BEHAVIORAL AND BRAIN SCIENCES (2007) **30**, 241–297
Printed in the United States of America
DOI: 10.1017/S0140525X07001653

Base-rate respect: From ecological rationality to dual processes

Aron K. Barbey
Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, Bethesda, MD 20892-1440
barbeya@ninds.nih.gov

Steven A. Sloman
Cognitive and Linguistics Science, Brown University, Providence, RI 02912
Steven_Sloman@brown.edu
http://www.cog.brown.edu/~sloman/

Abstract: The phenomenon of base-rate neglect has elicited much debate. One arena of debate concerns how people make judgments under conditions of uncertainty. Another more controversial arena concerns human rationality. In this target article, we attempt to unpack the perspectives in the literature on both kinds of issues and evaluate their ability to explain existing data and their conceptual coherence. From this evaluation we conclude that the best account of the data should be framed in terms of a dual-process model of judgment, which attributes base-rate neglect to associative judgment strategies that fail to adequately represent the set structure of the problem. Base-rate neglect is reduced when problems are presented in a format that affords accurate representation in terms of nested sets of individuals.

Keywords: Base-rate neglect; Bayesian reasoning; dual process theory; nested set hypothesis; probability judgment

1. Introduction

Diagnosing whether a patient has a disease, predicting whether a defendant is guilty of a crime, and other everyday as well as life-changing decisions reflect, in part, the decision-maker's subjective degree of belief in uncertain events. Intuitions about probability frequently deviate dramatically from the dictates of probability theory (e.g., Gilovich et al. 2002). One form of deviation is notorious: people's tendency to neglect base-rates in favor of specific case data. A number of theorists (e.g., Brase 2002a; Cosmides & Tooby 1996; Gigerenzer & Hoffrage 1995) have argued that such neglect reveals little more than experimenters' failure to ask about uncertainty in a form that naïve respondents can understand – specifically, in the form of a question about natural frequencies. The brunt of our argument in this target article is that this perspective is far too narrow. After surveying the theoretical perspectives on the issue, we show that both data and conceptual considerations demand that judgment be understood in terms of dual processing systems: one that is responsible for systematic error and another that is capable of reasoning not just about natural frequencies, but about relations among any kind of set representation.

Base-rate neglect has been extensively studied in the context of Bayes' theorem, which provides a normative standard for updating the probability of a hypothesis in light of new evidence. Research has evaluated the extent to which intuitive probability judgment conforms to the theorem by employing a Bayesian inference task in which the respondent is presented a word problem and has to infer the probability of a hypothesis (e.g., the presence versus absence of breast cancer) on the basis of an observation (e.g., a positive mammography). Consider the following Bayesian inference problem presented by Gigerenzer and Hoffrage (1995; adapted from Eddy 1982):

> The probability of breast cancer is 1% for a woman at age forty who participates in routine screening [base-rate]. If a woman has breast cancer, the probability is 80% that she will get a positive mammography [hit-rate]. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography [false-alarm rate]. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer? _%. (Gigerenzer & Hoffrage 1995, p. 685)

ARON BARBEY is a doctoral student of Lawrence W. Barsalou in the Cognition and Development Program at Emory University. His research addresses the cognitive and neural bases of human learning and inference. Upon graduation, he will join the Cognitive Neuroscience Section of the National Institute of Neurological Disorders and Stroke as a post-doctoral fellow under the supervision of Jordan Grafman.

STEVEN SLOMAN is Professor of Cognitive and Linguistic Sciences at Brown University. He is the author of *Causal Models: How People Think About the World and Its Alternatives* (Oxford University Press, 2005). He has also authored numerous publications in the areas of reasoning, categorization, judgment, and decision making.

© 2007 Cambridge University Press 0140-525x/07 $40.00

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

According to Bayes' theorem,[1] the probability that the patient has breast cancer given that she has a positive mammography is 7.8%. Evidence that people's judgments on this problem accord with Bayes' theorem would be consistent with the claim that the mind embodies a calculus of probability, whereas the lack of such a correspondence would demonstrate that people's judgments can be at variance with sound probabilistic principles and, as a consequence, that people can be led to make incoherent decisions (Ramsey 1964; Savage 1954). Thus, the extent to which intuitive probability judgment conforms to the normative prescriptions of Bayes' theorem has implications for the nature of human judgment (for a review of the theoretical debate on human rationality, see Stanovich 1999). In the case of Eddy's study, fewer than 5% of the respondents generated the Bayesian solution.

Early studies evaluating Bayesian inference under single-event probabilities also showed systematic deviations from Bayes' theorem. Hammerton (1973), for example, found that only 10% of the physicians tested generated the Bayesian solution, with the median response approximating the hit-rate of the test. Similarly, Casscells et al. (1978) and Eddy (1982) found that a low proportion of respondents generated the Bayesian solution: 18% in the former and 5% in the latter, with the modal response in each study corresponding to the hit-rate of the test. All of this suggests that the mind does not normally reason in a way consistent with the laws of probability theory.

1.1. Base-rate facilitation

However, this conclusion has not been drawn universally. Eddy's (1982) problem concerned a single event, the probability that a particular woman has breast cancer. In some problems, when probabilities that refer to the chances of a single event occurring (e.g., 1%) are reformulated and presented in terms of natural frequency formats (e.g., 10 out of 1,000), people more often draw probability estimates that conform to Bayes' theorem. Consider the following mammography problem presented in a natural frequency format by Gigerenzer and Hoffrage (1995):

> 10 out of every 1,000 women at age forty who participate in routine screening have breast cancer [base-rate]. 8 out of every 10 women with breast cancer will get a positive mammography [hit-rate]. 95 out of every 990 women without breast cancer will also get a positive mammography [false-alarm rate]. Here is a new representative sample of women at age forty who got a positive mammography in routine screening. How many of these women do you expect to actually have breast cancer? __ out of __ . (Gigerenzer & Hoffrage 1995, p. 688)

The proportion of responses conforming to Bayes' theorem increased by a factor of about three in this case, 46% under natural frequency formats versus 16% under a single-event probability format. The observed facilitation has motivated researchers to argue that coherent probability judgment depends on representing events in the form of natural frequencies (e.g., Brase 2002a; Cosmides & Tooby 1996; Gigerenzer & Hoffrage 1995).

Cosmides and Tooby (1996) also conducted a series of experiments that employed Bayesian inference problems that had previously elicited judgmental errors under single-event probability formats. In Experiment 1, they replicated Casscells et al. (1978), demonstrating that only 12% of their respondents produced the Bayesian answer when presented with single-event probabilities. Cosmides and Tooby then transformed the single-event probabilities into natural frequencies, resulting in a remarkably high proportion of Bayesian responses: 72% of respondents generated the Bayesian solution, supporting the authors' conclusion that Bayesian inference depends on the use of natural frequencies.

Gigerenzer (1996) explored whether physicians, who frequently assess and diagnose medical illness, would demonstrate the same pattern of judgments as that of clinically untrained college undergraduates. Consistent with the judgments drawn by college students (e.g., Gigerenzer & Hoffrage 1995), Gigerenzer found that the sample of 48 physicians tested generated the Bayesian solution in only 10% of the cases under single-event probability formats, whereas 46% did so with natural frequency formats. Physicians spent about 25% more time on the single-event probability problems, which suggests that they found these problems more difficult to solve than problems presented in a natural frequency format. Thus, the physician's judgments were consistent with those of non-physicians, suggesting that formal training in medical diagnosis does not lead to more accurate Bayesian reasoning and that natural frequencies facilitate probabilistic inference across populations.

Further studies have demonstrated that the facilitory effect of natural frequencies on Bayesian inference observed in the laboratory has the potential for improving the predictive accuracy of professionals in important real-world settings. Gigerenzer and his colleagues have shown, for example, that natural frequencies facilitate Bayesian inference in AIDS counseling (Gigerenzer et al. 1998), in the assessment of statistical information by judges (Lindsey et al. 2003), and in teaching Bayesian reasoning to college undergraduates (Kuzenhauser & Hoffrage 2002; Sedlmeier & Gigerenzer 2001). In summary, the reviewed findings demonstrate facilitation in Bayesian inference when single-event probabilities are translated into natural frequencies, consistent with the view that coherent probability judgment depends on natural frequency representations.

1.2. Theoretical accounts

Explanations of facilitation in Bayesian inference can be grouped into five types that can be arrayed along a continuum of cognitive control, from accounts that ascribe facilitation to processes that have little to do with strategic cognitive processing to those that appeal to general-purpose reasoning procedures. The five accounts we discuss can be contrasted at the coarsest level on five dimensions (see Table 1). We do not claim that theorists have consistently made these distinctions in the past, only that these distinctions are in fact appropriate ones.

A parallel taxonomy for theories of categorization can be found in Sloman et al. (in press). We briefly introduce the theoretical frameworks here. The discussion of each will be elaborated as required to reveal assumptions and derive predictions in the following sections in order to compare and contrast them.

1.2.1. Mind as Swiss army knife. Several theorists have argued that the human mind consists of a number of specialized modules (Cosmides & Tooby 1996; Gigerenzer

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Table 1. *Prerequisites for reduction of base-rate neglect according to 5 theoretical frameworks*

	Mind as Swiss army knife	Natural frequency algorithm	Natural frequency heuristic	Non-evolutionary natural frequency heuristic	Nested sets and dual processes
Cognitive impenetrability	X				
Informational encapsulation	X	X			
Appeal to evolution	X	X	X		
Cognitive process uniquely sensitive to natural frequency formats	X	X	X	X	
Transparency of nested set relations	X	X	X	X	X

Note. The prerequisites of each theory are indicated by an 'X'.

& Selten 2001). Each module is assumed to be unavailable to conscious awareness or deliberate control (i.e., cognitively impenetrable), and also assumed to be able to process only a specific type of information (i.e., informationally encapsulated; see Fodor 1983). One module in particular is designed to process natural frequencies. This module is thought to have evolved because natural frequency information is what was available to our ancestors in the environment of evolutionary adaptiveness. In this view, facilitation occurs because natural frequency data are processed by a computationally effective processing module.

Two arguments have been advanced in support of the ecological validity of natural frequency data. First, as natural frequency information is acquired, it can be "easily, immediately, and usefully incorporated with past frequency information via the use of natural sampling, which is the method of counting occurrences of events as they are encountered and storing the resulting knowledge base for possible use later" (Brase 2002b, p. 384). Second, information stored in a natural frequency format preserves the sample size of the reference class (e.g., 10 out of 1,000 women have breast cancer), and are arranged into subset relations (e.g., of the 10 women that have breast cancer, 8 are positively diagnosed) that indicate how many cases of the total sample there are in each subcategory (i.e., the base-rate, the hit-rate, and false-alarm rate). Because natural frequency formats entail the sample and effect sizes, posterior probabilities consistent with Bayes' theorem can be calculated without explicitly incorporating base-rates, thereby allowing simple calculations[2] (Kleiter 1994). Thus, proponents of this view argue that the mind has evolved to process natural frequency formats over single-event probabilities, and that, in particular, it includes a cognitive module that "maps frequentist representations of prior probabilities and likelihoods onto a frequentist representation of a posterior probability in a way that satisfies the constraints of Bayes' theorem" (Cosmides & Tooby 1996, p. 60).

Theorists who take this position uniformly motivate their hypothesis via a process of natural selection. However, the cognitive and evolutionary claims are in fact conceptually independent. The mind could consist of cognitively impenetrable and informationally encapsulated modules whether or not any or all of those modules evolved for the specific reasons offered.

1.2.2. Natural frequency algorithm. A weaker claim is that the mind includes a specific algorithm for effectively processing natural frequency information (Gigerenzer & Hoffrage 1995). Unlike the mind-as-Swiss-army-knife view, this hypothesis makes no general claim about the architecture of mind. Despite their difference in scope, however, these two theories adopt the same computational and evolutionary commitments.

Consistent with the mind-as-Swiss-army-knife view, the algorithm approach proposes that coherent probability judgment derives from a simplified form of Bayes' theorem. The proposed algorithm computes the number of cases where the hypothesis and observation co-occur, N(H and D), out of the total number of cases where the observation occurs, N(H and D) + N(not-H and D) = N(D) (Gigerenzer & Hoffrage 1995; Kleiter 1994). Because this form of Bayes' theorem expresses a simple ratio of frequencies, we refer to it as "the Ratio."

Following the mind-as-Swiss-army knife view, proponents of this approach have ascribed the origin of the Bayesian ratio to evolution. Gigerenzer and Hoffrage (1995, p. 686), for example, state

> The evolutionary argument that cognitive algorithms were designed for frequency information, acquired through natural sampling, has implications for the computations an organism needs to perform when making Bayesian inferences .... Bayesian algorithms are computationally simpler when information is encoded in a frequency format rather than a standard probability format.

As a consequence, the algorithm view predicts that "Performance on frequentist problems will satisfy some of the constraints that a calculus of probability specifies, such as Bayes' rule. This would occur because some inductive reasoning mechanisms in our cognitive architecture embody aspects of a calculus of probability" (Cosmides & Tooby 1996, p. 17).

The proposed algorithm is necessarily informationally encapsulated, as it operates on a specific information format – natural frequencies; but it is not necessarily cognitively impenetrable, as no one has claimed that other cognitive processes cannot affect or cannot use the algorithm's computations. The primary motivation for the existence of this algorithm has been computational (Gigerenzer & Hoffrage 1995; Kleiter 1994). As reviewed above, the value of natural frequencies is that these

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

formats entail the sample and effect sizes and, as a consequence, simplify the calculation of Bayes' theorem: Probability judgments are coherent with Bayesian prescriptions even without explicit consideration of base-rates.

1.2.3. Natural frequency heuristic. A claim which puts facilitation under more cognitive control is that people use heuristics to make judgments (Gigerenzer & Selten 2001; Tversky & Kahneman 1974) and that the Ratio is one such heuristic (Gigerenzer et al. 1999). According to this view, "heuristics can perform as well, or better, than algorithms that involve complex computations .... The astonishingly high accuracy of these heuristics indicates their ecological rationality; fast and frugal heuristics exploit the statistical structure of the environment, and they are adapted to this structure" (Gigerenzer 2006). Advocates of this approach motivate the proposed heuristic by pointing to the ecological validity of natural frequency formats, as Gigerenzer further states (p. 52):

> To evaluate the performance of the human mind, one needs to look at its environment and, in particular, the external representation of the information. For most of the time during which the human mind evolved, information was encountered in the form of natural frequencies ...

Thus, this view proposes that the mind evolved to process natural frequencies and that this evolutionary adaptation gave rise to the proposed heuristic that computes the Bayesian Ratio from natural frequencies.

1.2.4. Non-evolutionary natural frequency heuristic. Evolutionary arguments about the ecological validity of natural frequency representations provide part of the motivation for the preceding theories. In particular, proponents of the theories argue that throughout the course of human evolution natural frequencies were acquired via natural sampling (i.e., encoding event frequencies as they are encountered, and storing them in the appropriate reference class).

In contrast, the non-evolutionary natural frequency theory proposes that natural sampling is not necessarily an evolved procedure for encoding statistical regularities in the environment, but rather, a useful sampling method that, one way or another, people can appreciate and use. The natural frequency representations that result from natural sampling, on this view, simplify the calculation of Bayes' theorem and, as a consequence, facilitate Bayesian inference (Kleiter 1994). Thus, this non-evolutionary view differs from the preceding accounts by resting on a purely computational argument that is independent of any commitments as to which cognitive processes have been selected for by evolution.

This theory proposes that the computational simplicity afforded by natural frequencies gives rise to a heuristic that computes the Bayesian Ratio from natural frequencies. The proposed heuristic implies a higher degree of cognitive control than the preceding modular algorithms.

1.2.5. Nested sets and dual processes. The most extreme departure from the modular view claims that facilitation is a product of general-purpose reasoning processes (Evans et al. 2000; Fox & Levav 2004; Girotto & Gonzales 2001; Johnson-Laird et al. 1999; Kahneman & Frederick 2002; 2005; Over 2003; Reyna 1991; Sloman et al. 2003). In this view, people use two systems to reason (Evans & Over 1996; Kahneman & Frederick 2002; 2005; Reyna & Brainerd 1994; Sloman 1996a; Stanovich & West 2000), often called Systems 1 and 2. But in an effort to use more expressive labels, we will employ Sloman's terms "associative" and "rule-based."

The dual-process model attributes responses based on associative principles like similarity or retrieval from memory to a primitive associative judgment system. It attributes responses based on more deliberative processing that involves working memory, such as the elementary set operations that respect the logic of set inclusion and facilitate Bayesian inference, to a second rule-based system. Judgmental errors produced by cognitive heuristics are generated by associative processes, whereas the induction of a representation of category instances that makes nested set relations transparent also induces use of rules about elementary set operations – operations of the sort perhaps described by Fox and Levav (2004) or Johnson-Laird et al. (1999).

According to this theory, base-rate neglect results from associative responding and facilitation occurs when people correctly use rules to make the inference. Rule-based inference is more cognitively demanding than associative inference, and is therefore more likely to occur when participants have more time, more incentives, or more external aids to make a judgment and are under fewer other demands at the moment of judgment. It is also more likely for people who have greater skill in employing the relevant rules. This last prediction is supported by Stanovich and West (2000) who find correlations between intelligence and use of base rates.

Rules are effective devices for solving a problem to the extent that the problem is represented in a way compatible with the rules. For example, long division is an effective method for solving division problems, but only if numbers are represented using Arabic numerals; division with Roman numerals requires different rules. By analogy, this view proposes that natural frequencies facilitate use of base-rates because the rules people have access to and are able to use to solve the specific kind of problem studied in the base-rate neglect literature are more compatible with natural frequency formats than single-event probability formats.

Specifically, people are adept at using rules consisting of simple elementary set operations. But these operations are only applicable when problems are represented in terms of sets, as opposed to single events (Reyna 1991; Reyna & Brainerd 1995). According to this view, facilitation in Bayesian inference occurs under natural frequencies because these formats are an effective cue to the representation of the set structure underlying a Bayesian inference problem. This is the nested sets hypothesis of Tversky and Kahneman (1983). In this framework, natural frequency formats prompt the respondent to adopt an outside view by inducing a representation of category instances (e.g., 10 out of 1,000 women have breast cancer) that reveals the set structure of the problem and makes the nested set relations transparent for problem solving.[3] We refer to this hypothesis as the nested sets theory (Ayton & Wright 1994; Evans et al. 2000; Fox & Levav 2004; Girotto & Gonzalez 2001; 2002; Johnson-Laird et al. 1999; Reyna 1991; Tversky & Kahneman 1983; Macchi 2000; Mellers & McGraw 1999; Sloman et al. 2003). Unlike

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

the other theories, it predicts that facilitation should be observable in a variety of different tasks, not just posterior probability problems, when nested set relations are made transparent.

2. Overview of empirical and conceptual issues reviewed

We now turn to an evaluation of these five theoretical frameworks. We evaluate a range of empirical and conceptual issues that bear on the validity of these frameworks.

2.1. Review of empirical literature

The theories are evaluated with respect to the empirical predictions summarized in Table 2. The predictions of each theory derive from (1) the degree of cognitive control attributed to probability judgment (see Table 1), and (2) the proposed cognitive operations that underlie estimates of probability.

Theories that adopt a low degree of cognitive control – proposing cognitively impenetrable modules or informationally encapsulated algorithms – restrict Bayesian inference to contexts that satisfy the assumptions of the processing module or algorithm. In contrast, theories that adopt a high degree of cognitive control – appealing to a natural frequency heuristic or a domain general capacity to perform set operations – predict Bayesian inference in a wider range of contexts. The latter theories are distinguished from one another in terms of the cognitive operations they propose: The evolutionary and non-evolutionary natural frequency heuristics depend on structural features of the problem, such as question form and reference class. They imply the accurate encoding and comprehension of natural frequencies and an accurate weighting of the encoded event frequencies to calculate the Bayesian ratio. In contrast, the nested sets theory does not rely on natural frequencies and, instead, predicts facilitation in Bayesian inference, and in a range of other deductive and inductive reasoning tasks, when the set structure of the problem is made transparent, thereby promoting use of elementary set operations and inferences about the logical (i.e., extensional) properties they entail.

2.2. Information format and judgment domain

The preceding review of the literature found that natural frequency formats consistently reduced base-rate neglect relative to probability formats. However, the size of this effect varied considerably across studies (see Table 3).

Cosmides and Tooby (1996), for example, observed a 60-point % difference between the proportions of Bayesian responses under natural frequencies versus single-event probabilities, whereas Gigerenzer and Hoffrage (1995)

Table 2. *Empirical predictions of the five theoretical frameworks*

	Mind as Swiss army knife	Natural frequency algorithm	Natural frequency heuristic	Non-evolutionary natural frequency heuristic	Nested sets and dual processes
Facilitation with natural frequencies (information format and judgment domain)	X	X	X	X	X
Facilitation with questions that prompt the respondent to compute the Bayesian ratio (question form)			X	X	X
Facilitation with statistical information organized in a partitive structure (reference class)			X	X	X
Facilitation with diagrammatic representations that highlight the set structure of the problem			X	X	X
Inaccurate frequency judgments					X
Equivalent comprehension of natural frequencies and single-event probabilities					X
Non-normative weighting of likelihood ratio and prior odds					X
Facilitation with set representations in deductive and inductive reasoning					X

Note. The predictions of each theory are indicated by an 'X.'

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Table 3. *Percent correct for Bayesian inference problems reported in the literature (sample sizes in parentheses)*

	Information format and judgment domain	
Study	Probability	Frequency
Casscells et al. (1978)	18 (60)	—
Cosmides & Tooby (1996; Exp. 2)	12 (25)	72 (25)
Eddy (1988)	5 (100)	—
Evans et al. (2000; Exp. 1)	24 (42)	35 (43)‡
Gigerenzer (1996)	10 (48)	46 (48)
Gigerenzer & Hoffrage (1995)	16 (30)	46 (30)
Macchi (2000)	6 (30)	40 (30)
Sloman et al. (2003; Exp.1)	20 (25)	51 (45)
Sloman et al. (2003; Exp. 1b)	—	31 (48)‡

Note. Probability problems require that the respondent compute a conditional-event probability from data presented in a non-partitive form, whereas frequency problems include questions that prompt the respondent to evaluate the two terms of the Bayesian ratio and present data that is partitioned into these components.
‡ $p > 0.05$

reported a difference only half that size. The wide variability in the size of the effects makes it clear that in no sense do natural frequencies eliminate base-rate neglect, though they do reduce it.

Sloman et al. (2003) conducted a series of experiments that attempted to replicate the effect sizes observed by the previous studies (e.g., Cosmides & Tooby 1996; Experiment 2, Condition 1). Although Sloman et al. found facilitation with natural frequencies, the size of the effect was smaller than that observed by Cosmides and Tooby: The percent of Bayesian solutions generated under single-event probabilities (20%) was comparable to Cosmides and Tooby (12%), but the percentage of Bayesian answers generated under natural frequencies was smaller (i.e., 72% versus 51% for Sloman et al.). In a further replication, Sloman et al. found that only 31% of their respondents generated the Bayesian solution, a statistically non-significant advantage for natural frequencies.

Evans et al. (2000, Experiment 1) similarly found only a small effect of information format. They report 24% Bayesian solutions under single-event probabilities and 35% under natural frequencies, a difference that was not reliable.

Brase et al. (2006) examined whether methodological factors contribute to the observed variability in effect size. They identified two factors that modulate the facilitory effect of natural frequencies in Bayesian inference: (1) the academic selectivity of the university the participants attend, and (2) whether or not the experiment offered a monetary incentive for participation. Experiments whose participants attended a top-tier national university and were paid reported a significantly higher proportion of Bayesian responses (e.g., Cosmides & Tooby 1996) than experiments whose participants attended a second-tier regional university and were not paid (e.g., Brase et al. 2006, Experiments 3 and 4). These results suggest that a higher proportion of Bayesian responses is observed in experiments that (a) select participants with a higher level of general intelligence, as indexed by the academic selectivity of the university the participant attends (Stanovich & West 1998a), and (b) increase motivation by providing a monetary incentive. The former observation is consistent with the view that Bayesian inference depends on domain general cognitive processes to the degree that intelligence is domain general. The latter suggests that Bayesian inference is strategic, and not supported by automatic (e.g., modularized) reasoning processes.

