quite reasonable – it is, in fact, at the heart of the protocols for approving medicines. The (relative) unacceptability of the first cannot, therefore, be reduced to its logical invalidity since the second, while not logically valid, provides reasons to increase belief in the hypothesis in question. The difference in strength between the two arguments is easy to understand if we interpret them as inductive inferences, which consist in revising our degree of belief in a hypothesis *H* (the conclusion) in the light of evidence *E* (the premise). In Bayesian terms, the likelihood ratio $P(E/H) / P(E/\neg H)$ is greater in the second case (drug) than in the first (life outside the Earth). In fact, $P(E/\neg H)$, the risk of failing to detect a toxic effect in clinical trials, can be considered quite low, although of course not zero – and certainly much lower than the risk of failing to detect life beyond Earth, if such life exists. However, these evaluations essentially depend on the agent's prior beliefs, which in turn depend on the context (it may be that advances in our means of exploring space will one day make the first argument convincing) – which makes it possible to understand the relativity of the strength of an argument to its addressee, without a divergence of evaluation between subjects necessarily revealing a violation of rationality.

This normative approach to argumentation is coupled with an experimental programme designed to test whether agents follow Bayesian prescriptions.[37] This helps to show that our alleged mistakes, both in the laboratory and in everyday life, are likely to have a rational

[36] For example U. HAHN and M. OAKSFORD, A normative theory of argument strength, *Informal Logic*, 2006, 26 (1), pp. 1-24; A. CORNER and U. HAHN, Normative theories of argumentation: are some norms better than others?, *Synthese*, 2013, 190, p. 3579-610.

[37] U. HAHN, A. J. L. HARRIS, A. CORNER, Argument content and argument source: an exploration, *Informal Logic*, 2009, 29 (4), p. 337-67.

basis. Let me now take a closer look at the connection between the normative and the descriptive aspects of the Bayesian approach.

The link between the normative and the descriptive – presuming rationality

It is worth making a few remarks about the descriptive dimension of the Bayesian approach. First, by studying whether agents' responses match the predictions of Bayesian models, this approach does not claim to say much about agents' ability to solve explicitly probabilistic tasks. Just as we can observe that the cognitive processing of chromatic signals in the retina at the basis of visual perception is based on complex computations, without concluding that we are capable of performing these computations to solve mathematical problems, we can affirm that the mind is a probabilistic computational tool, without claiming that it enables us to solve problems explicitly presented in probabilistic terms.[38]

Second, Bayesian psychologists are concerned not with the nature of the inferential processes, but with the nature of the task they seek to accomplish. In David Marr's terms[39], the Bayesian approach is at the level of computation rather than implementation. The aim is to show that it is possible to interpret the behaviour of agents as conforming to Bayesian rules without saying anything about the inference processes actually implemented.

Indeed, the observation that it is possible to explain behaviour in Bayesian terms is one of the motivations for this approach: the adoption of the Bayesian norm is (partly) empirically justified. But isn't there a risk of a circular or self-justifying approach? There are several answers to this legitimate concern.

Firstly, it must be recognised that the rationality of an agent's behaviour cannot be assessed without presupposing the task she is trying to fulfil. How would it be possible to interpret a behaviour as revealing a reasoning process without adopting a model of rationality that corresponds to the correct way of solving the problem at hand? If this is seen as a risk of circularity, it is no less threatening to the logicist approach and the more descriptive approach of social psychology.

But the very possibility of describing agents' responses as following the prescriptions of a normative model should create a presumption in favour of such a model. In other words, rather than criticising the Bayesian approach for rationalising agents' behaviour 'by force', the burden of proof should be on the accusation of irrationality. In fact, the

[38] M. OAKSFORD and N. CHATER, *Bayesian rationality: the probabilistic approach to human reasoning*, Oxford, Oxford University Press, 2007.
[39] D. MARR, *Vision: a computational investigation into the human representation and processing of visual information*, New York, W. H. Freeman and Company, 1982.

endorsement of the presumption of rationality that guides the Bayesian approach is quite explicit in the works of its proponents.[40] Let me now consider a few criticisms that can be made to some studies of bias in social psychology (particularly confirmation bias) from a Bayesian perspective.