2.3. Question form

One methodological factor that may mediate the effect of problem format is the form of the Bayesian inference question presented to participants (Girotto & Gonzalez 2001). The Bayesian solution expresses the ratio between the size of the subset of cases in which the hypothesis and observation co-occur and the total number of observations. Thus, it follows that the respondent should be more likely to arrive at this solution when prompted to adopt an outside view by utilizing the sample of category instances presented in the problem (e.g., "Here is a new sample of patients who have obtained a positive test result in routine screening. How many of these patients do you expect to actually have the disease? __ out of __") versus a question that presents information about category properties (e.g., "... Pierre has a positive reaction to the test ...") and prompts the respondent to adopt an inside view by considering the fact about Pierre to compute a probability estimate. As a result, the form of the question should modulate the observed facilitation.

In the preceding studies, however, information format and judgment domain were confounded with question form: Only problems that presented natural frequencies prompted use of the sample of category instances presented in the problem to compute the two terms of the Bayesian solution, whereas single-event probability problems prompted the use of category properties to compute a conditional probability.

To dissociate these factors, Girotto and Gonzalez (2001) proposed that single-event probabilities (e.g., 1%) can be represented as chances[4] (e.g., "One chance out of 100"). Under the chance formulation of probability, the respondent can be asked either for the standard conditional probability or for values that correspond more closely to the ratio expressed by Bayes' theorem. The latter question asks the respondent to evaluate the chances that Pierre has a positive test for a particular infection, out of the total chances that Pierre has a positive test, thereby prompting consideration of the chances that Pierre – who could be anyone with a positive test in the sample – has the infection. In addition to encouraging an outside view by prompting the respondent to represent the sample of category instances presented in the problem, this question prompts the computation of the Bayesian ratio in two clearly defined steps: First calculate the overall number of chances where the conditioning event is observed, then compare this quantity to the number of chances where the conditioning event is observed in the presence of the hypothesis.

To evaluate the role of question form in Bayesian inference, Girotto and Gonzalez (2001, Study 1) conducted an experiment that manipulated question form

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

independently of information format and judgment domain. The authors presented the following Bayesian inference scenario to 80 college undergraduates of the University of Provence, France:

> A person who was tested had 4 chances out of 100 of having the infection. 3 of the 4 chances of having the infection were associated with a positive reaction to the test. 12 of the remaining 96 chances of not having the infection were also associated with a positive reaction to the test (Girotto & Gonzalez 2001, p. 253).

Half of the respondents were then asked to compute a conditional probability (i.e., "If Pierre has a positive reaction, there will be __ chance(s) out of __ that the infection is associated with his positive reaction"), whereas the remaining respondents were asked to evaluate the ratio of probabilities expressed in the Bayesian solution (i.e., "Imagine that Pierre is tested now. Out of the total 100 chances, Pierre has __ chances of having a positive reaction, __ of which will be associated with having the infection").

Girotto and Gonzalez (2001) found that only 8% of the respondents generated the Bayesian solution when asked to compute a conditional probability, consistent with the earlier literature. But the proportion of Bayesian answers increased to 43% when the question prompted the respondent to evaluate the two terms of the Bayesian solution. The same pattern was observed with the natural frequency format problem. Only 18% of the respondents generated the Bayesian solution when asked to compute a conditional frequency, whereas this proportion increased to 58% when asked to evaluate the two terms separately. This level of performance is comparable to that observed under standard natural frequency formats (e.g., Gigerenzer & Hoffrage 1995), and supports Girotto and Gonzalez's claim that the two-step question approximates the question asked with standard natural frequency formats. In further support of Girotto and Gonzalez's predictions, there were no reliable effects of information format or judgment domain across all the reported comparisons.

These findings suggest that people are not predisposed against using single-event probabilities but instead appear to be highly sensitive to the form of the question: When asked to reason about category instances to compute the two terms of the Bayesian ratio, respondents were able to draw the normative solution under single-event probabilities. Facilitation in Bayesian inference under natural frequencies need not imply that the mind is designed to process these formats, but instead can be attributed to the facilitory effect of prompting use of the sample of category instances presented in the problem to evaluate the two terms of the Bayesian ratio.

2.4. Reference class

To assess the role of problem structure in Bayesian inference, we review studies that have manipulated structural features of the problem. Girotto and Gonzalez (2001) report two experiments that systematically assess performance under different partitionings of the data: Defective frequency partitions and non-partitive frequency problems. Consider the following medical diagnosis problem, which presents natural frequencies under what Girotto and Gonzalez (2001, Study 5) term a defective partition:

> 4 out of 100 people tested were infected. 3 of the 4 infected people had a positive reaction to the test. 84 of the 96 uninfected people did not have a positive reaction to the test. Imagine that a group of people is now tested. In a group of 100 people, one can expect __ individuals to have a positive reaction, __ of whom will have the infection.

In contrast to the standard partitioning of the data under natural frequencies, here the frequency of uninfected people who did not have a positive reaction to the test is reported, instead of the frequency of uninfected, positive reactions. As a result, to derive the Bayesian solution, the first value must be subtracted from the total population of uninfected individuals to obtain the desired value ($96 - 84 = 12$), and the result can be used to determine the proportion of infected, positive people out of the total number of people who obtain a positive test (e.g., $3/15 = 0.2$). Although this problem exhibits a partitive structure, Girotto and Gonzalez predicted that the defective partitioning of the data would produce a greater proportion of errors than observed under the standard data partitioning, because the former requires an additional computation. Consistent with this prediction, only 35% of respondents generated the Bayesian solution, whereas 53% did so under the standard data partitioning. Nested set relations were more likely to facilitate Bayesian reasoning when the data were partitioned into the components that are needed to generate the solution.

Girotto and Gonzalez (2001, Study 6) also assessed performance under natural frequency formats that were not partitioned into nested set relations (i.e., unpartitioned frequencies). As in the case of standard natural frequency format problems (e.g., Cosmides & Tooby 1996), these multiple-sample problems employed natural frequencies and prompted the respondent to compute the two terms of the Bayesian solution.[5] Such a problem must be treated in the same way as a single-event probability problem (i.e., using the conditional probability and additivity laws) to determine the two terms of the Bayesian solution. Girotto and Gonzalez therefore predicted that performance under multiple samples would be poor, approximating the performance observed under standard probability problems. As predicted, none of the respondents generated the Bayesian solution under the multiple sample or standard single-event probability frames. Natural frequency formats facilitate Bayesian inference only when they partition the data into components needed to draw the Bayesian solution.

Converging evidence is provided by Macchi (2000), who presented Bayesian inference problems in either a partitive or non-partitive form. Macchi found that only 3% of respondents generated the Bayesian solution when asked to evaluate the two terms of the Bayesian ratio with non-partitive frequency problems. Similarly, only 6% of the respondents generated the Bayesian solution when asked to compute a conditional probability under non-partitive probability formats (see also Sloman et al. 2003, Experiment 4). But when presented under a partitive formulation and asked to evaluate the two terms of the Bayesian ratio the proportions increased to 40% under partitive natural frequency formats, 33% under partitive single-event probabilities, and 36% under the modified partitive single-event probability problems. The findings reinforce the nested sets view that information structure is the factor determining predictive accuracy.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

To further explore the contribution of information structure and question form in Bayesian inference, Sloman et al. (2003) assessed performance using a conditional chance question. In contrast to the standard conditional probability question that presents information about a particular individual (e.g., "Pierre has a positive reaction to the test"), their conditional probability question asked the respondent to evaluate "the chance that *a person* found to have a positive test result actually has the disease." This question requests the probability of an unknown category instance and therefore prompts the respondent to consult the data presented in the problem to assess the probability that this person – who could be any randomly chosen person with a positive result in the sample – has the disease. In Experiment 1, Sloman et al. looked for facilitation in Bayesian inference on a partitive single-event probability problem by prompting use of the sample of category instances presented in the problem to compute a conditional probability, as the nested sets hypothesis predicts. Forty-eight percent of the 48 respondents tested generated the Bayesian solution, demonstrating that making partitive structure transparent facilitates Bayesian inference.

In summary, the reviewed findings suggest that when the data are partitioned into the components needed to arrive at the solution and participants are prompted to use the sample of category instances in the problem to compute the two terms of the Bayesian ratio, the respondent is more likely to (1) understand the question, (2) see the underlying nested set structure by partitioning the data into exhaustive subsets, and (3) select the pieces of evidence that are needed for the solution. According to the nested sets theory, accurate probability judgments derive from the ability to perform elementary set operations whose computations are facilitated by external cues (for recent developmental evidence, see Girotto & Gonzalez, in press).

2.5. Diagrammatic representations

Sloman et al. (2003, Experiment 2) explored whether Euler circles, which were employed to construct a nested set structure for standard non-partitive single-event probability problems (e.g., Cosmides & Tooby 1996), would facilitate Bayesian inference (see Fig. 1 here). These authors found that 48% of the 25 respondents tested generated the Bayesian solution when presented non-partitive single-event probability problems with an Euler diagram that depicted the underlying nested set relations. This finding demonstrates that the set structure of standard non-partitive single-event probability problems can be represented by Euler diagrams to produce facilitation. Supporting data can be found in Yamagishi (2003) who used diagrams to make nested set relations transparent in other inductive reasoning problems. Similar evidence is provided by Bauer and Johnson-Laird (1993) in the context of deductive reasoning.

2.6. Accuracy of frequency judgments

Theories based on natural frequency representations (i.e., the mind-as-Swiss-army-knife, natural frequency algorithm, natural frequency heuristic, and non-evolutionary natural frequency heuristic theories) propose that "the mind is a frequency monitoring device" and that the cognitive algorithm that computes the Bayesian ratio encodes and processes event frequencies in naturalistic settings (Gigerenzer 1993, p. 300). The literature that evaluates the encoding and retrieval of event frequencies is large and extensive and includes assessments of frequency judgments under well-controlled laboratory settings based on relatively simple and distinct stimuli (e.g., letters, pairs of letters, or words), and naturalistic settings in which respondents report the frequency of their own behaviors (e.g., the medical diagnosis of patients). Laboratory studies tend to find that frequency judgments are surprisingly accurate (for a recent review, see Zacks & Hasher 2002), whereas naturalistic studies often find systematic errors in frequency judgments (see Bradburn et al. 1987). Recent efforts have been made to integrate these findings under a unified theoretical framework (e.g., Schwartz & Sudman 1994; Schwartz & Wanke 2002; Sedlmeier & Betsch 2002).

Are frequency judgments relatively accurate under the naturalistic settings described by standard Bayesian inference problems? Bayesian inference problems tend to involve hypothetical situations that, if real, would be based on autobiographical memories encoded under naturalistic conditions, such as the standard medical diagnosis problem in which a particular set of patients is hypothetically encountered (cf. Sloman & Over 2003). Hence, the present review focuses on the accuracy of frequency judgments for the autobiographical events alluded to by standard Bayesian inference problems (see sects. 2.1, 2.2, and 2.3) to assess whether Bayesian inference depends on the accurate encoding of autobiographical events.

Gluck and Bower (1988) conducted an experiment that employed a learning paradigm to assess the accuracy of frequency judgments in medical diagnosis. The respondents in the experiment learned to diagnose a rare (25%) or a common (75%) disease on the basis of four potential symptoms exhibited by the patient (e.g., stomach cramps, discolored gums). During the learning phase, the respondents diagnosed 250 hypothetical patients and in each case were provided feedback on the accuracy of their diagnosis. After the learning phase, the respondents estimated the relative frequency of patients who had the diseases given each symptom. Gluck and Bower found that relative frequency estimates of the disease were determined by the diagnosticity of the symptom (the degree to which the respondent perceived that the symptom provided useful information in diagnosing the disease) and not the base-rate frequencies of the disease. These findings were replicated by Estes et al. (1989, Experiment 1) and Nosofsky et al. (1992, Experiment 1).

Bradburn et al. (1987) evaluated the accuracy of autobiographical memory for event frequencies by employing a range of surveys that assessed quantitative facts, such as "During the last two weeks, on days when you drank liquor, about how many drinks did you have?" These questions require the simple recall of quantitative facts, in which the respondent "counts up how many individuals fall within each category" (Cosmides & Tooby 1996, p. 60). Recalling the frequency of drinks consumed over the last two weeks, for example, is based on counting the total number of individual drinking occasions stored in memory.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

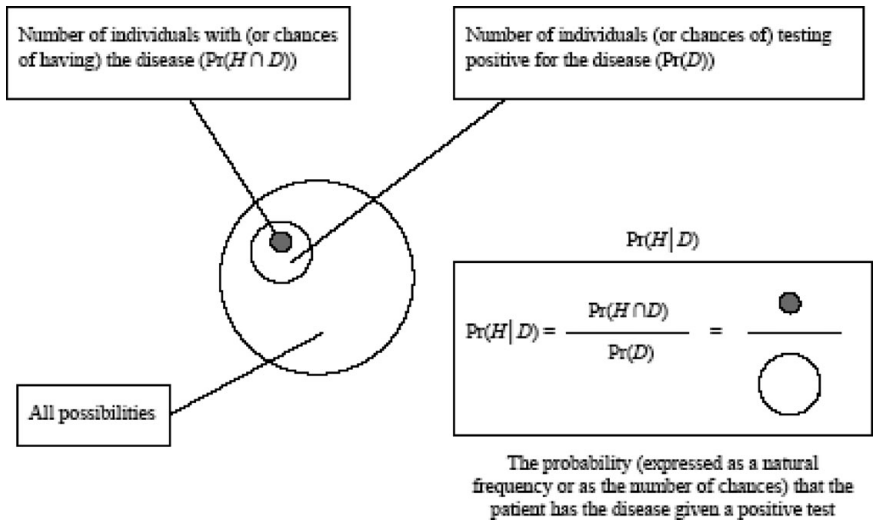


Figure 1. A diagrammatic representation of Bayes theorem: Euler circles (Sloman et al., 2003).

Bradburn et al. (1987) found that autobiographical memory for event frequencies exhibits systematic errors characterized by (a) the failure to recall the entire event or the loss of details associated with a particular event (e.g., Linton 1975; Wagenaar 1986), (b) the combining of similar distinct events into a single generalized memory (e.g., Linton 1975; 1982), or (c) the inclusion of events that did not occur within the reference period specified in the question (e.g., Pillemer et al. 1986). As a result, Bradburn et al. propose that the observed frequency judgments do not reflect the accurate encoding of event frequencies, but instead entail a more complex inferential process that typically operates on the basis of incomplete, fragmentary memories that do not preserve base-rate frequencies.

These findings suggest that the observed facilitation in Bayesian inference under natural frequencies cannot be explained by an (evolved) capacity to encode natural frequencies. Apparently, people don't have that capacity.

2.7. Comprehension of formats

Advocates of the nested sets view have argued that the facilitation of Bayesian inference under natural frequencies can be fully explained via elementary set operations that deliver the same result as Bayes' theorem, without appealing to (an evolved) capacity to process natural frequencies (e.g., Johnson-Laird et al. 1999). The question therefore arises whether the ease of processing natural frequencies goes beyond the reduction in computational complexity of Bayes' theorem that they provide (Brase 2002a). To assess this issue, we review evidence that evaluates whether natural frequencies are understood more easily than single-event probabilities.

Brase (2002b) conducted a series of experiments to evaluate the relative clarity and ease of understanding a range of statistical formats, including natural frequencies (e.g., 1 out of 10) and percentages (e.g., 10%). Brase distinguished natural frequencies that have a natural sampling structure (e.g., 1 out 10 have the property, 9 out of 10 do not) from "simple frequencies" that refer to single numerical relations (e.g., 1 out of 10 have the property). This distinction, however, is not entirely consistent with the literature, as natural frequency theorists have often used single numerical statements for binary hypotheses to express natural frequencies (e.g., Zue & Gigerenzer 2006). In any case, for binary hypotheses the natural sampling structure can be directly inferred from simple frequencies. If we observe, for example, that I win the weekly poker game "1 out of 10 nights," we can infer that I lose "9 out of 10 nights" and construct a natural sampling structure that represents the size of the reference class and is arranged into subset relations. Thus, single numerical statements of this type have a natural sampling structure, and, therefore, we refer to Brase's "simple frequencies" as natural frequencies in the following discussion.

Percentages express single-event probabilities in that they are normalized to an arbitrary reference class (e.g., 100) and can refer to the likelihood of a single-event (Brase 2002b; Gigerenzer & Hoffrage 1995). We therefore examine whether natural frequencies are understood more easily and have a greater impact on judgment than percentages.

To test this prediction, Brase (2002b, Experiment 1) assessed the relative clarity of statistical information presented in a natural frequency format versus percentage format at small, intermediate, and large magnitudes. Respondents received four statements in one statistical format, each statement at a different magnitude, and rated the clarity, impressiveness, and "monetary pull" of the presented statistics according to a 5-point scale. Example questions are shown in Table 4.

Brase (2002b) found that across all statements and magnitudes both natural frequencies and percentages were rated as "Very Clear," with average ratings of 3.98 and 3.89, respectively. These ratings were not reliably different, demonstrating that percentages are perceived as clearly and are as understandable as natural frequencies. Furthermore, Brase found no reliable differences in the impressiveness ratings (from question 2) of natural frequencies and percentages at intermediate and large

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Table 4. *Example questions presented by Brase (2002b)*

Statement
It is estimated that by the year 2020, one of every 100 Americans will have been exposed to Flu strain X [natural frequency format of low magnitude]
It is estimated that by the year 2020, 33% of all Americans will have been exposed to Flu strain X [single-event probability of intermediate magnitude]
Questions
1. How clear and easy to understand is the statistical information presented in the above sentence? [Clarity rating]
2. How serious do you think the existence of virus X is [Impressiveness rating]
3. If you were in charge of the annual budget for the U.S. Department of Health, how much of every $100 would you dedicate to dealing with virus X? __ out of every $100 [Monetary pull rating]

statistical magnitudes, suggesting that these formats are typically viewed as equally impressive. A significant difference between these formats was observed, however, at low statistical magnitudes: On average, natural frequencies were rated as "Impressive," whereas percentages were viewed as "Fairly Impressive." The observed difference in the impressiveness ratings at low statistical magnitudes did not accord with the respondent's monetary pull ratings – their willingness to allocate funds to support research studying the issue at hand – which were approximately equal for the two formats across all statements and magnitudes. Hence the difference in the impressiveness ratings at low magnitudes does not denote differences in people's willingness to act.

These data are consistent with the conclusion that percentages and natural frequency formats (a) are perceived equally clearly and are equally understandable; (b) are typically viewed as equally impressive (i.e., at intermediate and large statistical magnitudes); and (c) have the same degree of impact on behavior. Natural frequency formats do apparently increase the perceptual contrast of small differences. Overall, however, the two formats are perceived similarly, suggesting that the mind is not designed to process natural frequency formats over single-event probabilities.

2.8. Are base-rates and likelihood ratios equally weighted?

Does the facilitation of Bayesian inference under natural frequencies entail that the mind naturally incorporates this information according to Bayes' theorem, or that elementary set operations can be readily computed from problems that are structured in a partitive form? Natural frequencies preserve the sample size of the reference class and are arranged into subset relations that preserve the base-rates. As a result, judgments based on these formats will entail the sample and effect sizes; the respondent need not calculate them. To assess whether the cognitive operations that underlie Bayesian inference are consistent with the application of Bayes' theorem, studies that evaluate how the respondent derives Bayesian solutions are reviewed.

Griffin and Buehler (1999) employed the classic lawyer-engineer paradigm developed by Kahneman and Tversky (1973), involving personality descriptions randomly drawn from a population of either 70 engineers and 30 lawyers or 30 engineers and 70 lawyers. Participants' task in this study is to predict whether the description was taken from an engineer or a lawyer (e.g., "My probability that this man is one of the engineers in this sample is __%"). Kahneman and Tversky's original findings demonstrated that the respondent consistently relied upon category properties (i.e., how representative the personality description is of an engineer or a lawyer) to guide their judgment, without fully incorporating information about the population base-rates (for a review, see Koehler 1996). However, when the base-rates were presented via a counting procedure that induces a frequentist representation of each population and the respondent is asked to generate a natural frequency prediction (e.g., "I would expect that __ out of the 10 descriptions would be engineers"), base-rate usage increased (Gigerenzer et al. 1988).

To assess whether the observed increase in base-rate usage reflects the operation of a Bayesian algorithm that is designed to process natural frequencies, Griffin and Buehler (1999) evaluated whether participants derived the solution by utilizing event frequencies according to Bayes' theorem. This was accomplished by first collecting estimates of each of the components of Bayes' theorem in odds form[6]: Respondents estimated (a) the probability that the personality description was taken from the population of engineers or lawyers; (b) the degree to which the personality description was representative of these populations; and (c) the perceived population base-rates. Each of these estimates was then divided by their compliment to yield the posterior odds, likelihood ratio, and prior odds, respectively. Theories based on the Bayesian ratio predict that under frequentist representations, the likelihood ratios and prior odds will be weighted equally (Griffin & Buehler 1999).

Griffin and Buehler evaluated this prediction by conducting a regression analysis using the respondent's estimated likelihood ratios and prior odds to predict their posterior probability judgments (cf. Keren & Thijs 1996). Consistent with the observed increase in base-rate usage under frequentist representations (Gigerenzer et al. 1988), Griffin and Buehler (1999, Experiment 3b) found that the prior odds (i.e., the base-rates) were weighted more heavily than the likelihood ratios, with corresponding regression weights (β values) of 0.62 and 0.39. The failure to weight them equally violates Bayes' theorem. Although frequentist representations may enhance base-rate usage, they apparently do not induce the operation of a mental analogue of Bayes' theorem.

Further support for this conclusion is provided by Evans et al. (2002) who conducted a series of experiments demonstrating that probability judgments do not reflect equal weighting of the prior odds and likelihood ratio. Evans et al. (2002, Experiment 5) employed a paradigm that extended the classic lawyer-engineer experiments by assessing Bayesian inference under conditions where the base-rates are supplied by commonly held beliefs and

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

only the likelihood ratios are explicitly provided. These authors found that when prior beliefs about the base-rate probabilities were rated immediately before the presentation of the problem, the prior odds (i.e., the base-rates) were weighted more heavily than the likelihood ratios, with corresponding regression weights (β values) of 0.43 and 0.19.

Additional evidence supporting this conclusion is provided by Kleiter et al. (1997) who found that participants assessing event frequencies in a medical diagnosis setting employed statistical evidence that is irrelevant to the calculation of Bayes' theorem. Kleiter et al. (1997, Experiment 1) presented a list of event frequencies to respondents, which included those that were necessary for the calculation of Bayes' theorem (e.g., $Pr(D \mid H)$) and other statistics that were irrelevant (e.g., $Pr(\sim D)$). Participants were then asked to identify the event frequencies that were needed to diagnose the probability of the disease, given the symptom (i.e., the posterior probability). Of the four college faculty and 26 graduate students tested, only three people made the optimal selection by identifying only the event frequencies required to calculate Bayes' theorem.

These data suggest that the mind does not utilize a Bayesian algorithm that "maps frequentist representations of prior probabilities and likelihoods onto a frequentist representation of a posterior probability in a way that satisfies the constraints of Bayes' theorem" (Cosmides & Tooby 1996, p. 60). Importantly, the findings that the prior odds and likelihood ratio are not equally weighted according to Bayes' theorem (Evans et al. 2002; Griffin & Buehler 1999) imply that Bayesian inference does not rely on Bayesian computations per se.

Thus, the findings are inconsistent with the mind-as-Swiss-army-knife, natural frequency algorithm, natural frequency heuristic, and non-evolutionary natural frequency heuristic theories, which propose that coherent probability judgment reflects the use of the Bayesian ratio. The finding that base-rate usage increases under frequentist representations (Evans et al. 2002; Griffin & Buehler 1999) supports the proposal that the facilitation in Bayesian inference from natural frequency formats is due to the property of these formats to induce a representation of category instances that preserves the sample and effect sizes, thereby clarifying the underlying set structure of the problem and making the relevance of base-rates more obvious without providing an equation that generates Bayesian quantities.

2.9. Convergence with disparate data

A unique characteristic of the dual process position is that it predicts that nested sets should facilitate reasoning whenever people tend to rely on associative rather than extensional, rule-based processes; facilitation should be observed beyond the context of Bayesian probability updating. The natural frequency theories expect facilitation only in the domain of probability estimation.

In support of the nested sets position, facilitation through nested set representations has been observed in a number of studies of deductive inference. Grossen and Carnine (1990) and Monaghan and Stenning (1998) reported significant improvement in syllogistic reasoning when participants were taught using Euler circles. The effect was restricted to participants who were "learning impaired" (Grossen & Carnine 1990) or had a low GRE score (Monaghan & Stenning 1998). Presumably, those that did not show improvement did not require the Euler circles because they were already representing the nested set relations.

Newstead (1989, Experiment 2) evaluated how participants interpreted syllogisms when represented by Euler circles versus quantified statements. Newstead found that although Gricean errors of interpretation occurred when syllogisms were represented by Euler circles and quantified statements, the proportion of conversion errors, such as converting "Some A are not B" to "Some B are not A," was significantly reduced in the Euler circle task. For example, less than 5% of the participants generated a conversion error for "Some ... not" on the Euler circle task, whereas this error occurred on 90% of the responses for quantified statements.

Griggs and Newstead (1982) tested participants on the THOG problem, a difficult deductive reasoning problem involving disjunction. They obtained a substantial amount of facilitation by making the problem structure explicit, using trees. According to the authors, the structure is normally implicit due to negation and the tree structure facilitates performance by cuing formation of a mental model similar to that of nested sets.

Facilitation has also been obtained by making extensional relations more salient in the domain of categorical inductive reasoning. Sloman (1998) found that people who were told that all members of a superordinate have some property (e.g., all flowers are susceptible to thrips), did not conclude that all members of one of its subordinates inherited the property (e.g., they did not assert that this guaranteed that all roses are susceptible to thrips). This was true even for those people who believed that roses are flowers. But if the assertion that roses are flowers was included in the argument, then people did abide by the inheritance rule, assigning a probability of one to the statement about roses. Sloman argued that this occurred because induction is mediated by similarity and not by class inclusion, unless the categorical – or set – relation is made transparent within the statement composing the argument (for an alternative interpretation, see Calvillo & Revlin 2005).

Facilitation in other types of probability judgment can also be obtained by manipulating the salience and structure of set relations. Sloman et al. (2003) found that almost no one exhibited the conjunction fallacy when the options were presented as Euler circles, a representation that makes set relations explicit. Fox and Levav (2004) and Johnson-Laird et al. (1999) also improved judgments on probability problems by manipulating the set structure of the problem.

2.10. Empirical summary and conclusions

In summary, the empirical review supports five main conclusions. First, the facilitory effect of natural frequencies on Bayesian inference varied considerably across the reviewed studies (see Table 3), potentially resulting from differences in the general intelligence level and motivation of participants (Brase et al. 2006). These findings support

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

the nested sets hypothesis to the degree that intelligence and motivation reflect the operation of domain general and strategic – rather than automatic (i.e., modular) – cognitive processes.

Second, questions that prompt use of category instances and divide the solution into the sets needed to compute the Bayesian ratio, facilitate probability judgment. This suggests that facilitation depends on cues to the set structure of the problem rather than (an evolved) capacity to process natural frequencies. In further support of this conclusion, partitioning the data into nested sets facilitates Bayesian inference regardless of whether natural frequencies or single-event probabilities are employed (see Table 5).

Third, frequency judgments are guided by inferential strategies that reflect incomplete, fragmentary memories that do not entail the base-rates (e.g., Bradburn et al. 1987; Gluck & Bower 1988). This suggests that Bayesian inference does not derive from the accurate encoding and retrieval of natural frequencies. In addition, natural frequencies and single-event probabilities are rated similarly in their perceived clarity, understandability, and impact on the respondent's behavior (Brase 2002b), further suggesting that the mind does not embody inductive reasoning mechanisms (that are designed) to process natural frequencies.

Fourth, people (a) do not accurately weight and combine event frequencies, and (b) utilize event frequencies that are irrelevant in the calculation of Bayes' theorem (e.g., Griffin & Buehler 1999; Kleiter et al. 1997). This suggests that the cognitive operations that underlie Bayesian inference do not conform to Bayes' theorem. Furthermore, base-rate usage increases under frequentist representations (e.g., Griffin & Buehler 1999), suggesting that facilitation results from the property of natural frequencies to represent the sample and effect sizes, which highlight the set structure of the problem and make transparent what is relevant for problem solving.

Finally, nested set representations facilitate reasoning in a range of classic deductive and inductive reasoning tasks. This supports the nested set hypothesis that the mind embodies a domain general capacity to perform elementary set operations and that these operations can be induced by cues to the set structure of the problem to facilitate reasoning in any context where people tend to rely on associative rather than extensional, rule-based processes.

3. Conceptual issues

This section provides a conceptual analysis that addresses (1) the plausibility of the natural frequency assumptions, and (2) whether natural frequency representations support properties that are central to human inductive reasoning competence, including reasoning about statistical independence, estimating the probability of unique events, and reasoning on the basis of similarity, analogy, association, and causality.

3.1. Plausibility of natural frequency assumptions

The natural sampling framework was established by the seminal work of Kleiter (1994), who assessed "the correspondence between the constraints of the statistical model of natural sampling on the one hand, and the constraints under which human information is acquired on the other" (p. 376). Kleiter proved that under natural sampling and other conditions (e.g., independent identical sampling), the frequencies corresponding to the base-rates are redundant and can be ignored. Thus, conditions of natural sampling can simplify the calculation of the relevant probabilities and, as a consequence, facilitate Bayesian inference (see Note 2 of the target article). Kleiter's computational argument does not appeal to evolution and was advanced with careful consideration of the assumptions upon which natural sampling are based. Kleiter noted, for example, that the natural sampling framework (a) is limited to hypotheses that are mutually exclusive and exhaustive, and (b) depends on collecting a sufficiently large sample of event frequencies to reliably estimate population parameters.

Although people may sometimes treat hypotheses as mutually exclusive (e.g., "this person is a Democrat, so they must be anti-business"), this constraint is not always satisfied: many hypotheses are nested (e.g., "she has breast cancer" vs. "she has a particular type of breast cancer") or overlapping (e.g., "this patient is anxious or depressed"). People's causal models typically provide a

Table 5. *Percent correct for Bayesian inference problems reported in the literature (sample sizes in parentheses)*

	Information structure			
	Non-partitive		Partitive	
Study	Probability	Frequency	Probability	Frequency
Girotto & Gonzalez (2001; Study 5)	—	—	—	53 (20)
Girotto & Gonzalez (2001; Study 6)	0 (20)	0 (20)	—	—
Macchi (2000)	6 (30)*	3 (30)	36 (30)	40 (30)
Sloman et al. (2003; Exp. 1)	20 (25)*	—	48 (48)*	51 (45)
Sloman et al. (2003; Exp. 2)	—	—	48 (25)*	—
Sloman et al. (2003; Exp. 4)	—	21 (33)	—	—

Note. Studies that present questions that require the respondent to compute a conditional-event probability are indicated by*. The remaining studies present questions that prompt the respondent to compute the two terms of the Bayesian solution.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

wealth of knowledge about classes and properties, allowing consideration of many kinds of hypotheses that do not necessarily come in mutually exclusive, exhaustive sets. As a consequence, additional principles are needed to broaden the scope of the natural sampling framework to address probability estimates drawn from hypotheses that are not mutually exclusive and exhaustive. In this sense, the nested sets theory is more general: It can represent nested and overlapping hypotheses by taking the intersection (e.g., "she has breast cancer *and* it is type X") and union (e.g., "the patient is anxious *or* depressed) of sets, respectively.

As Kleiter further notes, inferences about hypotheses from encoded event frequencies are warranted to the extent that the sample is sufficiently large and provides a reliable estimate of the population parameters. The efficacy of the natural sampling framework therefore depends on establishing (1) the approximate number of event frequencies that are needed for a reliable estimate, (2) whether this number is relatively stable or varies across contexts, and (3) whether or not people can encode and retain the required number of events.

3.2. Representing qualitative relations

In contrast to single-event probabilities, natural frequencies preserve information about the size of the reference class and, as a consequence, do not directly indicate whether an observation and hypothesis are statistically independent. For example, probability judgments drawn from natural frequencies do not tell us that a symptom present in (a) 640 out of 800 patients with a particular disease and (b) 160 out of 200 patients without the disease, is not diagnostic because 80% have the symptom in both cases (Over 2000a; 2000b; Over & Green 2001; Sloman & Over 2003). Thus, probability estimates drawn from natural frequencies do not capture important qualitative properties.

Furthermore, in contrast to the cited benefits of non-normalized representations (e.g., Gigerenzer & Hoffrage 1995), normalization may serve to simplify a problem. For example, is someone offering us the same proportion if he tries to pay us back with 33 out of 47 nuts he has gathered (i.e., 70%), after we have earlier given him 17 out of 22 nuts we have gathered (i.e., 77%)? This question is trivial after normalization, as it is transparent that 70 out of 77 out of 100 are nested sets (Over 2007).

3.3. Reasoning about unique events and associative processes

One objection to the claim that the encoding of natural frequencies supports Bayesian inference is that intuitive probability judgment often concerns (a) beliefs regarding single events, or (b) the assessment of hypotheses about novel or partially novel contexts, for which prior event frequencies are unavailable. For example, the estimated likelihoods of specific outcomes are often based on novel and unique one-time events, such as the likelihood that a particular constellation of political interests will lead to a coalition. Hence, Kahneman and Tversky (1996, p. 589) argue that the subjective degree of belief in hypotheses derived from single events or novel contexts "cannot be generally treated as a random sample from some reference population, and their judged probability cannot be reduced to a frequency count."

Furthermore, theories based on natural frequency representations do not allow for the widely observed role of similarity, analogy, association, and causality in human judgment (for recent reviews of the contribution of these factors, see Gilovich et al. 2002 and Sloman 2005). The nested sets hypothesis presupposes these determinants of judgment by appealing to a dual-process model of judgment (Evans & Over 1996; Sloman 1996a; Stanovich & West 2000), a move that natural frequency theorists are not (apparently) willing to make (Gigerenzer & Regier 1996). The dual-process model attributes responses based on associative principles, such as similarity, or responses based on retrieval from memory, such as analogy, to a primitive associative judgment system. It attributes responses based on more deliberative processing involving rule-based inference, such as the elementary set operations that respect the logic of set inclusion and facilitate Bayesian inference, to a second deliberative system. However, this second system is not limited to analyzing set relations. It can also, under the right conditions, do the kinds of structural analyses required by analogical or causal reasoning.

Within this framework, natural frequency approaches can be viewed as making claims about rule-based processes (i.e., the application of a psychologically plausible rule for calculating Bayesian probabilities), without addressing the role of associative processes in Bayesian inference. In light of the substantial literatures that demonstrate the role of associative processes in human judgment, Kahneman and Tversky (1996, p. 589) conclude, "there is far more to inductive reasoning and judgment under uncertainty than the retrieval of learned frequencies."

4. Summary and conclusions

The conclusions drawn from the diverse body of empirical and conceptual issues addressed by this review consistently challenge theories of Bayesian inference that depend on natural frequency representations (see Table 2), demonstrating that coherent probability estimates are not derived according to an equational form for calculating Bayesian posterior probabilities that requires the use of such representations.

The evidence instead supports the nested sets hypothesis that judgmental errors and biases are attenuated when Bayesian inference problems are represented in a way that reveals underlying set structure, thereby demonstrating that the cognitive capacity to perform elementary set operations constitutes a powerful means of reducing associative influences and facilitating probability estimates that conform to Bayes' theorem. An appropriate representation can induce people to substitute reasoning by rules with reasoning by association. In particular, the review demonstrates that judgmental errors and biases were attenuated when (a) the question induced an outside view by prompting the respondent to utilize the sample of category instances presented in the problem, and when (b) the sample of category instances were represented in a nested set structure that partitioned the data into the components needed to compute the Bayesian solution.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Although we disagree with the various theoretical interpretations that could be attributed to natural frequency theorists regarding the architecture of mind, we do believe that they have focused on and enlightened us about an important phenomenon. Frequency formulations are a highly efficient way to obtain drastically improved reasoning performance in some cases. Not only is this an important insight to improve and teach reasoning, but it also focuses theorists on a deep and fundamental problem: What are the conditions that compel people to overcome their natural associative tendencies in order to reason extensionally?

ACKNOWLEDGMENTS

This work was supported by National Science Foundation Grants DGE-0536941 and DGE-0231900 to Aron K. Barbey. We are grateful to Gary Brase, Jonathan Evans, Vittorio Girotto, Philip Johnson-Laird, Gernot Kleiter, and David Over for their very helpful comments on prior drafts of this paper. Barbey would also like to thank Lawrence W. Barsalou, Sergio Chaigneau, Brian R. Cornwell, Pablo A. Escobedo, Shlomit R. Finkelstein, Carla Harenski, Corey Kallenberg, Patricia Marsteller, Robert N. McCauley, Richard Patterson, Diane Pecher, Philippe Rochat, Ava Santos, W. Kyle Simmons, Irwin Waldman, Christine D. Wilson, and Phillip Wolff for their encouragement and support while writing this paper.

NOTES

1. The respondent's subjective degree of belief in the hypothesis (H) that the patient has breast cancer, given the observed datum (D) that she has a positive mammography (i.e., the posterior probability, Pr(H | D)) can be expressed numerically as the ratio between (a) the probability that the patient has the disease and obtains a positive mammogram (Pr(H $\cap$ D)), and (b) the probability that the patient obtains a positive mammogram (Pr(D)). To calculate this ratio, Bayes' theorem incorporates two axioms of mathematical probability theory: the conditional probability and additivity laws. According to the former, (a) can be expressed by the probability that the patient has the disease (i.e., the base-rate of the hypothesis) multiplied by the probability that the patient obtains a positive mammogram, given that she has the disease (i.e., the hit-rate of the test): Pr(H $\cap$ D) = Pr(H) Pr(D | H). The additivity rule is then applied to express (b) as the probability that the patient has the disease and obtains a positive mammogram, plus the probability that the patient does not have the disease and obtains a positive mammogram: Pr(D) = Pr(H $\cap$ D) + Pr($\sim$H $\cap$ D). The conditional probability rule can be further applied to express this latter quantity as the complement of the base-rate multiplied by the probability that the patient obtains a positive mammogram, given that she does not have the disease (i.e., the false alarm rate of the test): Pr($\sim$H $\cap$ D) = Pr($\sim$H) Pr(D | $\sim$ H). Thus, according to Bayes' theorem, the probability that the patient has breast cancer, given that she has a positive mammography, equals Pr(H | D) = Pr(H | D) / Pr(D) = Pr(H) Pr(D | H) / Pr(H) Pr(D | H) + Pr($\sim$H) Pr(D | $\sim$ H) = (0.01)(0.80) / [(0.01)(0.80) + (0.99)(0.096)], or 7.8 per cent.

2. When estimated from natural frequency formats or formats expressing numbers of chances, because they entail the sample and effect sizes, posterior probabilities can be calculated in a way that does not require that the probabilities be multiplied by the base-rates. The following simple form can be used to calculate the probability of a hypothesis (H) given datum (D): $Pr(H \mid D) = [N(H \cap D)/N(H \cap D) + N(\sim H \cap D)]$, where $N(H \cap D)$ is the number of cases having the datum in the presence of the hypothesis, and $N(\sim H \cap D)$ is the number of cases having the datum in the absence of the hypothesis. This form requires that the respondent attend only to the $N(H \cap D)$ and the $N(\sim H \cap D)$, whereas estimating posteriors with percentages requires transforming percentage values into conditional probabilities by incorporating base-rates, making the calculation more complex than under natural frequency formats.

3. There may be an important relation between sensitivity to nested-set structure and the law of the excluded middle that appears in logic. By this rule, all propositions of the form "p or not-p" hold. We apply the rule, for example, to infer that everyone either has a disease or does not have the disease. We use it again to infer that everyone has some symptom or does not have it. Thus, the logical trees cited by natural frequency theorists are consistent with this fundamental logical rule (Over 2007).

4. Girotto and Gonzalez (2001) point out that the chance representation of probability is commonly employed in everyday situations, such as when someone says, "A tossed coin has one out of two chances of landing head up" or that "there is one out of a million chances of winning the lottery." Chances preserve information about the size of the reference class (i.e., the total population of chances). Hoffrage et al. (2002) argue that chances are just frequencies. This is false (see Girotto & Gonzalez 2002). Chances refer to the probability of a single event and are based on the total population of chances rather than a finite sample of observations. The chances, for example, of drawing an ace from a standard deck of playing cards are "4 out of 52": There are four ways that an ace can be drawn from the deck of 52 cards. In contrast to natural frequencies, the size of the reference class represents the total population (i.e., the deck of 52 cards). We might observe, for example, that one out of 10 cards randomly drawn from the deck is an ace, but this method of "natural sampling" would not represent the chance or number of ways of drawing an ace from the full deck. Chances cannot be directly assessed by "counting occurrences of events as they are encountered and storing the resulting knowledge base for possible use later" (i.e., natural sampling; Brase 2002b, p. 384). Chances are thus distinct from natural frequencies.

5. The mind-as-Swiss-army-knife, natural frequency algorithm, and natural frequency heuristic theories do not concern the encoding of event frequencies under naturalistic settings in general, but focus only on event frequencies that have a partitive structure. Therefore, these approaches do not address the encoding of non-partitive event frequencies (e.g., the event frequency of naturally occurring independent events). Given that both frequencies exist in nature, it is unclear why only frequencies of the latter type are deemed important.

6. Bayes' theorem in odds form refers to the probability in favor of a hypothesis (H) over the probability of an alternative hypothesis ($\sim H$), given observed datum (D) (i.e., the posterior odds: $[Pr(H \mid D) / Pr(\sim H \mid D)]$. To compute the posterior odds, Bayes' theorem incorporates two factors: the likelihood ratio and the prior odds. The likelihood ratio is a measure of whether the datum is diagnostic with respect to the hypothesis (H). If the evidence is diagnostic then the likelihood ratio will be positive, demonstrating that the observed datum is more likely to occur under the presence of the hypothesis (H) than under the alternative hypothesis ($\sim H$). The prior odds is the ratio of base-rate probabilities $[Pr(H) / Pr(\sim H)]$. Bayes' theorem in odds form states that the product of these quantities yields the posterior odds, $Pr(H \mid D) / Pr(\sim H \mid D) = [Pr(D \mid H) / Pr(D \mid \sim H)] * [Pr(H) / Pr(\sim H)]$. To directly estimate the relative weight of the likelihood ratios and prior odds, Bayes' theorem in odds form can be logarithmically transformed to yield $\log [Pr(H \mid D) / Pr(\sim H \mid D)] = \log [Pr(D \mid H) / Pr(D \mid \sim H)] + \log [Pr(H) / Pr(\sim H)]$. Under this formulation, the likelihood ratios and prior odds can be treated as independent variables in a regression analysis to assess the relative contribution of each factor in Bayesian inference.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Open Peer Commentary

A statistical taxonomy and another "chance" for natural frequencies

DOI: 10.1017/S0140525X07001665

Adrien Barton,[a,b] Shabnam Mousavi,[a,c] and Jeffrey R. Stevens[a]

[a]*Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, 14195 Berlin, Germany;* [b]*Institut d'Histoire et de Philosophie des Sciences et des Techniques, Paris I, and CNRS/ENS – UMR 8590, 75006 Paris, France;* [c]*Department of Statistics, The Pennsylvania State University, University Park, PA 16802.*

barton@mpib-berlin.mpg.de http://www.mpib-berlin.mpg.de/en/forschung/abc/index.htm mousavi@mpib-berlin.mpg.de sxm70@psu.edu http://www.stat.psu.edu/people/faculty/smousavi.html/jstevens@mpib-berlin.mpg.de http://www.abc.mpib-berlin.mpg.de/users/jstevens/

Abstract: The conclusions of Barbey & Sloman (B&S) crucially depend on evidence for different representations of statistical information. Unfortunately, a muddled distinction made among these representations calls into question the authors' conclusions. We clarify some notions of statistical representations which are often confused in the literature. These clarifications, combined with new empirical evidence, do not support a dual-process model of judgment.

We disagree with Barbey & Sloman's (B&S's) claim that data on Bayesian reasoning support their dual-process model of human judgment. First, we clarify the dimensions along which statistical information can be expressed, and then point to how this common conceptual confusion can influence B&S's interpretation of existing data. Second, we explain how new evidence contradicts this model.

1. Statistical representation. Statistical information can be represented in multiple ways along two orthogonal dimensions: the number of events and the numerical format. First, the information may concern only one event (single-event probability) or a set of events (frequency). For example, the probability that a person has a positive test if she is ill is a single-event probability, in contrast to the frequency of people having a positive test among those who are ill. Second, the numerical information can be represented as percentages (20%); fractions (20/100); real numbers between 0 and 1 (0.2); or pairs of integers ("20 chances out of 100" for single-event probabilities, and "20 people out of 100" for frequencies).

Consider now a Bayesian task of computing the probability of a hypothesis H, given the data D, such as the probability of being ill, given the result of a test. In this context, there is yet another orthogonal dimension along which the statistical information can vary: the information can be expressed in a *conjunctive* or in a *normalized* format. The conjunctive format gives the relevant conjunctive information $P(H \& D)$ and $P(\textit{not-}H \& D)$, or $P(H \& D)$ and $P(D)$. In this case the Bayesian computations are rather simple (see Eq. 1):

$$P(H \mid D) = \frac{P(H\&D)}{P(H\&D) + P(\textit{not}\text{-H\&D})} = \frac{P(H\&D)}{P(D)} \qquad (1)$$

Alternatively, information can be expressed in a normalized format giving the normalized information $P(D \mid H)$ and $P(D \mid \textit{not-}H)$, in addition to $P(H)$ – and not giving the relevant conjunctive information. The normalized format complicates computing the Bayesian results (see Eq. 2):

$$P(H \mid D) = \frac{P(D \mid H)P(H)}{P(D \mid H)P(H) + P(D \mid not-H)(1 - P(H))} \qquad (2)$$

Because the number of events and conjunctive/normalized format dimensions are orthogonal, one can give statistical information in a Bayesian task in four possible ways, each of which can be represented as percentages, fractions, real numbers or pair of integers (see Table 1 for examples of Bayesian tasks, each represented in different numerical formats).

Confusion among these three orthogonal dimensions is common in the literature and poses particular problems in the

Table 1 (Barton et al.). *Taxonomy of statistical information: examples of Bayesian tasks*

	Normalized format	Conjunctive format
Single event probabilities	A 40-year-old woman who participates in routine screening has 10 out of 1,000 chances to have breast cancer. [$P(H)$]	The probability of breast cancer is 1% for a 40-year-old woman who participates in a routine screening. [$P(H)$]
	If such a woman has breast cancer, she has 800 out of 1,000 chances to have a positive mammography [$P(D \mid H)$]	The probability that such a woman has a positive mammography and has breast cancer is 0.8% [$P(H \& D)$]
	If such a woman does not have breast cancer, she has 96 out of 1,000 chances to have a positive mammography. [$P(D \mid \textit{not-}H)$]	The probability that she has a positive mammography and does not have breast cancer is 9.5%. [$P(\textit{not-}H \& D)$]
Frequencies	A proportion of 0.01 of women at age 40 who participate in routine screening have breast cancer. [$P(H)$]	(labeled *Natural frequencies* when represented as pairs of integers) 10 out of 1,000 women at age forty who participate in routine screening have breast cancer [$P(H)$]
	A proportion 0.8 of women with breast cancer will have positive mammographies. [$P(D \mid H)$]	8 out of these 1,000 women have a positive mammography and have breast cancer [$P(H \& D)$]
	A proportion 0.096 of women without breast cancer will also have positive mammographies. [$P(D \mid \textit{not-}H)$]	95 out of these 1,000 women have a positive mammography and do not have breast cancer [$P(\textit{not-}H \& D)$]

Note. Each of the four numerical formats can apply to the four probability/frequency and normalized/conjunctive combinations. Here, we arbitrarily assigned numerical formats for each cell.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

B&S target article, because the authors draw false conclusions on this basis. First, they mention "natural frequency formats that were not partitioned into nested set relations" (B&S, sect. 2.4, para. 2). But non-partitioned frequency formats are simply frequencies expressed in a normalized format; therefore, natural frequencies *must* be partitioned into nested set relations. Consequently, B&S's critiques of the so-called non-partitioned natural frequencies apply only to normalized frequencies (i.e., frequencies in a normalized format). Second, a sentence like "33% of all Americans will have been exposed to Flu" (target article, Table 5) concerns a whole population, not a single individual; therefore, it conveys frequencies, not single-event probabilities, contrary to the authors' equating of percentages (referring to the numerical representation) and single-event probabilities (referring to the number of events). B&S misinterpret a key result of Brase (2002b), who showed that subjects perceived simple frequencies (i.e., frequencies represented as pairs of integers) as clearer, more understandable, and more impressive than single-event probabilities represented as non-integer numbers (Brase 2002b, pp. 388–89).