CRITICAL PERSPECTIVES ON CONFIRMATION BIAS

As Hahn and Harris point out, in order to claim the existence of cognitive biases that are more than mere occasional errors and that constitute a genuine disposition to irrationality, it is necessary to demonstrate systematic and robust – and therefore predictable – deviations from the prescriptions of a well-defined normative model.[41] To constitute a departure from rationality, such deviations must be harmful to us: they must be costly *on average*, in the sense that they must threaten the accuracy of our representations in general, or at least in a large number of situations. Indeed, an inference procedure that maximises the accuracy of our representations on average, while occasionally leading us to make mistakes, would still be optimal.[42]

From this perspective, Hahn and Harris not only question the robustness of the deviations highlighted in the social psychological literature – pointing out that the effects observed depend very much on the experimental context and can be cancelled out or even reversed in certain situations[43], which undermines the foundations of a general discourse on them – but also claim that the social psychological literature is far from having highlighted systematic deviations from any particular norm. For example, Harris and Hahn[44] show that Bayesian agents who are given a classical task designed to reveal an optimism bias, such as Weinstein's[45],

[40] They claim to follow the 'rational analysis' programme of John Anderson (1991), which aims to study the function and goals of cognitive processes on the assumption that the mind is adapted to its environment. Nevertheless, as in any other field, it is legitimate to remain vigilant as to whether or not the hypotheses inherent in the experimental set-up are ad hoc.

[41] U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, p. 68

[42] *Id.* These assertions by Hahn and Harris are based in part on their analysis of the statistical notion of bias, which corresponds to a predictable systematic deviation from accuracy that is not necessarily costly (for example, when the costs associated with false positives and negatives are unequal). Depending on how the accuracy of representations is assessed, a statistical bias may be optimal, in the sense that it may be the best procedure for maximising accuracy. See p. 60.

[43] Thus, the tendency towards unrealistic optimism gives way to a pessimistic bias depending on the nature of the events concerned (see J. R. CHAMBERS, P. D. WINDSCHITL, J. SULS, Egocentrism, event frequency, and comparative optimism: when what happens frequently is "more likely to happen to me, *Personality and Social Psychology Bulletin*, 2003, 29 (11), p. 1343-56; J. KRUGER, J. BURRUS, Egocentrism and focalism in unrealistic optimism (and pessimism), *Journal of Experimental Social Psychology*, 2004, 40 (3), p. 332-40).

[44] A. J. L. HARRIS and U. HAHN, Unrealistic optimism about future life events: a cautionary note, *Psychological Review*, 2011, 118 (1), pp. 135-54.

[45] N. D. WEINSTEIN, Unrealistic optimism about future life events, *Journal of Personality*

would themselves appear unreasonably optimistic. These discrepancies between normative predictions and the experimenter's intuitions suggest that the experimental design itself is underpinned by flawed assumptions and that making a normative standard explicit is an essential prerequisite. Finally, the Bayesian approach makes it possible to question the validity of the norms implicit in experiments that claim to demonstrate confirmation bias at the information assimilation stage: as we shall see from the example of Lord et al.'s experiments, the adoption of a Bayesian perspective suggests that the results of this type of study are not sufficient to conclude that agents are irrational.

Questioning the 'neutral evidence principle'

Lord et al. consider that the participants in their study are irrational in the sense that they appear to violate a form of impartiality according to which the mixed evidence presented to them should be 'neutral' – it should have no effect on their beliefs. But this 'neutral evidence principle', to use Jonathan Baron's[46] terms, is based on an unjustified abstraction. One of the basic tenets of Bayesianism is that the impact of evidence is not – and should not be – the same for subjects with different prior beliefs.

Not only is it normal for the beliefs of people starting from different positions to evolve differently, but the evidential value of information should also vary according to their initial beliefs. In fact, each of the 'ingredients' of Bayes' theorem is likely to be evaluated differently. Consequently, the experimenters' assumption that the information given to participants for and against the death penalty has exactly the same weight – and should cancel each other out – is not justified. This becomes particularly clear, as Hahn and Harris[47] show, when we consider the reliability of the source of the information and its relationship to the plausibility of its content.