B&S contrast what they call "theoretical frameworks" (sect. 1.2, para. 2) based on natural frequency representations with the "nested set hypothesis" (sect. 2.10, para. 5). However, some of these contrasts appear a bit artificial. Consider the Gigerenzer and Hoffrage (1995) study, which predicted and showed that natural frequencies facilitate Bayesian inference. B&S claim this effect results from the clarification of the nested-sets structure of the problem. But Gigerenzer and Hoffrage (1995) had already made a more specific, related argument, stating that the facilitation of natural frequencies results from simplifying the Bayesian computations by giving the relevant conjunctive information. B&S's idea that this facilitation results from the clarification of the nested-sets structure of the problem stands more in opposition to the use of an evolutionary argument (Cosmides & Tooby 1996) to predict the facilitating effect of natural frequencies, than in opposition to Gigerenzer and Hoffrage's argument. However, some of the arguments used by B&S against this evolutionary stance are not valid. For example, they say that since both normalized and natural "frequencies exist in nature, it is unclear why only frequencies of the latter type are deemed important" (target article, Note 5). This is not true: Natural frequencies "exist in nature" in the sense that, in a natural sample of the population, counting the number of individuals belonging to the groups $H \& D$ and *not-H* $\& D$ yields natural frequencies (conjunctive information); whereas, counting in such a natural sample cannot result in normalized information that does not contain the conjunctive information. Therefore, normalized frequencies, which only give the normalized information, do not exist in nature in this sense. So, an evolutionary argument that predicts the facilitating effect of natural frequencies, but no such effect for normalized frequencies, can be defended against this charge that both normalized and natural frequencies exist in nature (even if one sees this evolutionary argument as speculative and feels uncomfortable about making precise predictions on this basis).

2. Facilitating effects. B&S propose a general dual-process model of judgment, which denies any facilitating effect of frequencies per se, because "facilitation is a product of general-purpose reasoning processes" (sect. 1.2.5, para. 1). As evidence against such an effect, B&S cite Girotto and Gonzalez's (2001) facilitating effect when the information is given as "number of chances" in a conjunctive format, which is a way of expressing single-event probabilities. Recent work by Brase (2007), however, demonstrates that many people interpret such chances as natural frequencies, despite instructions to the contrary. Moreover, those who interpret chances as natural frequencies have higher rates of success than those who judge the information as single-event probabilities. This suggests that frequencies can have a facilitating effect in some circumstances, in addition to the facilitating effect of computational simplicity. If this facilitating effect of frequencies is confirmed, it would make the dual-process model much more difficult to defend.

Conclusion. Though we do not dismiss the idea of a dual-process model outright, we think that B&S have not made a robust argument in support of such a model. The authors misinterpret data used to reject alternatives to the nested-set hypothesis. Further, the connection between the nested-set hypothesis and the dual-process model of judgment is not as crisp as one would like. Perhaps this is due to the rather vague nature of the dual-process model itself (cf. the criticisms of Gigerenzer & Regier 1996). The general project of building such a model sounds exciting, but we look forward to a more rigorous, clearly defined (and therefore falsifiable) dual-process model of judgment.

ACKNOWLEDGMENTS

We thank Nils Straubinger and Gerd Gigerenzer for their comments on the manuscript.

From base-rate to cumulative respect

DOI: 10.1017/S0140525X07001677

C. Philip Beaman and Rachel McCloy

Department of Psychology, University of Reading – Earley Gate, Whiteknights, Reading RG6 6AL, United Kingdom.

c.p.beaman@reading.ac.uk http://www.personal.rdg.ac.uk/~sxs98cpb/philip_beaman.htm r.a.mccloy@reading.ac http://www.psychology.rdg.ac.uk/people/lecturing/Dr_Rachel_McCloy.php

Abstract: The tendency to neglect base-rates in judgment under uncertainty may be "notorious," as Barbey & Sloman (B&S) suggest, but it is neither inevitable (as they document; see also Koehler 1996) nor unique. Here we would like to point out another line of evidence connecting ecological rationality to dual processes, the failure of individuals to appropriately judge *cumulative* probability.

Recent data in studies by McCloy and colleagues (McCloy et al. 2007; McCloy et al., submitted) show that judgment of cumulative, disjunctive risk (i.e., the probability of avoiding an adverse event over a period of time during which one continually engages in a risky activity) benefits from presentation in a frequency, rather than a probability, format (McCloy et al., submitted). It does this in a similar manner to the way in which judgments of conditional probability avoid the base-rate neglect fallacy if presented in natural frequency format (Gigerenzer & Hoffrage 1995). Further, training in translation from probability to frequency formats shows similar improvements relative to baseline for both types of judgment (McCloy et al. 2007). However, the effects of both format and training are mimicked by presenting information in a partitive or "nested" set structure (in our studies, diagrammatically represented by probability trees rather than Euler circles). This suggests that similar processes may be involved in both problem types, and we applaud Barbey & Sloman (B&S) for attempting to break down the nature of those processes rather than remaining satisfied with a "natural frequency" label. However, we do not believe that B&S have (yet) produced a full and complete account of the means by which dual processes may produce rationality within certain given ecologies.

One worry is the assertion that people do not have an (evolved or otherwise) capacity to encode frequency information. This is a claim concerning a failure to observe a particular capability and might reflect failures in the observational technique employed as much as a failure in the capability itself. The data reviewed by B&S are based upon individuals' inability to produce, on demand, explicit knowledge obtained of frequencies from episodes of incidental learning. This is known to be problematic

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

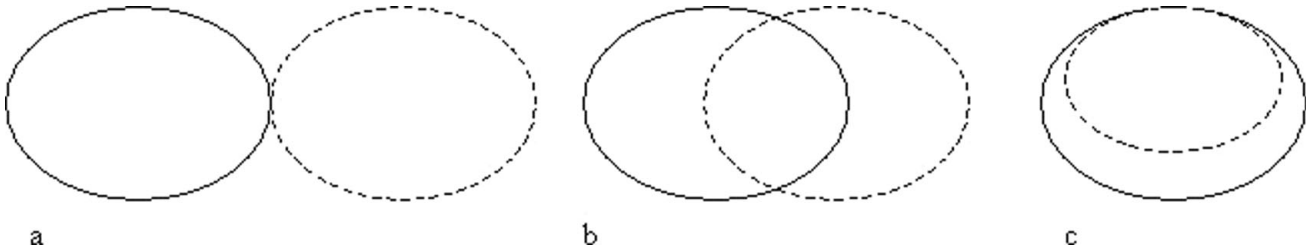


Figure 1 (Beaman & McCloy). Three situations (a, b, and c) representing the relationship between the set of car accidents in year 1 (solid circle, SC) and year 2 (dashed circle, DC).

(Morton 1968). This may – perhaps – explain why well-controlled laboratory studies (e.g., Sedlmeier et al. 1998) have had better success in showing accurate frequency judgments than studies in naturalistic settings do, although, as B&S note, such lab studies have typically not covered autobiographical events to which a Bayesian inference structure can be applied.

A second concern is directly related to the issue of training individuals to translate from probabilities to frequencies (McCloy et al. 2007). In this study we examined whether training people to represent the data to themselves in a partitive structure allowed for accurate responding to cumulative risk judgments. This was contrasted with equivalent judgment based on single-event probabilities expressed relative to a single time period. Participants were given statements such as: "Suppose that a person who drives fast whilst using a cell phone has a 90% probability of not being involved in a car accident in any one year. What is the probability that they avoid being involved in an accident at all if they continue to drive in the same way, over the same roads, for three years?" This statement produced, on average, only 25% correct responding, although the statistical "rule" for disjunctive cumulative probability ($1 - p^n$) is considerably simpler in form than Bayes' theorem. Following training in recoding this data into either a probability or a frequency tree, however, performance improved to approximately 67% correct, and stayed at this level after a one-week interval. This suggests that problem structure is important for more than just base-rate neglect problems – but it also begs the question of *how* problem structure is represented if not in frequency terms. Our participants were taught to use tree-structures and proved reasonably able at learning and using these. Sloman et al. (2003) employed Euler circles to likewise make nested set relations transparent for Bayes' theorem. Unfortunately, although tree structures and Euler circles may be interchangeable as aids to conditional reasoning, they are not necessarily equivalent when applied to cumulative probability judgments, as the diagram in Figure 1 makes clear.

With tree structures, the branching of different states of the world over time can be represented within a single tree regardless of whether the problem is conditional or cumulative (a fact that caused our participants some difficulty when unexpectedly confronted with problems of an unexpected type). In contrast, for conditional probabilities the number of (or chances of) having a disease, given a positive test, is represented by only one possible set of Euler circles (see target article's Fig. 1). Following Sloman et al. (2003), the probability of avoiding a car accident over the two-year period is 1 minus p (SC union DC) but the number of (or chances of) avoiding a car crash over a two-year period is, potentially, the complement of any one of three pairs of Euler circles (a, b, and c in our Fig. 1). Although for any realistic probability values circle b is considerably more probable, this may not be immediately apparent using solely partitive information.

Despite this difficulty, framing cumulative probability judgments in such a way that nested relations are transparent improves performance in a manner similar to making sorts (or kinds) of relations transparent within Bayesian judgments. This leaves us with the questions: If not all diagrammatic representations of nested relations are equal, what type of mental representation(s) of such relations are being employed in order to reason extensionally; and, crucially, when and how is this representational system employed?

Kissing cousins but not identical twins: The denominator neglect and base-rate respect models

DOI: 10.1017/S0140525X07001689

C. J. Brainerd

Departments of Human Development and Psychology and Cornell Law School, Cornell University, Ithaca, NY 148453.

cb299@cornell.edu

http://www.human.cornell.edu/che/bio.cfm?netid=cb299

Abstract: Barbey & Sloman's (B&S's) base-rate respect model is anticipated by Reyna's denominator neglect model. There are parallels at three levels: (a) explanations are grounded in a general cognitive theory (rather than in domain-specific ideas); (b) problem structure is treated as a key source of reasoning errors; and most importantly, (c) nested set relations are seen as the cause of base-rate neglect.

Science presents occasional examples of parallel development of the same ideas to explain the same findings. I comment here on such an example found in the psychology of judgment and decision making: the dual-process model of base-rate neglect proposed in the target article by Barbey & Sloman (B&S) and a dual-process model of base-rate neglect that was developed in the 1990s by Reyna (1991; Reyna & Brainerd 1993; 1994; 1995).

The key properties of the Barbey-Sloman model are its assumptions that (a) an explanation of base-rate neglect must be grounded in a general cognitive theory (not domain-specific ideas), (b) structural features of base-rate problems are what cause errors, and (c) set nesting is the structural feature that is directly responsible for errors. These are also properties of Reyna's dual-process *denominator neglect model*. The research program that led to Reyna's theory was the first to develop a process model for Tversky and Kahneman's (1983) suggestion that set nesting produces errors in the conjunction fallacy, and was the first to elucidate specific cognitive difficulties that nested sets foment in reasoners.

Concerning assumption (a), the aim of Reyna's research program was to identify cognitive mechanisms that cause various classes of reasoning illusions. From the beginning, the guiding principles were that models of reasoning illusions, such as base-rate neglect, should be grounded in a general cognitive theory and that domain-specific accounts are at best unparsimonious and at worst untestable. This work produced a general framework, known as fuzzy-trace theory (FTT), that explains reasoning illusions by focusing on relations between memory processes and reasoning operations. FTT's level of generality is such that it has been widely used to explain basic memory processes (e.g., FTT's models of recognition and recall) as well as reasoning illusions.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Table 1 (Brainerd). *Memory advantages that make intuitive reasoning advanced*

Advantange	Description
Memory availability	Intuitive reasoning processes the types of memories (gist) that are stable over time. The types of memories processed by logical reasoning are not stable.
Memory accessibility	The gist memories that intuition processes are accessible by a broad range of retrieval cues.
Memory malleability	The sketchy nature of gist memories makes them especially easy to transform into solutions during reasoning.
Processing simplicity	The simplicity of gist memories makes intuitive reasoning relatively uncomplicated.
Processing accuracy	Intuitive processing of gist memories is usually just as accurate (and is often more accurate) than logical processing of more elaborate, detailed memories.
Processing effort	Intuitive processing of gist memories is easier to execute than logical processing of more elaborate, detailed memories.

Concerning assumption (b), FTT explains reasoning illusions by isolating structural properties of reasoning problems that interfere with three general stages of cognitive processing: (1) storing the correct problem representation (its "gist"), (2) retrieving that representation and the appropriate processing operations on reasoning problems, and (3) executing the steps that are required for the processing operations to deliver solutions. This approach is exemplified in FTT's explanations of many types of reasoning failures, such as arithmetic errors (Brainerd & Reyna 1988) and transitive inference errors (Reyna & Brainerd 1990).

Concerning assumption (c), extensive research was conducted on the cognitive mechanisms that cause errors in the family of illusions to which base-rate neglect belongs: *inclusion illusions*. Much of that work involved a prototypical task that produces such errors: Piaget's class-inclusion problem. The mechanisms that were identified were then generalized to base-rate problems, conjunction problems, probability problems, expected-value problems, and other tasks in the inclusion illusions family. Class-inclusion problems are structurally simple but cognitively impenetrable: Children are presented with an array of objects, subdivided into two (or more) familiar sets, such as 7 cows and 3 horses, that belong to a common superordinate set (10 animals), and are asked "Are there more animals or more cows?" Young children consistently respond: "more cows." This error persists for many years, with the error rate at age 10 still being 50% (Winer 1980), and adults routinely make the error on slightly more complex versions of the problem (Reyna 1991). Why are such problems so difficult?

The answer that emerged, following many experiments (e.g., Brainerd & Reyna 1990; 1995), is that nested sets interfere massively with the aforementioned processing stages. Reyna (1991) summarized the cognitive effects as follows: "processing focuses on the subset mentioned in the question, the superordinate set recedes, and the question appears to involve nothing more than . . . a subset-subset comparison" (p. 325). These effects were found to be rooted in the fact that problems in the inclusion illusions family have two-dimensional structures, with one dimension (the subset-subset) being salient and easy to process and the other (the subset-superordinate set), which is crucial to solution, being obscure. The obscurity is caused by the containment relation, which creates "mental booking" problems in which subsets disappear whenever the mind focuses on the superordinate set and the superordinate set disappears whenever the mind focuses on the subsets. Yet, correct reasoning demands subset-superordinate set comparisons. Reyna went on to formulate the denominator neglect model, wherein this difficulty was posited as the cause of base-rate neglect, the conjunction fallacy, and other errors that arise from comparing numerical parts to numerical wholes. The term "denominator" referred to the fact that denominator information is ignored because denominators are obscure wholes of part–whole relations.

A last point that illustrates the deep parallels between the Reyna and the Barbey-Sloman models is the centrality of formatting manipulations in tests of the models. These are manipulations that make problem structure more transparent and, crucially, enhance the salience of subset-superordinate set relations. Reyna noted that her model predicts that such manipulations reduce the mental bookkeeping problem and, therefore, should significantly reduce errors. A formatting manipulation called tagging provided dramatic confirmation. Young children, who failed problems such as the animal example across the board, performed nearly perfectly when simple tags (e.g., a hat on each animal's head, a bow on each animal's tail) were affixed to all the members of each subset, so that the superordinate set was just as salient as the subsets. Likewise, B&S stress the importance of presentation formats that allow "accurate representation in terms of nested sets of individuals" (target article, Abstract) in base-rate problems. With respect to the most effective presentation formats that they discuss, these formats, too, are ones that ought to reduce the mental bookkeeping problem of nested sets.

Although the Barbey-Sloman model is anticipated to a remarkable degree by the Reyna model, the dual-process frameworks that lie behind the models are different. The Barbey-Sloman framework is the traditional System 1/System 2 approach, which treats intuition as a primitive form of thinking that cognitive development and expertise evolve away from. The Reyna framework is FTT, which treats intuition as an advanced mode of thinking that cognitive development and expertise evolve *towards*. That intuition is advanced by virtue of memory considerations, though different ones than those which figure in the Barbey-Sloman model (see Table 1).

Omissions, conflations, and false dichotomies: Conceptual and empirical problems with the Barbey & Sloman account

DOI: 10.1017/S0140525X07001690

Gary L. Brase

Department of Psychology, Kansas State University, Manhattan, KS 66506-5302.

gbrase@ksu.edu

Abstract: Both the theoretical frameworks that organize the first part of Barbey & Sloman's (B&S's) target article and the empirical evidence marshaled in the second part are marked by distinctions that should not exist (i.e., false dichotomies), conflations where distinctions should be made, and selective omissions of empirical results – within the very studies discussed – that create illusions of theoretical and empirical favor.

Theoretical frameworks. The number of contrasting theoretical frameworks that Barbey & Sloman (B&S) face is impressive on the face of it – four against one.But are all these distinctions appropriate or scientifically useful? The short

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

way to demonstrate they are not is to simply note that the same theorists (Gigerenzer, Cosmides, and Tooby) are repeatedly invoked for all of the first three frameworks.

The longer demonstration is to dismantle these accounts sequentially. There is the "Mind as Swiss army knife" framework that is distinguished by unavailability to conscious awareness or deliberate control (cognitive impenetrability), and then there is the "Natural frequency algorithm" framework that is informationally encapsulated. But wait; those distinguishing characteristics are actually imposed by *others* (e.g., Fodor 1983) and rejected by the actual theorists under consideration here (e.g., Cosmides & Tooby 2003; Duchaine et al. 2001; Ermer et al. 2007; Tooby et al. 2005; Tooby & Cosmides 2005). The first and second theoretical frameworks therefore collapse into a third one.

The third framework (incorporating the previous two), is a "Natural frequency heuristic" account, and is probably closest to the one actual and appropriate opposing view for B&S. The fourth framework ("Non-evolutionary natural frequency heuristic") suggests that an appropriate position is to willfully disregard all evolutionary factors that have influenced the structure and function of the human mind. One can question the nature of the cognitive structures generated by evolutionary selection pressures, but it is not scientifically legitimate to simply deny evolution and replace viable evolutionary explanations with "one way or another, people can appreciate and use [natural sampling]" and that somehow "gives rise to" Bayesian reasoning (sect. 1.2.4). Such vague descriptive explanations would have effectively stagnated our understanding of visual processing or language acquisition, and will have that effect on other cognitive phenomena if unchecked.

This leaves us with two real frameworks, the final "nested sets/dual processes" framework and an ecological rationality framework – the two frameworks of the target article's title. It is not so much that there is no possibility of other frameworks, but rather, that the ones described by B&S are not useful.

The empirical literature. Having constructed artificial required properties for the theoretical frameworks of others, B&S then tout the inability of those shackled frameworks to account for empirical results. As easy as this should be, given such a set up, it is nevertheless seriously flawed. Due to space constraints, I focus here on how my own research is considered within this target article. B&S use the findings of Brase et al. (2006) to support a claim that "Bayesian inference depends on domain general cognitive processes" that are strategically employed (sect. 2.1). This was not the original purpose, findings, or conclusions of our work – and for good reason. As B&S note in that very same section, there have been differences in absolute performance levels on Bayesian reasoning tasks, when comparing across research programs. These different research programs, however, had used different participants and different methods for obtaining those participants (e.g., paid versus classroom activity participation). Brase et al. (2006) sought to determine the effects of participant selection and recruitment methods on performance on such tasks, and found that there were, indeed, significant effects that were capable of accounting for all the differences in previous works. In summary, B&S make a confusion between performance and competence (Chomsky 1965) when they try to infer cognitive abilities and structures from data showing that incentives affect performance (see also Crespi 1942; 1944).

There also appears to be some confusion about the nature of natural sampling and natural frequencies (i.e., naturally sampled frequencies). The use of a consistent reference class (sect. 2.3), also called using a partitive structure, nested sets, or subset relations, are all linguistic twists on what is, in fact, natural sampling (a point made many times by myself and others; Brase 2002a; 2002b; Brase & Barbey 2006; Gigerenzer & Hoffrage 1999; Hoffrage et al. 2002). Natural sampling refers to the sequential acquisition of information (as in a natural environment) along with categorization of that information into meaningful, often overlapping, groups (see Brase et al. 1998 for some limitations on easily constructible categories.).

This confusion is starkly illustrated when B&S try to re-define the numerical formats used in Brase (2002b). First, natural frequencies are equated with simple frequencies by providing an incorrect example of the former (this example belongs to B&S and is not, as they claim, an inconsistency with the literature on the part of Brase 2002b). In direct contradiction to B&S, a single numerical statement such as the simple frequencies used in Brase (2002b) *cannot* be identified as having a natural sampling structure. Second, B&S point out – correctly – that percentages can express single-event probabilities, but they then carry this too far in concluding that this is the *only* thing that probabilities can express. Indeed, as pointed out in Brase (2002b), percentages are also referred to as "relative frequencies" because they can be understood as frequencies that are normalized to a reference class of 100 (e.g., as when one says "90% of my students understand this topic").

With B&S having misconstrued natural frequencies into simple frequencies, and misconstrued relative frequencies into probabilities, it is almost possible to claim that the results of Brase (2002b) indicate that single event probabilities are perceived equally well compared to natural frequencies. The remaining necessary manipulation is for B&S to also completely omit the other numerical format conditions used in Brase (2002b), which included actual single-event probabilities (and, no, these actual single-event probabilities were *not* understood as well or clearly as simple frequencies and relative frequencies).

Why frequencies are natural

DOI: 10.1017/S0140525X07001707

Brian Butterworth

Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom.

b.butterworth@ucl.ac.uk

Abstract: Research in mathematical cognition has shown that *rates*, and other interpretations of x/y, are hard to learn and understand. On the other hand, there is extensive evidence that the brain is endowed with a specialized mechanism for representing and manipulating the numerosities of sets – that is, frequencies. Hence, base-rates are neglected precisely because they are rates, whereas frequencies are indeed natural.

Barbey & Sloman (B&S) are to be congratulated for laying out the explanations for base-rate neglect so clearly and systematically. However, to a researcher not from the field of normative rationality research, but from the field of mathematical cognition, it is surprising that none of the explanations make reference to what is known about how we process numerical quantities (Butterworth 2001). From this perspective, another type of explanation can be proposed for base-rate neglect. It is in the word "rate." Rates can be expressed formally as x/y, and it is well known from research in mathematical cognition and education that humans are very bad at understanding x/y however it is interpreted – as a fraction, as a proportion, or as a rate. For example, it is well known that children find it hard to learn and understand fractions and simple operations on them (Bright et al. 1988; Hartnett & Gelman 1998; Mack 1995; Smith et al. 2005). It has also been found that most third and fourth graders cannot order fractions by size and cannot explain why there are two numbers in a given fraction (Smith et al. 2005). In particular, they seem to have trouble getting away from whole numbers – for example, when they say that 1/56 is smaller than 1/75 because 56 is smaller than 75

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

(Stafylidou & Vosniadou 2004). This has been called "whole number bias" (Ni & Zhou 2005) and can be found in adults as well as children (Bonato et al., in press). Whole number bias is not simply a function of the symbolic form of the rate, for example, 3/5, because it appears also in non-symbolic formats such as arrays of dots (Fabbri et al., submitted).

The advantage of presentations in terms of frequencies, and therefore of whole numbers, rather than rates, again is well supported by research in mathematical cognition. This has nothing to do with the relative computational simplicity of representing the problem in terms of frequencies as compared with rate-based Bayesian formulations; rather, it has to do with the fact that the human brain is configured from birth to represent sets and their numerosities. Infants can discriminate small sets on the basis of their numerosity (Antell & Keating 1983; Starkey & Cooper 1980; Wynn et al. 2002). This seems to be an inherited capacity since other primates can do the same in the wild (Hauser et al. 1996), and can learn to do it relatively easily (Brannon & Terrace 2000). Indeed, monkeys readily learn to select the larger of two numerosities (Brannon & Terrace 1998; Matsuzawa 1985).

These primate capacities are not merely analogous to those of humans, but appear to have been inherited from a common ancestral system. Evidence for this comes from recent research showing that the primate brain areas for numerosity processing are homologous to human brain areas. Studies have demonstrated that the intraparietal sulcus (IPS) in humans processes the numerosities of sets (Piazza et al. 2002). It has recently been demonstrated that when monkeys are required to remember the numerosity of a set before matching to sample, the homologous IPS brain area is active (Nieder 2005). This is evidence that we have inherited the core of our system from the common ancestor of humans and macaques.