Source reliability and content plausibility

As David Schum[48] has pointed out, both the relevance and the credibility of evidence are important in assessing its probative force for a given hypothesis. To count as evidence, a testimony must not only be relevant to the evaluation of the hypothesis in question, but it must also be credible: we must be able to accept its contents without risk.

---

and Social Psychology, 1980, 39 (5), pp. 806-20.

[46] J. BARON, *Thinking and deciding*, Cambridge, Cambridge University Press, 2008.

[47] U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, p. 90.

[48] D. SCHUM, *The evidential foundations of probabilistic reasoning*, Evanston, Northwestern University Press, 1994.

There are several dimensions to the credibility of evidence. The *plausibility* of its content is an essential component: the claim that it is raining in London is more plausible than the claim that it is 55°C in London – it is therefore more credible, all other things being equal. However, the credibility of information cannot be reduced to the plausibility of its content: the *reliability* of its source also needs to be considered. In some cases, (one's assessment of) the reliability of the source of a piece of information will completely determine (one's assessment of) its credibility. In other cases, however, the plausibility of the content will be the only indication one has of the reliability of the source: if I hear someone say that they have just seen an elephant in the street, this tells me more about the (lack of) reliability of that person than about the presence of such an animal in the street. In most cases, plausibility and credibility are part of a complex dynamic that depends on prior beliefs. As easily modelled in Bayesian terms[49], receiving a message implies simultaneously updating one's assessment of the reliability of its source and of the plausibility of its content (i.e. one's degree of belief in it). Some experimental results suggest that subjects follow Bayesian rules quite well in this area.[50]

There is no such thing as a completely reliable source. This observation must have consequences for the credibility of the information we receive from others (as is the case for almost all the information that forms the basis of our knowledge[51]); a normative model of belief revision based on testimony cannot ignore this dimension.[52] However, the implicit standards against which the inferential practices of agents are assessed in most psychology experiments neglect it; in particular, the question of the reliability of the experimenter as a source of information is very often ignored, even though it is doubtful that subjects trust (and should trust) experimenters completely.[53]

To these considerations must be added the pragmatic aspects of the source's intentions, which can blur the line between credibility and relevance: the reliability of a source (e.g. an expert or lay witness) lies not only in their ability to provide us with accurate information but also in their ability to select the information that is most relevant to the

[49] See L. BOVENS, S. HARTMANN, *Bayesian epistemology*, Oxford, Oxford University Press, 2004.
[50] See A. JARVSTAD, U. HAHN, Source reliability and the conjunction fallacy, *Cognitive Science*, 2011, 35 (4), pp. 682-711.
[51] See C. A. J. COADY, Testimony and observation, *American Philosophical Quarterly*, 1973.
[52] U. HAHN, A. J. L. HARRIS, A. CORNER, Argument content and argument source: an exploration, *Informal Logic*, 2009, 29 (4), pp. 337-67 ; U. HAHN, M. OAKSFORD, A. J. L. HARRIS, Testimony and argument: a Bayesian perspective, *in* F. Zenker (ed.), *Bayesian argumentation: the practical side of probability*, New York, Springer, 2012, pp. 15-38.
[53] Adam Corner and his colleagues suggest that this could provide an explanation for the conservatism mentioned above (A. CORNER, A. J. L. HARRIS, U. HAHN, Conservatism in belief revision and participant skepticism, *in* S. Ohlsson, R. Catrambone (eds.), *Proceedings of the 32nd annual conference of the cognitive science society*, Austin, Texas, Cognitive Science Society, 2010, 32, p. 1625-30).

question at hand.[54] This is true in everyday life, but it takes on a particular significance in an experimental context.[55]

Lord et al. acknowledge that it is natural for an agent to attribute more credibility to studies that support his preferred hypothesis, but they add that these studies should not in turn be invoked to reinforce that hypothesis.[56] However, when we look more closely at what is involved in this notion of credibility, we find that the model underlying this judgement is too simplistic. In fact, in addition to the plausibility of the information provided (in the sense that it is more or less expected under a given hypothesis), we cannot ignore the question of the reliability of the source providing it – it would be naive to assume that the experimental design of Lord et al. itself does not influence the participants' assessment of the authenticity of the studies presented to them (and rightly so, since in this case they were created from scratch by the experimenters). If we add to this some considerations about relevance and possible inferences about the intentions of the experimenters that explain the choice of these studies, there is not much to support the claim that these allegedly scientific studies should have equivalent probative value regarding the deterrent effect of the death penalty for all participants. Of course, it is far from certain that the inferences made by the participants in this experiment actually followed a Bayesian model. But given the complexity involved in any inference, there seems little justification for concluding that there was a clear and manifest violation of any normative model (and certainly not that such a violation occurred in a systematic and robust way, as would be required to establish the existence of a genuine bias).

Biased assessment of the cost of assimilation

Even if it is difficult to identify a rational norm that is systematically violated by the responses of participants in Lord et al.'s experiment – and even assuming that it is possible to account for these responses in a Bayesian model – the fact remains that the inferences they draw are harmful in the sense that they clearly lead them to increasingly inaccurate representations of reality – as evidenced by the polarisation of opinions observed at the end of the experiment. Again, this is a conclusion that must be qualified.

[54] This is an under-explored dimension of the epistemology of testimony. See M. VORMS, Relevance and testimonial reliability, in preparation; M. VORMS, Expert advice for decision-making: the subtle boundary between informing and prescribing, *in* A. Bernal and G. Axtell (eds.), *Epistemic paternalism*, Lanham, Rowman & Littlefied, 2020, p. 45-60.

[55] This has been suggested as part of the explanation for the so-called 'dilution' effect. See J. KOED MADSEN, U. HAHN, M. VORMS, The dilution effect: conversational basis and witness reliability, *CogSci*, 2017, pp. 2663-8.

[56] C. G. LORD, L. ROSS, M. R. LEPPER, Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence, p. 2106-7.

As Hahn and Harris[57] point out, it is not enough to observe that some individuals are obviously trapped in false representations to conclude that this inference process is costly on average. Olsson[58] has shown that a Bayesian model can predict both the polarisation of opinions and a general convergence of the majority. To put it intuitively, in the words of Hahn and Harris, when some participants are "more wrong" after the experiment, others "have moved their beliefs in the direction of 'the truth'. On average, accuracy may thus readily increase".[59]

What is left of confirmation bias?

All this is not to deny that there is an effect of prior beliefs on the assessment of the probative value of information. But this is by no means a bias, if by 'bias' we mean a systematic departure from rationality; it is in fact perfectly rational to assess the plausibility of information (and hence its probative value, of which plausibility is a component) in the light of its relationship to what is otherwise held to be true. Nor am I denying here that this dynamic can have damaging consequences in certain contexts. But in the absence of a clear standard against which systematic violation would be established, and insofar as a normative model that maximises the accuracy of representations in the long run seems compatible with participants' responses, there seems no reason to conclude that they are irrational. In short, at the end of this review there does not seem to be much left in favour of the existence of cognitive biases with explanatory and predictive power, which would be the clear manifestation of a form of irrationality that condemns us to move further and further away from accurate representations of reality.

The 'dead end' of explanations in terms of cognitive biases

Krueger and Funder regretted the negative turn that social psychology research has taken, pointing out that an experimental programme "designed to uncover misbehavior or cognitive failures is sure to find some", and "may be approaching a dead end", "becoming progressively less informative as it continues to proliferate, causing human strengths and cognitive skills to be underestimated and impairing the development of a theory", thus yielding "a cynical outlook on human nature rather than usable guidance for behavior and judgement".[60] By analogy with the

[57] U. HAHN, A. J. L. HARRIS, What does it mean to be biased: motivated reasoning, and rationality, p. 91.

[58] OLSSON E. J., A Bayesian simulation model of group deliberation and polarization, in F. Zenker (ed.), *Bayesian argumentation*, New York, Springer, 2013, pp. 113-33.

[59] *Id.*

[60] J. I. KRUEGER, D. C. FUNDER, Towards a balanced social psychology: causes,

study of visual illusions, they suggested that our inferential errors should be seen as reflecting processes that enable us to obtain accurate representations in most contexts, rather than as generalised failures of our cognitive system.

This is what the Bayesian approach aims to do, as I hope this overview has shown. Without denying that we often make gross mistakes, it aims to resolve the apparent paradox between this observation and the fact that, in the vast majority of situations, we are well-adapted to our environment.