The concept of the numerosity of a set is abstract, because sets logically contain any type of member that can be individuated. Members need not be visible objects, and they need not be simultaneously present. It turns out that the human numerosity system in the IPS responds when members of the set are distributed as a sequence in time or simultaneously distributed in a spatial array (Castelli et al. 2006) and for auditory as well as visual sets (Piazza et al. 2006). Indeed, the neural process of extracting numerosity from sets of visible objects appears to be entirely automatic, since repeated presentation of different sets with same numerosity produces a reduction in neural firing in the IPS, called "adaptation," even when numerosity is task-irrelevant (Cantlon et al. 2006; Piazza et al. 2004; 2007).

"Frequency" is just a way of referring to this numerosity property of a set, and so it too is natural. "Natural sampling" can be interpreted to be a way of making an estimate of numerosity when the set is distributed in time or in space. Humans and other species are born with the capacity to make these estimates of the approximate size of a set, using a specialized brain system probably related to the system for exact numerosities. This system also responds to environmental stimuli in rapid and automatic manner (Cantlon et al. 2006; Dehaene et al. 1999; Lemer et al. 2003; Piazza et al. 2004). So natural sampling too is natural, in the sense that it depends on an innate system.

B&S note that accounts involving specialized modules (Cosmides & Tooby 1996), specialized frequency algorithms (Gigerenzer & Hoffrage 1995), or specialized frequency heuristics (Gigerenzer & Hoffrage 1995; Tversky & Kahneman 1974) appeal to evolution. However, these claims depend on general arguments about ecological rationality rather than on specific facts about the evolution of dedicated neural system. On the other hand, there is a clear account, well supported by a range of evidence, as I have indicated, for the evolution of numerosity processing. Indeed, the evidence suggests that numerosity processing is a classic Fodorian cognitive module: domain-specific, automatic, with a dedicated brain system, and innate (though Fodor himself cites the number domain as the responsibility of classic central processes; cf. Fodor 1983). Therefore, the critical difference between normative Bayesian reasoning and actual human preferences for sets and their frequencies appears to be rooted in the evolution of a specialized "number module" for processing numerosities (Butterworth 1999). As far as I know, there is no comparable evolutionary account of a specialized brain system for x/y.

Base-rate is neglected because rates are neglected.

Nested sets and base-rate neglect: Two types of reasoning?

DOI: 10.1017/S0140525X07001719

Wim De Neys

Experimental Psychology Lab, University of Leuven, 3000 Leuven, Belgium.

wim.deneys@psy.kuleuven.be

http://ppw.kuleuven.be/reason/wim/

Abstract: Barbey & Sloman (B&S) claim that frequency formats and other task manipulations induce people to substitute associative thinking for rule-based thinking about nested sets. My critique focuses on the substitution assumption. B&S demonstrate that nested sets are important to solve base-rate problems but they do not show that thinking about these nested sets relies on a different type of reasoning.

In the target article, Barbey & Sloman (B&S) argue against various versions of the popular natural frequency heuristic and claim that the best account of the data should be framed in terms of a dual-process model of judgment. Base-rate neglect with the standard problem format is attributed to the pervasiveness of the associative system. Frequency versions and other reviewed task manipulations are argued to boost performance because they would induce people to substitute associative thinking for rule-based thinking. Although I am sympathetic to the basic rationale behind the B&S framework, I want to point out that it lacks support for the crucial substitution assumption. The authors nicely clarify that representations in terms of nested sets reduce base-rate neglect but they do not show that thinking about these nested sets relies on a different type of reasoning. Such a claim requires an examination of the processing characteristics of the two postulated modes of thinking.

One of the core characteristics of rule-based reasoning is that it draws on executive, working-memory resources, whereas associative thinking is more automatic in nature (e.g., Stanovich & West 2000). If the good performance on the frequency versions is due to a switch to rule-based reasoning, one would at least need to show that people recruit executive resources when they solve the frequency versions. This would demonstrate that the kind of thinking that is triggered by the frequency format exhibits the hallmark of rule-based thinking.

The good news is that B&S's model leads to some clear-cut, testable predictions. It is not hard to see, for example, how the substitution claim could be directly tested in a dual-task study (e.g., see De Neys 2006a; 2006b, for a related approach). B&S argue that in the vast majority of cases people rely on automatic, associative thinking to solve the standard probability format problems. Hence, burdening peoples' working-memory resources while they solve the probability versions should hardly affect their responses any further. However, if the frequency versions indeed trigger executive-demanding, rule-based processing, then the good performance on Bayesian inference problems with frequency formats should decrease under concurrent working-memory load (i.e., show a larger decrease than with standard probability formats). Note that the natural frequency accounts make the exact opposite prediction because they attribute the good performance on the modified versions to the recruitment of an automatically operating, module-based

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

reasoning process. Hence, if the natural frequency theorists are right and people rely on an automatic module to solve the frequency versions, burdening peoples' working-memory resources should not hamper performance.

In theory, studies that look at the correlation between cognitive capacity and use of base rates may also help to test the substitution claim. If people recruit resource-demanding, rule-based thinking to solve the frequency versions, one expects that individual differences in working-memory capacity (or general intelligence) will mediate performance. The more resources that are available, the more likely that the correct response will be computed. The operation of an automatic, encapsulated module, on the other hand, should not depend on the size of a general cognitive resource pool. Unfortunately, the little correlational data we have is not conclusive. In their extensive research program, Stanovich and West included two Bayesian inference problems (i.e., so-called noncausal base-rate problems) with a standard probability format (i.e., the notorious "Cab" and "Aids" problems based on Casscells et al. 1978 and on Bar-Hillel 1980). In contrast to the bulk of their findings with other reasoning tasks, Stanovich and West did not find any systematic correlations with cognitive capacity measures.[1] However, they did not look at correlations between cognitive capacity and performance on the crucial frequency format versions. Brase et al. (2006) did observe that students in top-tier universities solved frequency versions better than students in lower-ranked universities. If we simply assume that students in higher-ranked universities have a higher working memory capacity, the Brase et al. findings might present support for B&S's position. However, this conclusion remains tentative since Brase et al. did not explicitly test their participants on a standardized cognitive capacity measure.

In summary, B&S present an innovative model that highlights the central role of the set structure of a problem in the base-rate neglect phenomenon. I have tried to show that the model leads to some clear-cut predictions that might help to validate the framework in future studies. Unfortunately, the bad news is that in the absence of such processing validation, the central substitution claim of B&S's dual-system account remains questionable. Based on the available evidence, one needs to refrain from strong claims about the involvement of a rule-based reasoning system as the key mediator of base-rate "respect."

ACKNOWLEDGMENT

The author's research is supported by the Fund for Scientific Research Flanders (FWO-Vlaanderen).

NOTE

1. B&S write that Stanovich and West (1998a; 2000) found correlations between intelligence and use of base rates. It should be clarified that these studies concerned problems where statistical aggregate information is plotted against a concrete case (i.e., so-called causal base-rate problems) and not the type of Bayesian inference problem (i.e., so-called noncausal base-rate problems) on which the debate focuses.

Dual-processing explains base-rate neglect, but which dual-process theory and how?

DOI: 10.1017/S0140525X07001720

Jonathan St. B. T. Evans[a] and Shira Elqayam[b]

[a]*Centre for Thinking and Language, School of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom;* [b]*School of Applied Social Sciences, Faculty of Health and Life Sciences, De Montfort University, Leicester LE1 9BH, United Kingdom.*

jevans@plymouth.ac.uk selqayam@dmu.ac.uk

http://www.plymouth.ac.uk/pages/dynamic.asp?page=staffdetails&id=jevans&size=l

Abstract: We agree that current evolutionary accounts of base-rate neglect are unparsimonious, but we dispute the authors' account of the effect in terms of parallel associative and rule-based processes. We also question their assumption that cueing of nested set relations facilitates performance due to recruitment of explicit reasoning processes. In our account, such reasoning is always involved, but usually unsuccessful.

Barbey & Sloman (B&S) argue that evolutionary accounts of base-rate neglect are unparsimonious, especially with regard to the facilitatory effects of frequency formats. We agree. They also propose that the phenomena can be best accounted for with a dual-processing framework. Here we also agree, in general, but not with the specific dual-processing account offered by these authors. This is based upon Sloman's (1996a) proposal of two parallel systems for reasoning, one *associative* and one *rule-based*. The impression that B&S give, that this theory is representative of dual-process models of reasoning and judgement in general, is not correct.

Parallel-form dual process theories occur commonly where authors propose two forms of *knowledge*. This applies in dual-process accounts of learning (Reber 1993; Sun et al. 2005) where implicit and explicit learning processes lead to something like associative neural nets and propositional rules, respectively. This form of theory is also common in social psychology (see Smith & DeCoster 2000) where researchers are often concerned to distinguish implicit stereotypes and attitudes, inferable from behaviour, from their explicit counterparts that are verbally stated. Other dual-process theories of reasoning and judgement, however, have a sequential form in which the implicit processing is *pragmatic* rather than associative and serves to *contextualise* explicit analytic thinking. This is the type of account offered by Stanovich (1999) and by Evans (2006). In these theories, System 2 monitors default intuitions arising in System 1 and may intervene with more effortful and abstract reasoning (see also Kahneman & Frederick 2002).

We actually think it quite plausible, as Cosmides and Tooby (1996) argue, that both humans and other animals would have evolved mechanisms for processing frequency information about their environments. What we find implausible is that any such cognitive module would be applied to quantitative word problems of the kind presented in experiments on Bayesian inference. In particular, we note that advocates of the frequency-processing module do not ask their participants to engage in experiential learning of frequencies, but present them with collated frequencies instead. Under such circumstances, people can only apply general-purpose procedures for reasoning by construction and manipulation of mental models. As B&S demonstrate in their review, this will only be successful when people are cued to construct models which encode the nested set relationships.

We can find no evidence in the target article for the authors' assertion that base-rate neglect is due to *associative* processing. What association is involved? That between the hypothesis and the diagnostic information, presumably – which would be a similar account to the one given in terms of representativeness (Kahneman & Tversky 1973). However, although the majority tendency may be to base an answer on diagnostic information, a substantial minority give the base rate as the answer instead (Cosmides & Tooby 1996; Evans et al. 2000). Also, when real-world beliefs rather than stated probabilities are used to convey Bayesian priors, they may have much stronger influence than the diagnostic data on the judgments made (Evans et al. 2002, Experiment 5). What is common to all these cases is a failure to *integrate* base rate and diagnostic information, which should be equally weighted. Only when set-inclusive mental models are cued, will analytic reasoning be successful.

In this article, B&S clearly imply that application of System 2 reasoning will lead to *correct* responding. For example, they state

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

that, "An appropriate representation can induce people to substitute reasoning by rules with reasoning by association" (sect. 4, para. 2). There has been a tendency for authors in the reasoning and judgement area to equate heuristic responding with biases and analytic reasoning with normatively correct responding, but this is a mistake that has been corrected in more recent dual-process accounts of reasoning (Evans 2007; Stanovich 2006). In these new accounts, biases can occur in System 2 as well as System 1. The problem is that no psychologically plausible definition of System 2 reasoning – say, as slow, effortful, and engaging executive working memory – can immediately imply that it will be successful in finding normatively correct solutions. Even if it is correct to think of such reasoning as "rule based," people may access rules that are normatively unsound, or fail to apply them in an appropriate manner. At best we could say that System 2 reasoning is necessary for logical reasoning; we certainly could not say that it is sufficient.

The literatures on reasoning and decision making are marked by confusion between competence or computational level accounts of reasoning and accounts based on normative rationality (see Elqayam 2007). Even if we compare people of high and low cognitive ability, it is not universally the case that higher ability participants give more normatively correct answers. Recent experiments, for example, have shown that the logically valid *modus tollens* in inference may be endorsed *less* often by participants with higher ability (Newstead et al. 2004). More directly relevant in the current context is the report by Stanovich and West (1998b) that participants who incorporated base rates had *lower* intelligence scores than participants who neglected them. It is clearly impossible to reconcile the following propositions: (a) System 2 necessarily produces normatively correct responses; (b) attention to base rate is a normatively correct response, and (c) System 2 is more strongly employed by those of higher ability. Of these three statements, (b) holds by definition, and there is overwhelming empirical evidence for (c). The only possible conclusion is that (a) is false.

In our view, the literature on base-rate neglect is amenable to a dual-processing account, but not of the kind proposed by B&S. We suggest that university student participants, instructed to engage in reasoning, will in fact do so but with varying degrees of conformity to Bayesian inference. Without formal training they will have no access to the rules of Bayesian inference and can therefore only attempt to use general-purpose analytic reasoning procedures which involve constructing and manipulating mental models to represent the problem information. With standard presentations, it will appear to them that either base rate or else (more commonly) diagnostic information is relevant, but they will have no means of integrating the two. Only when problem design cues them to construct set-inclusive mental models will they succeed in computing the normatively correct solution.

Enhancing sensitivity to base-rates: Natural frequencies are not enough

DOI: 10.1017/S0140525X07001732

Edmund Fantino and Stephanie Stolarz-Fantino

Department of Psychology, University of California San Diego, La Jolla, CA 92093-0109.

efantino@ucsd.edu **sfantino@psy.ucsd.edu**

Abstract: We present evidence supporting the target article's assertion that while the presentation of base-rate information in a natural frequency format can be helpful in enhancing sensitivity to base rates, method of presentation is not a panacea. Indeed, we review studies demonstrating that when subjects directly experience base rates as natural frequencies in a trial-by-trial setting, they evince large base-rate neglect.

In many studies of base-rate use (or neglect), human participants are asked to judge the likelihood of an event on the basis of information about past occurrences (base rates) and present diagnostic case cue information. Base rates and case cues are presented in a tabulated statistical format. As Barbey & Sloman (B&S) point out in the target article, and as has been found by various investigators, subjects are less likely to neglect base-rate information when data presentation is transparent with respect to the set relations involved. Natural frequency presentation is transparent in this way. In fact, in daily life humans experience base rates in terms of natural frequencies; however, they experience them one example at a time. We thought that an ideal way to assess people's sensitivity to base-rate information would be with a matching-to-sample (MTS) procedure (Stolarz-Fantino & Fantino 1990). In the typical MTS procedure selection of the comparison stimulus that matches the sample is reinforced. But the procedure may be modified to vary the "accuracy" of the sample, that is, the degree to which it predicts the correct answer and the base rates (the proportion of correct answers assigned to each of the two comparison stimuli). This procedure is analogous to Tversky and Kahneman's "taxicab problem" (Tversky & Kahneman 1982a) and other problems of similar type. In the MTS analog, the sample stimulus corresponds to the witness in the taxi problem, or the case-cue information in other base-rate neglect problems; the probabilities of reinforcement for selecting the comparison stimuli correspond to the base rates, or incidence of taxi types.

Our procedure was simple. The sample in a MTS task was either a blue or green light. After the sample was terminated, two comparison stimuli appeared: these were always a blue and a green light. Participants were instructed to choose either. We could present subjects with repeated trials rapidly (from 150 to 400 trials in a less than one-hour session, depending on the experiment), and we could readily manipulate the probability of reinforcement for selecting either color after a blue sample and after a green sample. In one condition, the blue and green samples were equi-probable. Following a blue sample, selection of the blue comparison stimulus was reinforced on 67% of trials and selection of the green comparison stimulus on 33% of trials. Following a green sample, selection of the green comparison stimulus was reinforced on 33% of trials and selection of the blue comparison stimulus on 67% of trials. In other words, the sample in this case had no discriminative (or informative) function, just as the witness testimony has no function in the cab problem when the witness is 50% accurate. If our participating college students were responding optimally, they should have come to select the blue comparison stimulus on every trial, regardless of the sample color, thereby obtaining reward on 67% of trials. In this condition, participants showed a huge base-rate neglect, matching green on 56% of trials. In fact, our human participants showed significant base-rate neglect over hundreds of trials in this condition and in several others, in a series of studies conducted primarily by Adam Goodie (e.g., Goodie & Fantino 1995; 1996; 1999). In contrast, Hartl and Fantino (1996) found that pigeons selected optimally in a comparable MTS task, with no evidence of base-rate neglect.

What might account for the drastic difference in the behavior of pigeons and college students? We have speculated that humans have acquired strategies for dealing with matching problems that are misapplied in our MTS problem (e.g., Stolarz-Fantino & Fantino 1995). For instance, from early childhood we learn to match like shapes and colors at home, in school, and at play (e.g., in picture books and in playing with blocks

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

and puzzles). If base-rate neglect is a learned phenomenon, we should be able to eliminate it by using sample stimuli that are physically unrelated to the comparison stimuli. Therefore, we repeated our earlier experiment (Goodie & Fantino 1995) with a MTS procedure in which the sample stimuli were line orientations and the comparison stimuli were again the colors blue and green. This change eliminated base-rate neglect in keeping with the learning hypothesis (Goodie & Fantino 1996). Instead, participants' choices were well described as probability matching. To further assess the learning hypothesis, we next introduced a MTS task in which the sample and comparison stimuli were physically different but were related by an extensive history: The samples were the words "blue" and "green"; the comparison stimuli the colors blue and green. A robust base-rate neglect was reinstated. These and other experiments led us to conclude with some confidence that pre-existing associations contribute to the tendency to ignore or underweight base-rate information. Our human participants were not sensitive to the frequencies of reinforced choices, whereas pigeons, unfettered by prior associations, were appropriately sensitive to the same frequencies. However, Fantino et al. (2005) have shown that pigeons will also neglect base-rate information if they are given sufficient prior MTS training with a 100% reliable sample – that is, a training history more similar to that of human subjects.

Whether or not we have evolved to process information in terms of frequencies, there is evidence that trial-by-trial presentation of information can be difficult for human participants to process. For example, Jenkins and Ward (1965) found that participants' judgments of contingency were more related to the number of successful trials (Response 1 followed by Outcome 1) than they were to the actual degrees of association between responses and outcomes. A second study (Ward & Jenkins 1965) showed that, when statistical information was presented in summary form, many more participants based their contingency judgments on logical rules (75%) than when they received trial-by-trial information (17%).

Stolarz-Fantino et al. (2006) found that even with base-rate story problems (similar to the taxicab problem) in which statistical information was presented in a probability format, human participants attended appropriately to base-rate information under certain conditions. For example, when they made likelihood judgments on a series of problems, their estimates of likelihood varied appropriately with base-rate and case-cue values; this was not the case when participants judged single problems. And when the case-cue information was unreliable (as when the "witness" was described as being correct 50% of the time), many participants ignored the case cue in favor of the base rate.

Presumably, trial-by-trial experience emphasizes associative processes (System 1), whereas information presented in tabulated statistical form can cue associative and/or rule-based (System 2) activity. A challenge for future research will be to learn more about how information learned through experience becomes the subject of rule-based reasoning. Like Goodie and Fantino's (1996) participants, people typically have more immediate experience with case-cue information than with base rates; appreciating the effect of base rates usually means integrating information over a period of time. In tasks involving tabulated statistical information, base-rate and case-cue information are available simultaneously; however, in life, as pointed out by Fiedler (2000), the base rates of many important events are unknown. Therefore, past experience may lead people to put more emphasis on case cues even when base-rate information is available.

As the target article suggests, there is more to performance on base-rate problems than the question of whether information is presented in the form of probabilities or frequencies. In fact, other variables may dwarf the effects of presentation.

Ecologically structured information: The power of pictures and other effective data presentations

DOI: 10.1017/S0140525X07001744

Wolfgang Gaissmaier,[a] Nils Straubinger,[a] and David C. Funder[a,b]

[a]*Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany;* [b]*Department of Psychology, University of California, Riverside, CA 92521.*

gaissmaier@mpib-berlin.mpg.de
www.abc.mpib-berlin.mpg.de/users/gaissmaier
straubinger@mpib-berlin.mpg.de
www.abc.mpib-berlin.mpg.de./users/straubinger#funder@ucr.edu
www.rap.ucr.edu

Abstract: The general principle behind the effects of nested sets on the use of base rates, we believe, is that the mind is prepared to take in "ecologically structured information." Without any need to assume two cognitive systems, this principle explains how the proper use of base rates can be facilitated and also accounts for occasions when base rates are overused.

A picture speaks a thousand words. Barbey & Sloman (B&S) demonstrate this ancient principle with research showing how Euler circles efficiently convey complex information and clarify relationships among nested sets. Many other examples would be possible; tree diagrams, Venn diagrams, pie-charts, bar graphs, and other familiar data representations frequently appear in publications ranging from the most revered scientific journals to the pages of *USA Today*. As every successful speaker, writer, teacher, or advocate knows, some ways of presenting information are especially effective. The best way to teach an abstract concept may be via concrete examples. Certain pictures as well as natural frequencies appear to be especially useful ways to convey complex information. According to Pinker, "graphic formats present information in a way that is easier for people to perceive and reason about. However, it is hard to think of a theory or principle in contemporary cognitive science that explains why this should be so" (Pinker 1990, p. 73). We will introduce a concept that might help to answer this question, while also helping to explain why some diagrams support problem solving and others do not (see Larkin & Simon 1987).

The general principle is that the human perceptual and cognitive system uses certain kinds of information readily because it has impressive evolved and learned capacities for pattern recognition and automatic categorization. Hence, people can try to reason through an intricate system of interlocking sets to determine a probability, *or* they can look at certain diagrams and see the result literally at a glance. Similarly, natural frequencies and other well-chosen representations can vastly simplify problems and calculations that would otherwise be difficult, if not impossible. As Gigerenzer and Hoffrage (1995) observed, "Cognitive algorithms, Bayesian or otherwise, cannot be divorced from the information on which they operate and how that information is represented" (p. 701). What effective representations have in common is that they are *ecologically structured*, a term we derive from the concept of *ecological rationality*, which stresses the match between the human mind and the environment (Gigerenzer et al. 1999). Ecologically structured information can simplify apparently complex problems because it fulfills this match: It is presented in a manner that exploits human capacities to recognize relations in certain representations of complex problems (e.g., pictures and frequency counts). It is information received in the same way that people have evolved to receive information over the millennia: through vividly sensed images and experiences, and via specific instances, rather than abstract descriptions.

There is no need to posit two cognitive systems, as B&S do, to explain the advantages of ecologically structured information,

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

and it is a serious mistake to describe the first of these systems as something "to overcome" (sect. 4) and the second as the ideal of sound reasoning. Ironically, if the pattern recognition capacities exploited by Euler circles were assigned to one of these putative systems, it would seem more reasonable to base them in the first, more perceptually based one. B&S nearly acknowledge this point when they conclude that "appropriate representations can induce people to substitute reasoning by rules with reasoning by association" (sect. 4, para. 2).

It is disappointing, therefore, that B&S rely on the traditional and simplistic dichotomy between "abstract reasoning = good" and "heuristics = bad." A vast amount of evidence demonstrates how heuristics can make us smart (e.g., Gigerenzer et al. 1999), and, in particular, how fast and frugal processes of the sort B&S would call "associative" are an important part of good performance. As Chase and Simon (1973) commented, concerning chess experts: "One key to understanding chess mastery, then, seems to lie in the immediate perceptual processing, for it is here that the game is structured" (p. 56). Other examples include recognizing a visual pattern (e.g., faces), listening to music, and acquiring a first language. Furthermore, thinking more "extensionally" may even lead one astray, depending on the environment (for an example from the domain of probability learning, see Gaissmaier et al. 2006).

B&S characterize prior research on the importance of "ask(ing) about uncertainty in a form that naïve respondents can understand" as "far too narrow" (sect. 1, para. 1). However, their article focuses on a single phenomenon, the underuse of base rates in probabilistic reasoning. Yet, early research on Bayesian inference observed the opposite phenomenon: *conservatism*, the overuse of base rates (e.g., Edwards 1968). A second phenomenon that can be seen as the opposite of base-rate neglect is described in the social psychological literature on *stereotypes*. Both base rates and stereotypes comprise beliefs about the prevalence of a characteristic in a population. But the current literature on base rates generally concludes that such beliefs are underused, whereas the stereotype literature almost uniformly concludes that they are *over*used (Funder 1996).

We believe that the concept of ecologically structured information can reconcile these two seemingly opposed phenomena. In studies of stereotypes, the belief about the population is vivid, accessible, and perhaps even emotionally tinged (e.g., the racial stereotype held by a bigot). The factual information opposed to that stereotype, by contrast, is typically rather pallid (e.g., crime rate statistics). In studies of base rates, the opposite pattern holds. The specific case information is vivid (e.g., a woman with a positive mammogram), while the base rate is pallid (e.g., the prevalence of disease in her demographic group; see Funder 1995; 1996, for a more complete analysis).

A picture speaks a thousand words, and ecologically structured information can communicate complex situations efficiently and clearly because it exploits elementary perceptual and cognitive capacities. The implications range far beyond the putatively uniform underuse of base rates upon which B&S focus so tightly; indeed, this principle can help to explain the apparently opposite phenomenon, the cases in which base rates (in a literature where they are labeled as stereotypes) are overused.

The role of representation in Bayesian reasoning: Correcting common misconceptions

DOI: 10.1017/S0140525X07001756

Gerd Gigerenzer[a] and Ulrich Hoffrage[b]

[a]*Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany;* [b]*Ecole des Haute Etudes Commerciales (HEC), University of Lausanne, Batiment Internef, 1015 Lausanne, Switzerland.*

gigerenzer@mpib-berlin.mpg.de **Ulrich.Hoffrage@unil.ch**

Abstract: The terms *nested sets, partitive frequencies, inside-outside view*, and *dual processes* add little but confusion to our original analysis (Gigerenzer & Hoffrage 1995; 1999). The idea of nested set was introduced because of an oversight; it simply rephrases two of our equations. Representation in terms of *chances*, in contrast, is a novel contribution yet consistent with our computational analysis – it uses exactly the same numbers as natural frequencies. We show that non-Bayesian reasoning in children, laypeople, and physicians follows multiple rules rather than a general-purpose associative process in a vaguely specified "System 1." It is unclear what the theory in "dual process theory" is: Unless the two processes are defined, this distinction can account post hoc for almost everything. In contrast, an ecological view of cognition helps to explain how insight is elicited from the outside (the external representation of information) and, more generally, how cognitive strategies match with environmental structures.

For many years researchers believed that people are "not Bayesian at all" (Kahneman & Tversky 1972, p. 450) and that "the genuineness, the robustness, and the generality of the base-rate fallacy are matters of established fact" (Bar-Hillel 1980, p. 215). In 1995, however, we showed that Bayesian reasoning depends on and can be improved by external representations (Gigerenzer & Hoffrage 1995). This ecological approach led to practical applications in medicine, law, and education; natural frequency representations are now part of evidence-based medicine, high-school mathematics textbooks, and cancer-screening information brochures, helping people to understand risks (Gigerenzer 2002; Hoffrage et al. 2000).

Our 1995 article was about the general question of how various external representations facilitate Bayesian computations, not about natural frequencies versus single-event probabilities, as Barbey & Sloman (B&S) suggest. It contained four main predictions (Gigerenzer & Hoffrage 1995, pp. 691–92):

Prediction 1: Natural frequencies (standard frequency formats) elicit a higher proportion of Bayesian algorithms than standard probability formats do.

Prediction 2: Short probability formats elicit a higher proportion of Bayesian algorithms than standard probability formats do.

Prediction 3: Natural frequencies, whether in the standard or short format, elicit the same proportion of Bayesian algorithms.

Prediction 4: Relative frequencies elicit the same (small) proportion of Bayesian algorithms as standard probability formats do.

These predictions follow from Equations 1 to 3 in Gigerenzer and Hoffrage (1995). If information is presented in the standard probability format or in normalized (relative) frequencies, then the following computations are necessary (H = hypothesis, D = data):

$$p(H\,|\,D) = p(H)p(D\,|\,H)/[p(H)p(D\,|\,H) + p(-H)p(D\,|\,(-H)] \quad (1)$$

If information is instead represented in natural frequencies (standard or short format), then Bayesian computations reduce to:

$$p(H\,|\,D) = a/(a+b) \quad (2)$$

Here, a and b are natural frequencies. If probabilities are presented in short probability format, then the computations reduce to:

$$p(H\,|\,D) = p(D\&H)/p(D) \quad (3)$$

B&S mistakenly present (i) experiments reporting facilitation with probability representations (as in our Prediction 2) and (ii) experiments finding no facilitation with relative frequencies (exactly our Prediction 4) as if these were contradicting or going beyond our position, without making any mention of our Predictions 2, 3, and 4. The upshot is that the "nested set structure" explicit in our Equations 2 and 3 – the observation that the numerator is a subset of the denominator – is then presented as a new, alternative explanation. The predictions in B&S's Table 2 are based on the erroneous idea that our computational analysis was restricted to natural frequencies, as is the claim in their Table 1 that our computational analysis was only about a "cognitive

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

process uniquely sensitive to natural frequency formats." In the remainder of this comment, we will clarify the key ideas for the reader.

What are natural frequencies? Our Figure 1 shows the differences between natural and normalized frequencies. Natural frequencies leave the naturally occurring base rates intact, whereas normalized frequencies standardize these. Note, first, that all natural frequencies have a "nested set structure" in the sense that they simplify Bayesian computations, as defined in Equation 2. Hence, when B&S talk of "natural frequency formats that were not partitioned into nested set relations" (sect. 2.4, para. 2), these are not natural frequencies but instead normalized frequencies. This conceptual confusion makes the notion of nested sets appear as a different and broader explanation when it in fact simply paraphrases Equation 2. Second, natural frequencies refer to joint events, such as H&D events, as shown by the four numbers at the bottom of Figure 1–1. It is the structure of the entire tree that distinguishes natural from normalized frequencies. In contrast, an isolated frequency statement, represented by one single branch in the tree (such as 10 out of 1,000), could be part of a tree with natural frequencies, or normalized frequencies, or – if there is no second variable – no tree at all. Therefore, it is misleading to call the isolated statement "one of every 100 Americans will have been exposed to Flu strain X" (Table 5 of the target article) a natural frequency, as B&S do. In the same table caption, the relative frequency "33% of all Americans" is wrongly called a "single-event probability." This incorrect use of terms causes B&S to draw erroneous conclusions, such as that "natural frequencies and single-event probabilities are rated similarly in their perceived clarity, understandability … [etc.]" (sect. 2.10).

Next, the term "single-event probability" is irrelevant to our computational analysis (see Equations 1–3). A single-event probability can refer to at least three different concepts: a conditional probability $p(D|H)$, which makes Bayesian computations difficult (Prediction 1 and Equation 1), a joint probability $p(D\&H)$, which makes Bayesian computations easier (Prediction 2 and Equation 3), and a simple single-event probability, such as a "30% chance of rain," which has nothing to do with Bayesian inference but invites misunderstandings, because, by definition, no reference class is specified (Gigerenzer et al. 2005).

B&S's distinction between a "natural frequency algorithm," "natural frequency heuristic," and a "non-evolutionary natural frequency heuristic" is emphatically not ours. We cannot see how these would lead to different predictions, since in each case the algorithm computes Equation 2. We recommend not using the term *heuristic* for a version of Bayes's rule, since a heuristic, like a shortcut, ignores information. However, the term *heuristics* applies to shortcuts that approximate Bayes's rule under specific conditions such as rare events, where they lead to fast and frugal Bayesian reasoning (Table 1). Martignon et al. (2003) analyzed the connection between natural frequency trees and fast and frugal trees.

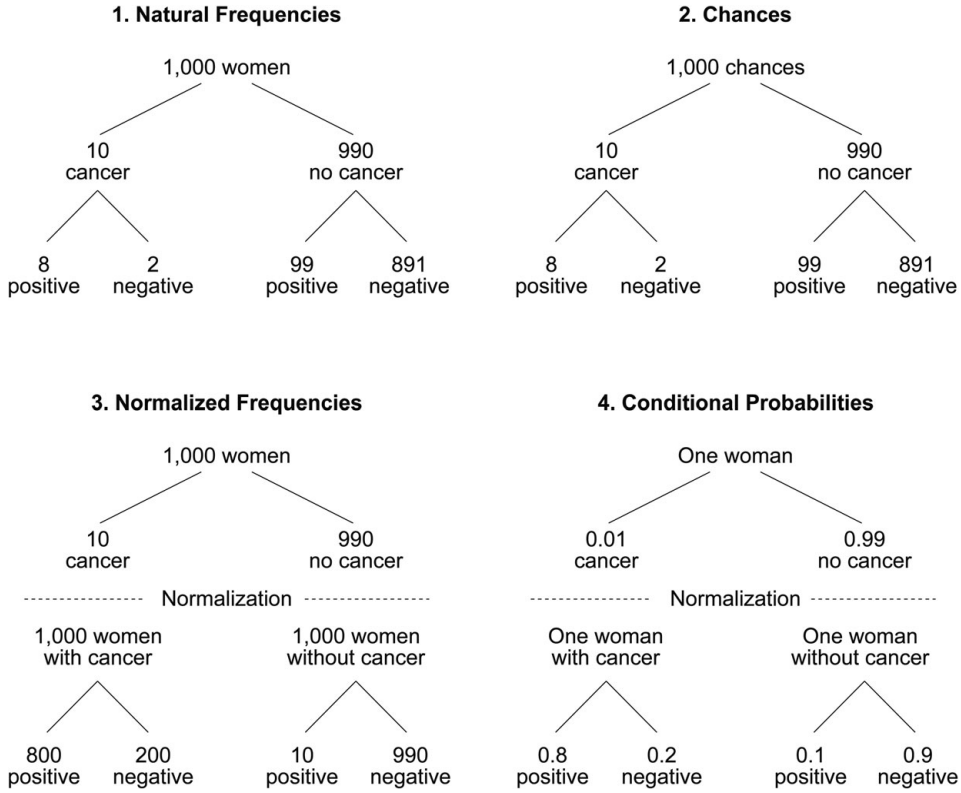


Figure 1 (Gigerenzer & Hoffrage). Natural frequencies, chances, normalized frequencies, and conditional probabilities. Note that B&S's "chances" are exactly the same numbers as natural frequencies and lead to identical computational demands (see Eq. 2). Contrary to B&S's interpretation, chances are *not* mathematical probabilities, since these cannot be normalized over the interval [0,1] – otherwise, chances would no longer facilitate Bayesian computations. The fact that "chances" refer to a single event does not transform them into mathematical probabilities: not all statements about singular events are probabilities. Normalized frequencies are derived from natural frequencies by normalizing the base rate frequencies to some common number (here: 1,000), and conditional probabilities normalize to the interval [0,1]. Note that our distinction is neither that between frequencies versus probabilities nor that between natural frequencies versus single-event probabilities, as B&S suggest; we distinguish between natural frequencies which facilitate Bayesian computations and normalized frequencies and conditional probabilities which do not.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Table 1 (Gigerenzer & Hoffrage). *Bayesian strategies and cognitive shortcuts for approximating Bayes' rule. Based on the experimental evidence in* Gigerenzer *and Hoffrage (1995, pp. 689–691).* n(D&H) *is the natural frequency of* D&H *cases. We suggest that Barbey & Sloman consider these rules as mechanisms for their System 2, to be interpreted as an adaptive toolbox rather than a single, general-purpose calculus*

Strategy/Shortcut	Formal Equivalent	Conditions in which the shortcut is ecologically rational
Conditional Probability Representation		
Bayesian Strategy	$p(H)p(D \mid H) \,/\, [p(H)p(D \mid H) + p(-H)p(D \mid -H)]$	
Rare-Event Shortcut	$p(H)p(D \mid H) \,/\, [p(H)p(D \mid H) + (D \mid -H)]$	$p(H)$ is rare and $p(-H)$ thus approaches 1
Big Hit-Rate Shortcut	$p(H)/[p(H) + p(-H)p(D \mid -H)]$	$p(D \mid H)$ is very large approaches 1
Comparison Shortcut	$p(H)p(D \mid H) \,/\, p(-H)(D \mid -H)$	$p(D\&-H)$ is much larger than $p(D\&H)$
Quick-and-Clean Shortcut	$p(H)/p(D \mid -H)$	All 3 conditions above
Natural Frequency Representation		
Bayesian Strategy	$n(D\&H)/[n(D\&H) + n(D\&-H)]$	
Comparison Shortcut	$n(D\&H)/n(D\&-H)$	$n(D\&-H)$ is much larger than n$(D\&H)$
Pre-Bayes	$n(H)/[n(D\&H) + n(D\&-H)]$	$n(D\&H)$ is close to $n(H)$

B&S repeatedly refer to our evolutionary argument that natural sampling characterizes the way people learned individually in human history. But we did not – nor can one – use this general argument to derive Predictions 1 to 4 or the seven results reported in our 1995 article; these derivations were based solely on a computational analysis. The evolutionary perspective, however, provides a general framework for finding the right questions. Instead of asking what cognitive deficits explain reasoning that deviates from Bayes's rule (such as an error-prone System 1), the question should be how and why reasoning depends on the external representation of information. An ecological framework postulates that thought does not simply emerge inside the mind. Every theory of reasoning needs to specify both cognitive strategies and the environmental structures under which these strategies work well (just as with the shortcuts in Table 1).

The "nested sets" explanation originated from an oversight. The authors credited by B&S as the originators of the "nested set theory" missed the distinction between natural and normalized frequencies, and implied that we had predicted that any kind of frequencies would facilitate reasoning. For instance, Johnson-Laird et al. (1999, p. 81) stated: "In fact, data in the form of frequencies by no means guarantee good Bayesian reasoning," and referred to a study reporting that normalized frequencies showed no facilitation. Since mental models theory cannot account for the facilitating effect of natural frequencies or "chances" (we discuss this further on), Johnson-Laird et al. introduced a "subset principle" identical to our 1995 Equation 2, without mentioning its source, and presented it as an alternative explanation to ours.

Macchi and Mosconi (1998) seem to have been the first who confused natural frequencies with any kind of frequencies and concluded that the facilitating effect is not due to "frequentist phrasing" (which they mistook as our explanation) but to computational simplification (our explanation, which they proposed as their alternative one). Like Johnson-Laird et al., Macchi (2000) independently rediscovered the proper explanation, and distinguished between "partitive" and "non-partitive" representations, where "partitive" – like the "subset principle" – is a new label for Equations 2 and 3. Lewis and Keren (1999) promoted the same confusion. In Gigerenzer and Hoffrage (1999), we pointed out that we had actually tested Prediction 4 about relative frequencies with 24 Bayesian problems in Experiment 2 of Gigerenzer and Hoffrage (1995). Nevertheless, Evans et al. (2000) embraced the same misconception, concluding that "we are not convinced that it is frequency information per se which is responsible for the facilitation" (p. 200). All of these authors overlooked that our predictions were not about frequencies per se.

To summarize, the "nested set theory" originated from an oversight that reproduced itself like a meme through various articles. It is identical to our Equations 2 and 3, rephrasing the computational explanation we had proposed.

What is new about the "chances representation"? In our 1995 article, we tested two natural frequency representations, three relative frequency representations, and three probability representations. One of the probability representations had the structure of Equation 1, another the nested structure defined by Equation 3, and a third one demanded computations of in-between complexity (Equation 4 in our article). Therefore, B&S's contention that "nested sets" would be more general than our computational account – because it covers not only frequencies but probabilities as well – ignores that we actually applied the computational account to various probability representations. Specifically, B&S present a "chances representation," which mimics the computational structure of natural frequencies precisely (see our Fig. 1), but is verbally phrased in terms of a single event. This representation is a new addition to the eight representations we already tested, and it leads to the same computational demands as in Equation 2. Hence, from our computational analysis, the prediction is that "chances" facilitate as well as natural frequencies because they involve exactly the same computations (although the occasionally odd-sounding wording may have a negative impact).

B&S call chances "single-event probabilities." However, like natural frequencies, these are not probabilities. Mathematical probabilities have a range between 0 and 1. If chances were expressed in this range, their facilitating effect would be gone (like the conditional probabilities in Fig. 1). In the example B&S give, one cannot express the chances "12 out of 96" as "1 out of 8" or .125, because chances are exactly like natural frequencies in that they do not allow normalization. To summarize, "chances" are the same numbers as natural frequencies and lead to the same computational demands specified in Equation 2. The "nested sets" notion does not seem to add anything further.

What processes underlie non-Bayesian judgments? B&S's answer is: the associative "System 1." Yet we have taken a closer look at non-Bayesian judgments and found that a substantial proportion of them follow several rules rather than one associative process. Specifically, 65% of all non-Bayesian

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Table 2 (Gigerenzer & Hoffrage).*Six cognitive rules underlying non-Bayesian judgments. Values are percentages of people classified as using a rule among all non-Bayesian judgments. The experiments with children (grades 4, 5, and 6) were conducted by Zhu & Gigerenzer (2006), with laypeople (students with median age 21–22) by Gigerenzer & Hoffrage (1995), with medical students (median age 25) by Hoffrage et al. (2000), and with physicians by Hoffrage & Gigerenzer (1998). Cognitive rules are reported here for natural frequencies and conditional probabilities (standard format) only; for other representations and how rules depend on representations, see the original studies. We suggest that Barbey & Sloman consider these rules as mechanisms for their System 1, to be interrupted as an adaptive toolbox rather than a single, general-purpose associative process*

		Conditional Probabilities			Natural Frequencies			
Cognitive Rule	Formal Equivalent	Psychology Students	Medical Students	Physicians	Children	Psychology Students	Medical Students	Physicians
Joint occurrence	$p(H\&D)$	10.7	2.5	1.2	0.0	8.3	9.8	3.8
Fisherian (Sensitivity)	$p(D\&H)$	27.2	17.7	20.9	2.9	22.8	4.9	9.6
Positives Only	$p(D)$	0.0	0.0	0.0	7.3	0.0	7.3	17.3
Pre-Bayes	$p(H) / [p(H\&D) + p(-H\&D)]$	0.0	0.0	0.0	18.3	5.4	2.4	0.0
Likelihood substraction (ΔR)	$p(\mathrm{D\backslash H}) - p(D\,|\,-H)$	8.2	12.7	23.3	0.0	1.7	0.0	9.6
Base rate only (Conservatism)	$p(H)$	1.6	3.8	1.2	8.5	5.4	29.3	28.8
Other non-Bayesian strategies		19.5	32.9	19.8	0.0	19.5	22.0	5.8
Not identified		32.7	30.4	33.7	63.1	36.9	24.4	25.0
Total of non-Bayesian (in %)		100.0	100.0	100.0	100.0	100.0	100.0	100.0

judgments across children, laypeople, and experts resulted from applying a rule, and our Table 2 shows the six most frequent ones. These rules allow for a better understanding of non-Bayesian reasoning than does the notion of base-rate neglect due to "System 1." In fact, one of these rules, base-rate only (conservatism), does not even entail base-rate neglect, but an over-reliance on the base rate. Moreover, a strategy such as the Fisherian one (or representativeness, which amounts to calculating p-values) ignores more than the base rate, namely, also p(D|-H). Ironically, when researchers use Fisher's null hypothesis tests to determine whether people follow Bayes's rule, they themselves use a non-Bayesian framework and commit base-rate neglect. Does this mean that researchers' "System 1" is in charge of hypothesis testing? In summary, there is experimental evidence that a substantial proportion of non-Bayesian judgments result from six rules; there is no reason to ignore these results and invoke some unknown general-purpose associative process instead.

What does the dual-processes notion explain?. Table 2 indicates that a handful of rules model non-Bayesian judgments. In general, people rely on multiple cognitive rules or heuristics, consciously or unconsciously, tending to switch between these in an adaptive way. Models of these heuristics and the environments in which they work have been published (e.g., Gigerenzer 2004; Payne et al. 1993; Rieskamp & Otto 2006). What does a distinction between a "System 1" and "System 2" add?

Sloman (1996a) proposed two systems of reasoning. Gigerenzer and Regier (1996) responded that there is a certain amount of slack in this distinction, that it collapses too many different dichotomies, and that it needs be sharpened by overt reference to explicit models of associative and rule-based processing. Sloman (1996b) willingly admitted that he left room for further precision and clarity in his dual-processes notion. Yet more than ten years later, the notion is still vague. What is the mechanism of "System 1": the delta rule, fuzzy set theory, fast and frugal heuristics, constrained neural networks, or something else? Since B&S assume a general-purpose process, there should be only one. And what is the nature of the rule-based system: first-order logic, Bayes's rule, signal-detection theory, or expected utility maximization? It cannot be all of these, since they are not the same. What do we gain from a dual-processes theory that does not develop a theory about the two processes?

Talking of two systems has become popular in some quarters. The "inside-outside view" is another case in point. According to Kahneman and Lovallo (1993, p. 25), an inside view focuses on "the case at hand," whereas an outside view focuses "on the statistics of a class of cases." Yet this distinction is too crude. For instance, it fails to predict the differential effect of natural versus normalized frequencies (Prediction 4), given that both invoke an "outside view," as well as the differential effects of various single-event representations, such as in Prediction 2, which all invoke an "inside view." B. F. Skinner asked us to refrain from building theories of cognition and to treat the mind as a black box. B&S's dual-systems notion is dangerously similar to two black boxes. What about replacing the two black boxes by an adaptive toolbox that contains multiple heuristics and logical tools?

Towards ecological rationality. In their title, B&S include the term *ecological rationality*. We have introduced this term to refer to the study of how cognitive processes map onto environmental structures. The Bayesian algorithms and shortcuts are part of this larger enterprise. It extends to heuristics that solve problems ranging from categorization to choice to inference, and from catching fly balls to making coronary care unit allocations or moral judgments (Gigerenzer 2007; Gigerenzer et al. 1999). The study of ecological rationality requires computational models of cognitive processes, in order to predict where they fail and succeed. It may actually help define the notion of dual processes more precisely.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

How to elicit sound probabilistic reasoning: Beyond word problems

DOI: 10.1017/S0140525X07001768

Vittorio Girotto[a] and Michel Gonzalez[b]

[a]*Department of Arts and Industrial Design, University IUAV of Venice, Convento delle Terese, 30123 Venice, Italy;* [b]*Laboratory of Cognitive Psychology, University of Provence and CNRS, Centre St Charles, 13331 Marseilles, France.*

vittorio.girotto@iuav.it **http://www.iuav.it/English-Ve/Department/dADI—dep/Faculty-te/Vittorio-G/index.htm**

michel.gonzalez@up.univ-mrs.fr

http://www.up.univ-mrs.fr/document.php?pagendx=3614&project=lpc

Abstract: Barbey & Sloman (B&S) conclude that natural frequency theorists have raised a fundamental question: What are the conditions that compel individuals to reason extensionally? We argue that word problems asking for a numerical judgment used by these theorists cannot answer this question. We present evidence that nonverbal tasks can elicit correct intuitions of posterior probability even in preschoolers.

Barbey & Sloman's (B&S's) comprehensive analysis of the literature shows that, given an accurate set representation, naive individuals are able to evaluate posterior probability by performing elementary set operations on various information types (e.g., natural frequencies or numbers of chance). We agree with B&S's conclusion that natural frequency theorists have raised a deep and fundamental question: What are the conditions that compel individuals to reason extensionally? However, we argue that natural frequency theorists have not used the appropriate methodology to answer this question.

It is ironic that the hypothesis "the mind is a frequency monitoring device" has been tested almost exclusively by means of complex word problems in which numerical information is conveyed by symbols. There are three reasons to doubt the suitability of this approach. First, verbal expressions of frequencies are potentially ambiguous. For instance, in some reputedly natural frequency problems, the statements are stated in the future tense: "8 out of every 10 women with breast cancer will get a positive mammography" (Gigerenzer & Hoffrage 1995). The point is that statements of this sort express expected rather than observed frequencies, and could be correctly interpreted as statements about probabilities. For example, the above statement could be understood as follows: "Women with breast cancer have 8 chances in 10 of getting a positive mammography." To make this point clearer, consider another statement that expresses expected frequencies: "This coin will land heads up one out of two times." Given that it is stated in the future tense, this statement does not mean that an individual has tossed the coin twice and observed that one of the times it landed heads up, but rather, that the coin has a priori one out of two chances of landing heads up. If readers interpret predicted frequencies as probabilities, answers based on information understood as chance data could be mistaken for answers based on natural frequencies. Second, Bayesian reasoning concerns the revision of probability in the light of new evidence. However, the standard problems used in the literature typically ask for a single judgment, rather than for two successive judgments, the first one based on prior information and the following one on posterior information. Hence, these problems do not investigate the ways in which individuals actually revise their judgment in light of new information. Third, word problems asking for a numerical judgment cannot be used with individuals lacking basic verbal and numerical skills, such as young children. In sum, despite their common use, standard word problems are not the best tool to seriously test general hypotheses about the nature of human judgment.