I would have missed my objective if, after reading this article, one were to conclude that none of what is commonly cited as worrying departures from common sense and formidable threats to democratic life, social peace and public health exist and that all is for the best in the most rational of worlds. My aim is rather to denounce the simplification that consists of explaining apparent irrationality by a generalised tendency towards irrationality without explaining what is meant by 'rationality' and, a fortiori, providing empirical evidence for any regularity (and hence generality) of the phenomena it is supposed to explain. I have also tried to suggest, on the basis of the work of Bayesian psychologists, that it is possible to explain many phenomena by representing subjects' inferences in terms of models that conform to a rational standard. This in no way detracts from the problematic and worrying nature of some of these phenomena. But if a normative model makes it possible to account for them, the burden of proof lies with the accusation of irrationality.

REFERENCES

ANDERSON J. R., Is human cognition adaptive?, *Behavioural and Brain Sciences*, 1991, 14, p. 471-517.

BARON J., *Thinking and deciding*, Cambridge, Cambridge University Press, 2008.

BOVENS L., HARTMANN S., *Bayesian epistemology*, Oxford, Oxford University Press, 2004.

COADY C. A. J., Testimony and observation, *American Philosophical Quarterly*, 1973.

CHAMBERS J. R., WINDSCHITL P. D., SULS J., Egocentrism, event frequency, and comparative optimism: when what happens frequently is 'more likely to happen to me', *Personality and Social Psychology Bulletin*, 2003, 29 (11), p. 1343-56.

CORNER A., HARRIS A. J. L., HAHN U., Conservatism in belief revision and participant skepticism, *in* S. Ohlsson, R. Catrambone (eds.), *Proceedings of the 32nd annual conference of the cognitive science*

consequences, and cures for the problem-seeking approach to social behavior and cognition, p. 2.

society, Austin, Texas, Cognitive Science Society, 2010, 32, p. 1625-30.

CORNER A., HAHN U., Normative theories of argumentation: are some norms better than others?, *Synthese*, 2013, 190, pp. 3579-610.

DE FINETTI B., *Theory of probability*, New York, Wiley, 1974.

GILBERT D. T., KRULL D. S., MALONE P., Unbelieving the unbelievable. Some problems in the rejection of false information, *Journal of Personal and Social Psychology*, 1990, 59 (4), p. 601-13.

HAHN U., HARRIS A. J. L., What does it mean to be biased: motivated reasoning, and rationality?, *Psychology of Learning and Motivation*, 2014, 61, pp. 41-102.

HAHN U., OAKSFORD M., A normative theory of argument strength, *Informal Logic*, 2006, 26 (1), p. 1-24.

HAHN U., OAKSFORD M., The rationality of informal argumentation: a Bayesian approach to reasoning fallacies, *Psychological Review*, 2007, 114 (3), p. 704-32.

HAHN U., HARRIS A. J. L., CORNER A., Argument content and argument source: an exploration, *Informal Logic*, 2009, 29 (4), p. 337-67.

HAHN U., OAKSFORD M., HARRIS A. J. L., Testimony and argument: a Bayesian perspective, in F. Zenker (ed.), *Bayesian argumentation: the practical side of probability*, New York, Springer, 2012, pp. 15-38.

HARRIS A. J. L., HAHN U., Unrealistic optimism about future life events: a cautionary note, *Psychological Review*, 2011, 118 (1), p. 135-54.

HART W., ALBARRACÍN D., EAGLY A. H., BRECHAN I., LINDBERG M. J., MERRILL L., Feeling validated versus being correct: a meta-analysis of selective exposure to information, *Psychological Bulletin*, 2009, 135 (4), p. 555-88.

HOWSON C., URBACH P., *Scientific reasoning: the Bayesian approach*, Chicago, Open Court, 1996.

HUNEMAN P., VORMS M., Is a unified theory of conspiracy possible? *Argumenta*, 2018, 3 (2), p. 49-72.

JARVSTAD A., HAHN U., Source reliability and the conjunction fallacy, *Cognitive Science*, 2011, 35 (4), pp. 682-711.

JONES E. E., HARRIS V. A., The attribution of attitudes, *Journal of Experimental Social Psychology*, 1967, 3 (1), p. 1-24.