Consider a situation in which participants are presented with a bag containing four round chips (all black) and four square chips (three white and one black). Before the experimenter draws a chip from the bag, participants have to make a prior bet on the drawing of a black versus a white chip. After this first bet, the experimenter draws a chip and, keeping it in her hand, informs participants that it is square, and asks them to make another, posterior bet. A task of this sort does not present the aforementioned three weaknesses. First, it does not have the potential ambiguities of verbal statements: it certainly asks individuals to reason about a set of prior possibilities, not about observed frequencies in a series of actual draws. Second, it requires participants to update their choice in light of new evidence. Third, it can be used to investigate whether young children, who cannot tackle complex word problems, are nonetheless able to use new evidence for evaluating an uncertain event.

Testing whether children possess some intuitions of posterior probability may sound paradoxical, given the difficulties of adults' reasoning discussed in B&S's review. But, if reasoning about uncertain events depends on the application of elementary set operations, then even children should be able to solve posterior probability tasks, at least from the age at which they are able to compare and add quantities. Indeed, Girotto and Gonzalez (in press, Study 1) have shown that from the age of about five, children perform correctly in the chip task. As found in previous studies (e.g., Brainerd 1981), children first answered "black," by reasoning about the initial set of possibilities in which there were five black and three white chips. Then they correctly updated their initial choice, by considering the subset of possibilities compatible with the new piece of information (i.e., the four squares). In sum, preschoolers are able to apply correct extensional procedures in reasoning about the random events produced by a chance device. But they do so even when they have to reason about a single, not repeatable event produced by an intentional agent (Girotto & Gonzalez, in press, Study 3). For example, children were presented with two boxes, each containing three animals (two cats and one dog vs. two dogs and one cat). The experimenter informed children that a troll secretly put one chocolate in the bag of one animal. Children had to choose a box in order to find the animal with the chocolate. There was no optimum choice, given that prior evidence did not favor one box over the other. After they made their choice, children were informed that a cat carried the chocolate and were asked to make a new choice. As predicted by the extensional reasoning hypothesis, children passed the test: even children who initially did not choose the more advantageous box, now chose the box favored by posterior evidence.

In sum, humans may be "developmentally and evolutionarily prepared to handle natural frequencies" (Gigerenzer & Hoffrage 1999, p. 430). However, they are not blind to single-case probability. Even preschoolers correctly draw posterior probability inferences about single events in nonverbal tasks asking for a choice or a non-numerical judgment. And they do so in the same situations in which adults succeed – that is, when they have to make a simple enumeration of possibilities.

ACKNOWLEDGMENT

Preparation of this commentary was funded in part by a COFIN grant (2005117840_003) from the Italian Ministry of Universities.

Frequency formats are a small part of the base rate story

DOI: 10.1017/S0140525X0700177X

Dale Griffin,[a] Derek J. Koehler,[b] and Lyle Brenner[c]

[a]*Sauder School of Business, University of British Columbia, Vancouver, BC V6T 1Z2, Canada;* [b]*Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1, Canada;* [c]*Warrington College of Business, University of Florida, Gainesville, FL 32611.*

dale.griffin@sauder.ubc.ca **dkoehler@watarts.uwaterloo.ca**

lbrenner@ufl.edu

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Abstract: Manipulations that draw attention to extensional or set-based considerations are neither sufficient nor necessary for enhanced use of base rates in intuitive judgments. Frequency formats are only one part of the puzzle of base-rate use and neglect. The conditions under which these and other manipulations promote base-rate use may be more parsimoniously organized under the broader notion of case-based judgment.

Although we agree that the two-system nested set account provides a better fit to the data reviewed in the target article than the alternative frequency-format accounts, we believe that the nested set account is an overly narrow lens through which to view base-rate use and its relation to probability and frequency judgments. In particular, manipulations making nested set representations more transparent may not be sufficient to improve base-rate use and such manipulations are not necessary to improve base-rate use. In terms of the dual systems model, base-rate use is not improved solely by rule-based processes, nor is base-rate neglect always driven by associative processes. By focusing only on areas where frequency formats increase base-rate use, the target article oversells the value of frequency formats – and rule-based or System 2 processes more generally – in improving intuitive judgment.

A case-based judgment account built on Kahneman and Tversky's early theorizing (e.g., Kahneman & Tversky 1973) provides a perspective on intuitive judgment that is compatible with yet broader than the nested set account. The case-based account provides a parsimonious explanation of patterns of base-rate use and neglect across both probability reasoning tasks and experience-based probability judgments, and also provides a more realistic view of the debiasing value of frequency formats. According to the case-based account, intuitive judgments focus on assessing the strength of evidence relevant to the current case at hand (Brenner et al. 2005; Griffin & Tversky 1992). Strength of evidence is commonly evaluated by associative processes such as similarity or fluency, but can also be evaluated by rule-based processes. However, to the extent that both associative and rule-based processes focus on the strength of impression favoring a particular hypothesis about the current case, background evidence about class or extensional relations is not included when the strength of evidence is mapped onto a probability (or related) scale. This produces neglect of base rates, as well as neglect of cue validity in intuitive judgments.

According to the case-based account, any evidence that influences the strength of impression regarding the case at hand will affect probability judgment. This explains why base rates that can be interpreted (associatively, via System 1 processes) as a propensity of the single case are highly influential. Racial or gender stereotypes, for example, can be interpreted as base rates but also can yield a strong expectation about a particular individual. Similarly, the win-loss record of a sports team can yield an impression of the strength of that team (Gigerenzer et al. 1988). The debate about "causal" base rates can also be interpreted in this way (Tversky & Kahneman 1980). When provided with a statistical summary of the number of blue versus green cabs in a city, people rely on the testimony of a fallible accident witness and disregard the base rate; however, when base rates are given a causal significance by describing the differential likelihood of accidents for the cabs, both the witness's testimony and the accident-proneness of cabs contribute to the strength of impression for this particular accident. In these contexts, the use of base rates per se does not indicate a System 2 rule-based process.

Furthermore, improved judgment resulting from a diagram or other aid to viewing a problem in terms of nested sets does not necessarily implicate rule-based reasoning. Diagrams prompting an immediate comparison of the size of circles may allow a low-level perceptual computation to solve the problem. If wording or outcome formats allow a judge to represent such relationships visually or symbolically, the line between associative and rule-based solutions becomes blurred. From the perspective of the case-based account, such manipulations may operate through their impact on the case-specific impression of evidence strength. The results of the Girotto and Gonzalez (2001) study described in the target article could be interpreted in this manner.

According to the evolutionary frequency module account, "our hunter-gatherer ancestors were awash in statistical information in the form of the encountered frequencies of real events: in contrast, the probability of a single event was inherently unobservable to them" (Cosmides & Tooby 1994, p. 330). In several recent studies (Brenner et al. 2005; 2006), we have examined probability judgment in a learning paradigm similar to the Gluck and Bower (1988) study described in the target article. In this simulated stock market study, case-specific evidence is provided in terms of a company's sales and costs. A participant's task is to estimate the probability that the stock price will increase, given the financial information and experience in the market which provide evidence about the base rate of stock increases and the validity of financial cues. Notably, participants were extremely accurate in estimating the base rates that they had experienced. However, despite this – and despite being awash in encountered frequencies – participants' probability judgments were largely unaffected by base rates or cue validity. When juxtaposed with case-specific information, apparently, such extensional considerations can be readily available, yet be viewed as largely irrelevant to the judgment. A more evolutionarily grounded outcome measure would assess the resources that an individual is willing to commit to a decision based on uncertain evidence. A natural measure is thus the price one is willing to pay for a stock certificate for a particular company. When price is used as an outcome measure in our learning paradigm, however, the neglect of base rate and cue validity remains.

Barbey & Sloman (B&S) offer a helpful reappraisal of the impact of frequency representations on base-rate use in probability reasoning tasks. We agree that the evidence clearly does not support the strong claim that frequency formulations yield effortless Bayesian reasoning. The view that base-rate use proceeds only or primarily through application of rules of set inclusion, however, may also be too strong. On the one hand, Bayesian solution rates are far from perfect when set relations are explicitly highlighted (see Table 4 of the target article). On the other hand, under the right circumstances, base rates may be used effortlessly, if they are captured in the immediate impression of the strength of evidence regarding the case at hand.

One wrong does not justify another: Accepting dual processes by fallacy of false alternatives

DOI: 10.1017/S0140525X07001781

Gideon Keren,[a] Iris van Rooij,[a] and Yaacov Schul[b]

[a]*Department of Technology Management, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands;* [b]*Department of Psychology, The Hebrew University, Mount Scopus, Jerusalem 91905, Israel.*

g.b.keren@tue.nl **i.v.rooij@tue.nl** **yschul@huji.ac.il**

Abstract: Barbey & Sloman (B&S) advocate a dual-process (two-system) approach by comparing it with an alternative perspective (ecological rationality), claiming that the latter is unwarranted. Rejecting this alternative approach cannot serve as sufficient evidence for the viability of the former.

The target article's title suggests two messages to take home. Current theories of ecological rationality rest on weak grounds (we generally agree), and data patterns of base-rate neglect provide empirical support for dual-process theory (we generally disagree). Barbey & Sloman's (B&S's) analysis is mistaken on two grounds.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

First, they commit the fallacy of false alternatives: Demonstrating that 4 out of 5 theoretical accounts are false does not necessarily imply the truth of the remaining one unless the list of hypotheses is exhaustive, which it is not (we label the 5 accounts considered by B&S as T1–T5). T1–T5 do not instantiate an exclusive set of (plausible) theoretical possibilities. This becomes evident when analyzing T5 into its conjunctive parts: (1) the hypothesis that explicating nested set relations facilitates Bayesian reasoning (T-NESTED), and (2) the hypothesis that the mind has a dual-process architecture with associative processing occurring in parallel with rule-based processing (T-DUAL). Clearly, T-NESTED and T-DUAL are two distinct and separable theoretical claims (hence we already have T1–T6 different accounts, with T6 being equal to T-NESTED without T-DUAL). There is no reason we can see – nor do the authors provide one – for the nested set hypothesis to be married specifically to a dual-process architecture of mind. It seems as plausible (a priori at least) that a single-, or multi-process architecture can implement the benefits from nested set representations.

This leads directly to B&S's second fallacy: Even if the non-rejected account (nested sets) has some merits, in no way does it imply or support a dual process (two-systems) perspective. It is well known that representation and computation can trade spaces, in the sense that computation can be facilitated, or otherwise affected, by changing between (logically equivalent) representational formats (Clark & Thornton 1997; Marr 1982). This general cognitive principle has been demonstrated in areas as diverse as problem solving, memory retrieval, and visual imagery. Also, the cognitive facilitation afforded by Venn diagrams, and diagrams in general (Larkin & Simon 1987), is well known (yet, unrelated to dual process theories). Framing effects in decision making also illustrate how changes in representational format affect cognitive judgments. The nested sets facilitation hypothesis, reported by B&S, seems to be yet another (potential) example.[1] As such, the hypothesis, though viable, is neither novel nor surprising. Because of its generalized flavor it seems particularly ill-suited as a basis for conjecturing a particular architecture of mind: almost any architecture of mind (whether single-, dual- or multi-process; whether associative, rule-based, both or neither) could accommodate the effect.

Apparently, B&S do consider evidence in favor of the nested set hypothesis as also constituting support for the idea that human minds have a dual-process architecture. Arguing for such a general theoretical position, based on the available performance data alone, is simply trying to do the impossible. This is also illustrated by the target article's Table 2. Close inspection of the table shows that the available data cannot decide between theories that assume modular or non-modular architectures (predictions for T1 and T2 are identical), and cannot decide between theories postulating evolutionary or non-evolutionary adaptations (predictions for T3 and T4 are identical). In the same vein, the available data cannot decide between theories that postulate dual- or single-process architectures. Table 2 may seem to suggest otherwise because the predictions of T5 appear to be unique. However, it should be noted that the table is missing a column and thus is incomplete. The authors should have included a sixth column listing predictions for T6 identical to the predictions for T5 (granting that T5 is really making the listed predictions – which seems questionable to begin with, yet is insubstantial for our claim that the reviewed findings cannot discriminate between T5 and T6). Including such a sixth column may have highlighted that the dual process assumption is superfluous in the authors' explanation of base-rate neglect.

Here B&S are confronted with the fact that theoretical frameworks in science generally cannot be justified on the basis of a small set of empirical phenomena (Lakatos 1977).[2] Rather, theoretical frameworks derive their explanatory power from making insightful a large corpus of seemingly unrelated findings that would otherwise be puzzling or anomalous. B&S make no attempt to argue for the explanatory superiority of dual-process architectures (compared to other architectures of mind); and as we have argued, effects of representational format (e.g. nested set relations) on cognitive processing are not puzzling in any event.

In short, B&S do not provide any argument for why support for the nested-set hypothesis constitutes evidence for dual-process (two-systems) theories. The presumed superiority of dual-process architectures is presumably established by citing other authors who advocate a two-systems theory (e.g., Evans & Over 1996; Kahneman & Fredrick 2002; Sloman 1996a; Stanovich & West 2000). Indeed, there has recently been an upsurge in theoretical frameworks alluding to the existence of two different processing systems that supposedly operate according to different rules. Recently, we (Keren & Schul, under review) have pointed to the lack of robust and reliable evidence that would support the two-systems architecture of the mind. The target article seems to offer arguments that question the viability of the natural frequencies approach, and more generally the ecological rationality framework. Yet, it does not add any forceful evidence in support of the alternative favored by the authors, namely the dual-process approach. The possibility that both theoretical frameworks (i.e., ecological rationality and dual processes) are undefendable, cannot be ruled out.

NOTES

1. B&S's attempt to rule out the possibility that explicating nested set relations simply affords easier computation is questionable. They draw on a study asking participants to judge ease of understanding of different presentation formats. Whether participants have introspective access to the nature and efficiency of their own cognitive processes is highly doubtful (Nisbett & Wilson 1977).

2. Certainly when the phenomenon under discussion remains controversial (Koehler 1996) on both theoretical and empirical grounds.

Implications of natural sampling in base-rate tasks

DOI: 10.1017/S0140525X07001793

Gernot D. Kleiter

Department of Psychology, Salzburg University, A-5020 Salzburg, Austria.
gernot.kleiter@sbg.ac.at

Abstract: The hypothesis that structural properties and not frequencies per se improve base-rate sensitivity is supported from the perspective of natural sampling. Natural sampling uses a special frequency format that makes base-rates redundant. Unfortunately, however, it does not allow us to empirically investigate human understanding of essential properties of uncertainty – most importantly, the understanding of conditional probabilities in Bayes' Theorem.

Barbey & Sloman (B&S) disentangle and systematize the various explanations of base-rate neglect/facilitation. They present strong arguments in favor of the hypothesis that the nested subset structure is responsible for facilitation effects. My comments try to further clarify the implications of natural sampling. Throughout the article, the authors adopt the terminology of "natural frequencies" used by Gigerenzer and his group. The adjective "natural" was transferred from "natural sampling." Let's therefore start with the origin of the latter concept.

The notion "natural sampling" was introduced by Aitchison and Dunsmore (1975) in their excellent book on statistical prediction analysis. In estimating probability parameters, frequencies are informative if and only if they are the outcome of a random sampling process and there is no missing data. Sampling is non-natural if, for example, sample sizes are planned by an experimenter. I used the term "natural sampling" in the Bayesian analysis of binomial sampling (Kleiter 1994) in the technical sense of Aitchison and Dunsmore. For several Bernoulli

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

processes and beta prior distributions for the binomial probability parameters, natural sampling results in a mathematically nice property: The posterior distribution turns out to be also a beta distribution, and, most important in the present context, the distribution does not depend on the between-group frequencies (base rates) but only on the frequencies within each group (e.g., hit and false alarm rates). As the between-group frequencies are estimators of base-rate probabilities, this result describes a situation in which the base rates are irrelevant. Under properly defined natural sampling conditions, base-rate neglect is rational. The result was generalized to multinomial sampling in combination with Dirichlet priors (Kleiter & Kardinal 1995) and used to propagate probabilities in Bayesian networks (Kleiter 1996). Recently, Hooper (2007) has shown that some claims about the generality of beta posteriors in Bayesian networks made in my 1996 paper are only asymptotically valid.

If base rates are irrelevant in a "normative" model, then base-rate neglect in psychological experiments is not necessarily an error but may be rational. If Bayes' Theorem is written in "frequency format," even elementary school math shows that the base rates in the numerator and in the denominator get cancelled when the within-group frequencies add up to the between-group frequencies. This property fitted extremely well within Gigerenzer's approach. In the early 1990s when Gerd Gigerenzer was at Salzburg University, during one of the weekly breakfast discussions held among Gigerenzer and members of his group, the mathematical result of base-rate cancellation was communicated and it was immediately taken up and integrated into his work. Natural sampling requires random sampling, additive frequencies in hierarchical tree-like sample/subsample structure (i.e., complete data), and a few more properties that belong to the statistical model. The notion of "natural frequencies" seems, in addition, to involve sequential sampling and thus acquires an evolutionary adaptive connotation.

The additivity in natural sampling goes hand in hand with the subset structure, the favorite explanation in the target article. The close relationship between natural sampling and the subset structure may have led to a confounding of the two in the past. If frequencies (and not subset structures) are the cause of facilitation effects, then critical experiments should investigate non-natural sampling conditions (Kleiter et al. 1997). Frequencies should still have a facilitating effect. Unfortunately, instead of non-natural sampling conditions, often "single-case probabilities" are taken for comparison to demonstrate the base-rate facilitation with natural sampling conditions.

How common are natural sampling conditions in everyday life? I have severe doubts about the ecological validity and the corresponding evolutionary adaptive value. From the perspective of ecological validity, it is important that the base-rate neglect has often been demonstrated for categories with low prevalence, such as rare diseases. Consequently, the prevalence of base-rate neglect will also be low. Base-rate effect certainly depends upon the actual numbers used in the experiments, a property not discussed in B&S's review.

The cognitive system of an intelligent agent capable of uncertainty processing and judgment requires competence in at least six domains. (1) Perception and processing of environmental information, such as numerosity, cardinalities of sets, relative frequencies, descriptive statistics of central tendency, variability, and covariation. (2) Understanding of randomness, of not directly observable states, of alternatives to reality and hidden variables, of the non-uniformities in the environment, and of the limited predictability of events and states. (3) Introspection of one's own knowledge states, and weighting and assessing one's own incomplete knowledge by degrees of beliefs (subjective probabilities). (4) An inference engine that derives conclusions about the uncertainty of a target event from a set of uncertain premises. Typical inference forms are diagnosis, prediction, or explanation. The conclusions often concern single events. The probabilities can be precise or imprecise (lower and upper probabilities, or second order probability distributions). Recently, classical deductive argument forms have also been modeled probabilistically (Oaksford & Chater 2007; Pfeifer & Kleiter 2005). (5) Modeling functional dependencies/independencies which are basic to causal reasoning. (6) Understanding of the knowledge states of other persons – a prerequisite for the effective communication of uncertainty in social settings.

Many base-rate studies present frequency information (belonging to item [1] in the list given above) and observe whether the subjects use "Bayes' Theorem" as an inference rule (belonging to item [5]). Bayes' Theorem degenerates to a rule for cardinalities, formulated not in terms of probabilities but in terms of frequencies (see Note 2 in the target article). This can of course be done, but we should be aware that we are dealing with the most elementary forms of uncertain reasoning, not involving any of the other items listed above. Moreover, if the response mode requires frequency estimates and not the probabilities of single events, another important aspect of uncertain reasoning is lost. If subjects are poor in the judgment of single event probabilities they have an essential deficit in uncertainty processing.

Conditional events and conditional probabilities are at the very heart of probability theory. Correspondingly, the understanding of conditional events and conditional probabilities should be central to investigations on human uncertain reasoning. Considering base-rate tasks in natural sampling conditions alone, misses this point completely. The B&S structural subset hypothesis shows that conditional probabilities are not needed in this case, and that structural task properties are the main cause of facilitation effects.

Dual concerns with the dualist approach

DOI: 10.1017/S0140525X0700180X

David A. Lagnado and David R. Shanks

Department of Psychology, University College London, London WC1E 6BT, United Kingdom.

d.lagnado@ucl.ac.uk **d.shanks@ucl.ac.uk**

Abstract: Barbey & Sloman (B&S) attribute all instances of normative base-rate usage to a rule-based system, and all instances of neglect to an associative system. As it stands, this argument is too simplistic, and indeed fails to explain either good or bad performance on the classic Medical Diagnosis problem.

Barbey & Sloman (B&S) claim that an associative system is responsible for errors in a range of probabilistic judgments. Although this is plausible in the case of the conjunction fallacy (where a similarity-based judgment substitutes for a probability judgment), it is less applicable to the Medical Diagnosis base-rate problem. What are the automatic associative processes that are supposed to drive incorrect responses in this case? Respondents reach incorrect solutions in various different ways (Brase et al. 2006; Eddy 1982), many of which involve explicit computations. Indeed, the modal answer is often equal to one minus the false positive rate (e.g., 95% when the false positive rate is 5%). This clearly involves an explicit calculation, not the output of an implicit process. Thus, errors can arise from incorrect application of rules (or application of incorrect rules), rather than just brute association.

The key point here is that base-rate neglect in the Medical Diagnosis problem provides little evidence for the exclusive operation of an implicit associative system. Indeed, it is arguable that adherents of classical statistics are guilty of similar base-rate neglect in their reliance on likelihood ratios (Howson & Urbach 2006). Presumably this is not due to an implicit associative system, but is based on explicit rules and assumptions.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

What about the claim that the rule-based system is responsible for correct responses in frequency-based versions of the task? This hinges on the idea that representing the task in a frequency format alerts people to the relevant nested set relations, and thus permits the operation of the rule-based system. In one sense, this is trivially true – those participants who reach the correct answer can always be *described* as applying appropriate rules. But what are these rules? And what is the evidence for their use, as opposed to associative processes that could also yield a correct response?

B&S explicitly block one obvious answer – that once people are presented with information in a format that reveals the nested set structure, they use a simplified version of Bayes' rule to compute the final solution. The authors' reasons for rejecting this answer, however, are unconvincing. The cited regression analyses (Evans et al. 2002; Griffin & Buehler 1999) were performed on a different task. And they were computed on grouped data, so it is possible that those who answered correctly did weight information equally. Furthermore, it is wrong to assume that a Bayesian position requires equal weighting – in fact, a full Bayesian treatment would allow differential weights according to the judged reliability of the sources.

More pertinently, if people are not using the frequency version of Bayes' rule, what are they doing? How do they pass from nested set relations to a correct Bayesian answer? B&S offer no concrete or testable proposal, and thus no reason to exclude an associative solution. Why can't the transparency of the nested set relations allow other associative processes to kick in? It is question-begging to assume that the associative system is a priori unable to solve the task.

Indeed, there are at least two arguments that support this alternative possibility. First, our sensitivity to nested sets relations might itself rest on System 1 (associative) processes. When we look at the Euler circles, we simply "see" that one set is included in the other (perhaps this is why they are so useful, because they recruit another System 1 process?). Second, it is not hard to conceive of an associative system that gives correct answers to the Medical Diagnosis problem. Such a system just needs to learn that the correct diagnosis covaries with the base rate as well as the test results. This could be acquired by a simple network model trained on numerous cases with varying base rates and test results. And a system (or person) that learned in this way could be described as implementing the correct Bayesian solution.

The dual-process framework in general makes a strong distinction between normative and non-normative behaviour. In so doing, it embraces everything and explains nothing. One simply cannot align the normative/non-normative and rule-based/associative distinctions. True, rule-based processes might often behave in accordance with a norm such as Bayes' theorem, and associative systems non-normatively (as in the example from Gluck & Bower 1988); but, as argued above, it is also possible for rule-based processes to behave irrationally (think of someone explicitly using an incorrect rule), and for associative systems to behave normatively (backpropagation networks are, after all, optimal pattern classifiers).

Moreover, we know that without additional constraints, each type of process can be enormously powerful. Imagine a situation in which patients with symptom A and B have disease 1, while those with symptoms A and C have disease 2, with the former being more numerous than the latter (i.e., the base-rate of disease 1 is greater). Now consider what inference to make for a new patient with only symptom A and another with symptoms B and C. Both cases are ambiguous, but if choice takes account of base-rate information, then disease 1 will be diagnosed in both cases. In fact, people reliably go counter to the base rate for the BC conjunction (hence the "inverse base-rate effect"), choosing disease 2, whereas they choose disease 1 for symptom A (Medin & Edelson 1988; Johansen et al., in press). Thus, in one and the same situation, we see both usage and counter-usage of base-rate information. But strikingly, these simultaneous patterns of behaviour have been explained both in rule-based systems (Juslin et al. 2001) and in associative ones (Kruschke 2001), emphasizing the inappropriateness of linking types of behaviour (normative, non-normative) to different processing "systems" (rule-based or associative).