KAHNEMAN D., *Thinking, fast and slow*, London, Penguin, 2011.

KLAYMAN J., HA Y., Confirmation, disconfirmation, and information in hypothesis testing, *Psychological Review*, 1987, 94 (2), p. 211-28.

KOED MADSEN J., HAHN U., VORMS M., The dilution effect: conversational basis and witness reliability, *CogSci*, 2017, pp. 2663-8.

KRUEGER J. I., FUNDER D. C., Towards a balanced social psychology: causes, consequences, and cures for the problem-seeking approach to social behavior and cognition, *Behavioral and Brain Sciences*, 2004, 27 (3), p. 313-27.

KRUGER J., BURRUS J., Egocentrism and focalism in unrealistic optimism (and pessimism), *Journal of Experimental Social Psychology*, 2004, 40 (3), p. 332-40.

KUNDA Z., The case for motivated reasoning, *Psychological Bulletin*, 1990, 108 (3), p. 480-98.

LEITGEB H., PETTIGREW R., An objective justification of Bayesianism II: the consequences of minimizing inaccuracy, *Philosophy of Science*, 2010, 77 (2), p. 201-35.

LEWANDOWSKY S., ECKER U., SEIFERT C., SCHWARZ N., COOK J., Misinformation and its correction: continued influence and successful debiasing, *Psychological Science in the Public Interest*, 2012, 13 (3), p. 106-31.

LORD C. G., ROSS L., LEPPER M. R., Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence, *Journal of Personality and Social Psychology*, 1979, 37 (11), p. 2098-109.

MARR D., *Vision: a computational investigation into the human representation and processing of visual information*, New York, W. H. Freeman and Company, 1982.

NICKERSON R. S., Confirmation bias: a ubiquitous phenomenon in many guises, *Review of General Psychology*, 1998, 2, p. 175-220.

OAKSFORD M., CHATER N., A rational analysis of the selection task as optimal data selection, *Psychological Review*, 1994, 101 (4), p. 608-31.

OAKSFORD M., CHATER N., *Bayesian rationality: the probabilistic approach to human reasoning*, Oxford, Oxford University Press, 2007.

OAKSFORD M., CHATER N., "New paradigms in the psychology of reasoning", *Annual Review of Psychology*, 2020, 71, p. 305-30.

OLSSON E. J., "A Bayesian simulation model of group deliberation and polarization", *in* F. Zenker (ed.), *Bayesian argumentation*, New York, Springer, 2013, pp. 113-33.

OVER D., New paradigm psychology of reasoning, *Thinking and Reasoning*, 2009, 15 (4), p. 249-65.

PETERSON C. R., BEACH L. R., Man as an intuitive statistician, *Psychological Bulletin*, 1967, 68 (1), pp. 29-46.

POPPER K., *The logic of scientific discovery*, London, Hutchinson, 1959.
SCHUM D. A., *The evidential foundations of probabilistic reasoning*, Evanston, Northwestern University Press, 1994.

SLOVIC P., LICHTENSTEIN S., Comparison of Bayesian and regression approaches to the study of information processing in judgment, *Organizational Behavior and Human Performance*, 1971, 6 (6), p. 649-744.

TVERSKY A., KAHNEMAN D., Judgment under uncertainty: heuristics and biases, *Science*, 1974, 185 (4157), pp. 1124-31.

VORMS M., Expert advice for decision-making: the subtle boundary

between informing and prescribing, *in* A. Bernal and G. Axtell (eds), *Epistemic paternalism*, Lanham, Rowman & Littlefied, 2020, p. 45-60.

VORMS M., Relevance and testimonial reliability, in preparation.

VORMS, M. HARRIS, A.J.L., TOPF, S., HAHN, U., Plausibility matters: A challenge to Gilbert's 'Spinozan' account of belief formation, *Cognition*, 2022, 220

WASON P. C., On the failure to eliminate hypotheses in a conceptual task, *Quarterly Journal of Experimental Psychology*, 1960, 12 (3), p. 129-40.

WEINSTEIN N. D., Unrealistic optimism about future life events, *Journal of Personality and Social Psychology*, 1980, 39 (5), pp. 806-20.