The crux of B&S's argument, that a dual-process framework explains people's performance on probability problems, is unconvincing both theoretically and empirically. This is not to dismiss their critique of the frequentist program, but to highlight the need for finer-grained analyses. A crude dichotomy between the associative-system and the rule-based system does not capture the subtleties of human inference.

Ordinary people do not ignore base rates

DOI: 10.1017/S0140525X07001811

Donald Laming

University of Cambridge, Department of Experimental Psychology, Downing Street, Cambridge, CB2 3EB, United Kingdom.

drjl@hermes.cam.ac.uk

Abstract: Human responses to probabilities can be studied through gambling and through experiments presenting biased sequences of stimuli. In both cases, participants are sensitive to base rates. They adjust automatically to changes in base rate; such adjustment is incompatible with conformity to Bayes' Theorem. "Base-rate neglect" is therefore specific to the exercises in mental arithmetic reviewed in the target article.

When participants are asked to reason about statistical data, they tend to ignore base rates. But there is a problem with the experiments that the target authors do not address. A probability is a mathematical abstraction and cannot be presented as a stimulus (though it can be realised as a property of an otherwise random sequence of stimuli). The research reviewed in the target article substitutes values for probabilities and presents participants with exercises in mental arithmetic. "Base-rate neglect" might therefore be either the result of a failure to understand Bayes' Theorem, or due to insufficient ability in mental arithmetic. The authors do not enquire which.

The same question can be presented at different levels of difficulty. The mammography example from Gigerenzer and Hoffrage (1995), in probability format, requires participants to multiply 0.01 by 0.8 and (1-0.01) by 0.096 and then compare the two. In the frequency version, participants merely have to compare 8 with the sum of 8 and 95. It is not surprising that the latter version elicited a greater number of correct (Bayesian) answers. The results summarised in Table 3 of the target article suggest that a substantial proportion of incorrect answers are consequent on difficulties in mental arithmetic. Performance needs to be related to ability in mental calculation. But participants have been uniformly drawn from university populations and the results lack generality. Except for Brase et al.'s (2006) study, the matter of participants' prior education has been ignored.

The "probability" problem can be circumvented in two different ways. First, as in gambling: Gamblers – not gamblers doing mental arithmetic, not even in a casino (Lichtenstein & Slovic 1973), but real gamblers chancing their own real money – are sensitive to "base rates" that do not even exist! "Roulette players believe that certain numbers are due, when they have not come up for a long time" (Wagenaar 1988, p. 112). Gamblers do not reason rationally, else there would be no bookmakers or casinos in business. Moreover, the assessment of probability divides into at most five categories (Laming 2004, Ch. 16).

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

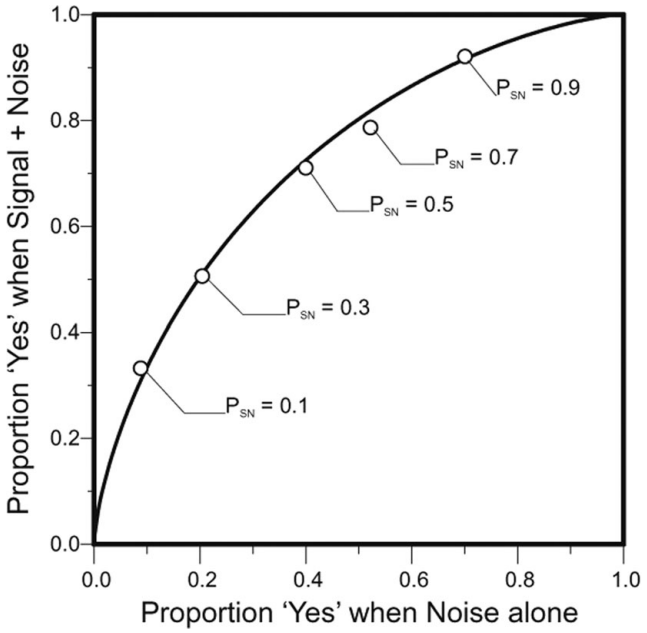


Figure 1 (Laming). Signal detection data from Tanner et al. (1956). (From *Signal Detection Theory and Psychophysics* by D. M. Green and J. A Swets, p. 88. © 1966, J. A Swets. Adapted with permission.)

The alternative is to realise probabilities as the relative frequencies of different stimuli in an otherwise random sequence. Figure 1 reproduces signal-detection data from Tanner et al. (1956). Different proportions of signal (SN) trials lead to different probabilities of detections and false-positives. Two-choice reaction times exhibit a similar phenomenon. The mean reaction times to two signals are different; the more frequent signal elicits a systematically faster response (see Laming 1968, Fig. 5.2). In 1994 the Public Health Laboratory Service (PHLS) in England introduced a saliva test for rubella. When a doctor notified the PHLS of a diagnosis, a kit was sent, and the swab, when returned, was tested for the disease. Figure 2 (right hand scale) shows the proportions of returned swabs that tested positive – this reflects the incidence of the disease – and (left hand scale) the numbers of notifications. Diagnoses of rubella followed the rise in incidence after a lag of eight weeks.

So, people do not ignore base rates. But do they update their expectations in line with Bayes' Theorem? Green (1960) calculated an optimum (Bayesian) placement of the criteria for the experiment in Figure 1, and found that the actual placements of the criteria did not vary so widely as Bayes' theorem prescribes. A much better prescription of criterion placement is provided by a scheme of probability matching suggested by Thomas and Legge (1970). Under this scheme, the numbers of different responses are adjusted to match the numbers of different stimuli, or, what is equivalent, there are equal numbers of errors of each kind, irrespective of the proportions of the different stimuli. The two schemes are compared in Laming (2001, Fig. 2).

An experiment by Tanner et al. (1970) reveals how this scheme is effected. These authors partitioned their data according to the event (detection, miss, correct rejection, false positive) on the preceding trial. This analysis showed that the effective operating point fluctuates; when there is no feedback errors can be seen to occur chiefly at extreme swings of the criterion (see Laming 2004, Fig. 12.4). But when feedback is supplied, performance on the trial following an error is different; that is, knowing one has just made an error generates a correction to the criterion. Given feedback (as in Fig. 1), participants adjust to the prevailing proportions of signal and noise trials by means of a substantial correction to criterion following each error. The effective criterion oscillates around a value at which the numbers (not proportions) of errors of each kind are equal. A second observer in the experiment by Tanner et al. (1956; see Green & Swets 1966, p. 95) displayed a highly asymmetric operating characteristic. Green's (1960) calculations no longer apply. But this second

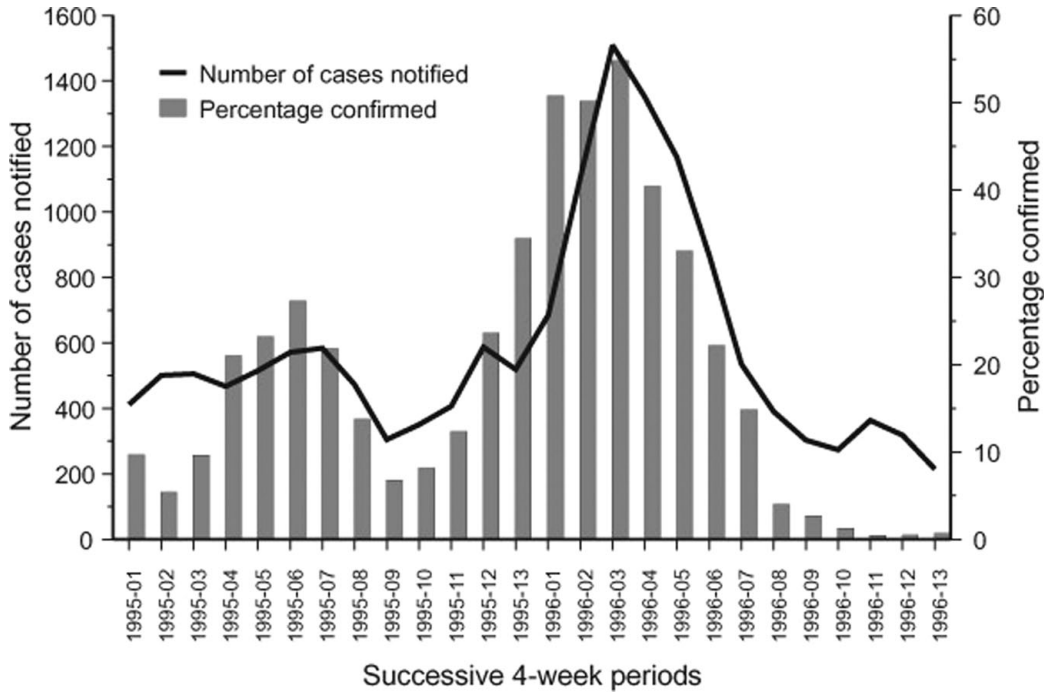


Figure 2 (Laming). Numbers of reported diagnoses of rubella and percentage confirmed. (Data from *Communicable Disease Report*, PHLS Communicable Disease Surveillance Centre, 1995–96.)

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

observer's data still showed approximate equality between the numbers of each kind of error, irrespective of the proportions of signal and noise trials (Laming 2001, Table 2). A similar relationship is found in two-choice reaction experiments; the variation of reaction time with signal probability can be reduced to a common scheme of sequential interactions operating on signal sequences of different composition (Laming 1968, Ch. 8).

To sum up: In real life people do not ignore base rates. "Base-rate neglect" is specific to exercises in mental arithmetic. People do not use Bayes' Theorem either. The nature of sensitivity to frequencies of events means that people adjust automatically to changes in base rate (e.g., Fig. 2), and automatic adjustment to changes in base rate is incompatible with the use of Bayes' Theorem itself. It follows that there is no reason why people should learn anything about Bayes' Theorem from natural experience – they learn only if they have (informal) lessons in probability theory. The research summarised in the target article tells us only about the prior education of the participants. It leads us astray in the matter of how people update their prior expectations.

The underinformative formulation of conditional probability

DOI: 10.1017/S0140525X07001823

Laura Macchi and Maria Bagassi

Department of Psychology, University of Milano-Bicocca, 20126 Milan, Italy.
laura.macchi@unimib.it **maria.bagassi@unimib.it**

Abstract: The formulation of the conditional probability in classical tasks does not guarantee the effective transmission of the independence of the hit rate from the base rate. In these kinds of tasks, data are all available, but subjects are able to understand them in the specific meanings proper to a specialized language only if these are adequately transmitted. From this perspective, the partitive formulation should not be considered a facilitation, but rather, a way of effectively transmitting the conditional probability.

> Consider the following two phrases:
>
> 1. The death-rate among men is twice that for women.
> 2. In the deaths registered last month there were twice as many men as women.
>
> Are these two different ways of saying the same or are these different events? In fact, they are different events. (Lindley 1985, p. 44)

The two phrases presented by Lindley describe two completely different probabilities which are connected to the same pair of events (P[Death/Men] and P[Men/Death]). In the first phrase, the probability is about the rate of mortality, given the gender; in the second one, the probability is about the rate of gender, given the mortality. Whereas the second phrase implies that $P(M/D)$ is equal to 2/3, the first does not.[1] The confusion between these two notions is a very common phenomenon, and has great implications for reasoning and decision making. According to Dawes (1988, p.80), "words are poor vehicles for discussing inverse probabilities." A main question is about the nature of this confusion and, consequently, the understanding and the use of the conditional probability.

We propose a pragmatic explanation of the phenomenon of the confusion between inverse probabilities and of the base-rate fallacy. In particular, concerning the kinds of problems considered in the literature on the base-rate fallacy, a pragmatic analysis of the texts/problems allowed us to identify, as responsible for the fallacy, the ambiguous formulation of a likelihood, instead of an intrinsic difficulty to reason in Bayesian terms (Macchi 1995; 2000; Macchi & Mosconi 1998).

Let us consider, for example, three kinds of problems formulated as follows:

> "If a woman has breast cancer, the probability is 80% that she will get a positive mammography." (Medical Diagnosis problem)
> "The percentage of deaths by suicide is three times higher among single individuals than among married individuals." (Suicide problem)
> "The witness made correct identifications in 80% of the cases and erred in 20% of the cases." (Cab problem)

This sort of formulation does not seem to express the intrinsic nature of a conditional probability, which conditions an event (A) to the occurrence of another event (B). Nor does this sort of formulation represent the cases in which, given the occurrence of B, A also occurs. In other words, it transmits just a generic association of events: A & B. For statistically naive subjects, this kind of formulation is not informative even if all the data were literally spelled out. The distinction between sentence and utterance is at the core of Grice's communication theory, according to which phrases imply and mean more than what they literally say (Grice 1975). What is implied is the outcome of an inferential process, in which what is said is interpreted in the light of the intentions attributed to the speaker and of the context (unavoidably elicited and determined by any communications).

Common language is ambiguous in itself. The understanding of what a speaker means requires a disambiguation process, centered on the attribution of intention. Differently, specialized languages (e.g., logical and statistical ones) presuppose an univocal, unambiguous interpretation (the utterance corresponds to the sentence).

The formulation of these kinds of problems uses common language. Data are all available, but subjects are able to understand them in the specific meanings proper to a specialized language, only if they are adequately transmitted. Then, the particular interpretation of the data, required for a correct solution, needs a "particularized," marked formulation (see Grice 1975; Levinson 1995; 2000).

In the sentences from Lindley (1985) quoted earlier, the effective transmission of the independence of the hit rate from the base rate does not seem guaranteed. This is a crucial assumption for proper Bayesian analysis (Birnbaum 1983), because the a posteriori probability $P(H/D)$ is calculated on the basis of the base rate and is therefore dependent upon it. If the hit rate depended on the base rate, it would already include it and, if this were the case, we would already have the a posteriori probability and it would be unnecessary to consider the base rate itself. This is what often underlies the base-rate fallacy, which consists of a failure to consider the base rate on account of the privileging of hit-rate information. In our view, the confusion sometimes generated between the hit rate and a posteriori probability is due to an unmarked formulation of conditional probability, or, in other words, to the absence of a partitive formulation.

If this is true, the partitive formulation should not be considered a facilitation, as Barbey & Sloman (B&S) argue, but a way of transmitting a particular information, able to translate a specialized language into a common, natural language, differently from the "step by step" question form used to compute the Bayesian ratio (adopted by Girotto & Gonzalez 2001). An example of partitive formulation of likelihood in the diagnoses problem is: "80 per cent of women who have breast cancer will get a positive mammography" (Macchi 2003).

The low performance with the Medical Diagnosis (MD) problem has usually been considered as evidence of the activation of System 1, which operates associatively (fast, automatic, effortless). Vice versa, the high performance with this kind of problem is ascribed to System 2, which is able to process rule-based inferences. However, in a recent study (Macchi et al.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

2007), statistically sophisticated subjects who solved the MD problem (66% of 35 subjects) were not able to solve the computationally less complex Linda problem (only 14% of the subjects did not commit the conjunction fallacy). So, the conceptual distinction between two reasoning systems, which explains the biases recurring in System 1 and the normative performance as the result of the activation of System 2, gives rise to some doubts.

According to us, the ability of statistically sophisticated subjects to grasp the informativeness of the data and the aim of the task in Bayesian tasks is a pragmatic ability. Also, when subjects don't give the logical-normative solution to the Linda problem, they are again considering the informativeness of the data, which, in this instance, hinders the intent of the experimenter (concerning the inclusion-class rule), because of a misleading contextualization of the task.

We could further speculate that, instead of having an ability for decontextualizing the task (Stanovich & West 2000), those gifted subjects who give the normative solution to the Linda problem (14%) would have a high ability to understand which context the experimenter intended, thereby revealing an interactional intelligence.

NOTE

1. Except when the number of men and women is the same.

Nested sets theory, full stop: Explaining performance on Bayesian inference tasks without dual-systems assumptions[1]

DOI: 10.1017/S0140525X07001835

David R. Mandel

Defence Research and Development Canada (Toronto), Toronto, ON M3M 3B9, Canada.

david.mandel@drdc-rddc.gc.ca

http://mandel.socialpsychology.org/

Abstract: Consistent with Barbey & Sloman (B&S), it is proposed that performance on Bayesian inference tasks is well explained by nested sets theory (NST). However, contrary to those authors' view, it is proposed that NST does better by dispelling with dual-systems assumptions. This article examines why, and sketches out a series of NST's core principles, which were not previously defined.

A voluminous literature documents the many shortcomings people exhibit in judging probability. Barbey & Sloman (B&S) focus on a subset of this research that explores people's abilities to aggregate statistical information in order to judge posterior probabilities of the form $P(H|D)$, where D and H represent data and a focal hypothesis, respectively. Some of this literature indicates that people neglect base rates, although some of the findings are consistent with other judgment errors, such as the *inverse fallacy* (Koehler 1996), which involves confusing $P(H|D)$ with $P(D|H)$. For instance, Villejoubert and Mandel (2002) observed that bias (i.e., systematic inaccuracy) and incoherence (i.e., nonadditivity) in posterior probability judgments was well explained by the inverse fallacy, even though base-rate neglect could not account for the observed performance decrements. Thus, there is some question regarding exactly how much of what has been called base-rate neglect *is* in fact base-rate neglect. A safer claim is that performance on such Bayesian inference tasks is often suboptimal and much of the error observed is systematic.

B&S challenge a set of theoretical positions oriented around the core notion that humans are better at judging probabilities when the information they are provided with is in the form of natural frequencies (Gigerenzer & Hoffrage 1995). They argue – convincingly, I believe – that variation in frequency versus probability formats neither explains away performance errors, nor does it account for errors as well as variation in the transparency of the nested set structure of an inference task. I shall not repeat their arguments here. Rather, my aim is, first, to sketch out some key propositions of nested sets theory (NST), which have yet to be described as a series of interlocking principles. Second, I will argue that NST would be on even firmer theoretical ground if the dual-systems assumptions that currently pervade B&S's version of it were jettisoned.

At its core, NST consists of a few simple propositions: First, performance on a range of reasoning tasks can be improved by making the partitions between relevant sets of events more transparent. I call this the *representation principle*. Second, because many reasoning tasks, such as posterior probability judgment, involve *nested* set relations, transparency often entails making those relations clear as well. I call this the *relational principle*. Third, holding transparency constant, nested set representations that minimize computational complexity will optimize performance. I call this the *complexity principle*. Fourth, the manner in which task queries are framed will affect performance by varying the degree to which default or otherwise salient representations minimize task complexity. In effect, this is the flip side of the complexity principle, and I call it the *framing principle*. Fifth, improvements in the clarity of nested set representations can be brought about through different modalities of expression (e.g., verbal description vs. visual representation). I call this the *multi-modal principle*. Sixth, within a given modality, there are multiple ways to improve the clarity of representation. I call this the *equifinality principle*. This list is almost certainly incomplete, yet it provides a starting point for developing a more explicit exposition of NST, which up until now has been more of an assemblage of hypotheses, empirical findings, and rebuttals to theorists proposing some form of the "frequentist mind" perspective. In the future, attempts to develop NST could link up with other recent attempts to develop a comprehensive theory of the representational processes in probability judgment (e.g., Johnson-Laird et al. 1999; Mandel, in press).

Although NST is not intrinsically a dual-systems theory (DST), B&S have tried to make it "DST-compatible." This is unfortunate for two main reasons. First, although DSTs are in vogue (for an overview, see Stanovich & West 2000) – perhaps because they offer a type of Aristotelian explanation long favored by psychologists (Lewin 1931) – they are not particularly coherent theoretical frameworks. Rather, they provide a rough categorization of the processes that guide reasoning and that influence performance through an effort-accuracy tradeoff. The second reason for preferring "pure NST" to an NST-DST hybrid is that the former is not only more parsimonious, it actually offers a better explanatory account. According to the hybrid theory, when the nested set structure of a reasoning task is unclear, people have difficulty applying the rigorous rule-based system (also called "System 2") and fall back on the more error-prone associative system (also called "System 1"). However, B&S say little about how judgment biases arise from those associative processes, or how system switching may occur.

Pure NST does not preclude the idea that impoverished representations of nested set relations can shift judgment towards a greater reliance on associative reasoning processes, but nor does it depend on that idea either. A viable alternative explanation is that impoverished representations lead to performance decrements because they increase one's chances of failing to access the correct solution to a problem. This does not necessarily mean that they switch to associative processes. It may simply mean that they fail to apply the correct principle or that they select the wrong information aggregation rule. Consider the inverse fallacy: It seems more likely that the error stems from a failure to understand how to combine $P(D \mid \neg H)$ with $P(D \mid H)$ (and with the base rates of H and $\neg H$ where they are unequal) than that it follows from use of associative processes.

Improving the representational quality of nested sets may also influence rule-based processes by simplifying the computations

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

required to implement a normative information aggregation strategy. Indeed, as B&S indicate, the performance decrements on Bayesian judgment tasks that Girotto and Gonzalez (2001) observed when participants were presented with "defective" (but nevertheless transparent) nested sets, appear to be attributable to the fact that such representations require at least one additional computational (subtraction) step. That computation itself may not be difficult to perform, but if it is missed the participant's judgment will surely be wrong.

In short, the types of errors that arise from impoverished representations of nested set relations are generally consistent with a rule-based system. NST should remain pure and single, unencumbered by a marriage to dual-systems assumptions.

NOTE

1. The author of this commentary carried out this research on behalf of the Government of Canada, and as such the copyright of the commentary belongs to the Canadian Crown and is not subject to copyright within the United States.

Naturally nested, but why dual process?

DOI: 10.1017/S0140525X07001847

Ben Newell and Brett Hayes

School of Psychology, University of New South Wales, Sydney 2052, Australia.
ben.newell@unsw.edu.au **b.hayes@unsw.edu.au**
http://www2.psy.unsw.edu.au/Users/BNewell

Abstract: The article by Barbey & Sloman (B&S) provides a valuable framework for integrating research on base-rate neglect and respect. The theoretical arguments and data supporting the nested set model are persuasive. But we found the dual-process account to be underspecified and less compelling. Our concerns are based on (a) inconsistencies within the literature cited by B&S, and (b) studies of base-rate neglect in categorization.

Why so low? A striking feature of the data reviewed by Barney & Sloman (B&S) is that the percentage of participants achieving the correct answer in the Medical Diagnosis problem rarely exceeds 50% (e.g., Table 3 of the target article). Thus, whether it is presenting information as natural frequencies or making nested set relations apparent that leads to improvements, overall the levels of performance remain remarkably low.

Potential reasons for this low level of overall performance are not discussed adequately in the target article. Although acknowledging in section 2.2 that "wide variability in the size of the effects makes it clear that in no sense do natural frequencies eliminate base-rate neglect" (para. 2), B&S fail to apply the same standard to their own proposal that "set-relation inducing formats" (be they natural frequencies or otherwise) facilitate a shift to a qualitatively different system of reasoning. The clear message of the article is that by presenting information appropriately, participants can "overcome their natural associative tendencies" (sect. 4, para. 3) and employ a reasoning system which applies rules to solve problems. Why does this system remain inaccessible for half of the participants in the studies reviewed? Is the rule system engaged, but the wrong rules are applied (e.g., Brase et al. 2006, Experiment 1)? Or do these participants remain oblivious to the nested sets relations and persevere with "inferior" associative strategies?

B&S cite evidence from studies of syllogistic reasoning, deductive reasoning, and other types of probability judgment in support of their contention that nested set improvements are domain-general. In these other tasks, however, the improvements are considerably more dramatic than in the base-rate studies (e.g., Newstead [1989] found a reduction in errors from 90% to 5% for syllogisms with Euler circle representations). The contrast between these large improvements in other domains and the modest ones in the base-rate neglect problems sits uncomfortably in a dual-process framework. Why doesn't the rule-based system overcome associative tendencies in similar ways across different tasks? In essence, the issue is one of specification – one needs to be able to specify what aspects of a problem make it amenable to being solved by a particular system for the notion of dual systems to have explanatory value. Why not simply appeal to "difficulty" and create a taxonomy or continuum of tasks differentially affected by various manipulations (diagrams, incentives, numerical format). Such a framework would not require recourse to dual processes or the vague rules of operation and conditions for transition that duality entails.

Incentives to "shift" systems? B&S report evidence from a study by Brase et al. (2006) in which it was found that monetary incentives improved performance on a base-rate problem. These data might be useful in gaining a clearer understanding of the factors that induce a shift between systems. It is difficult to make clear predictions, but under one interpretation of the dual-process position, one might expect incentives to have a larger effect in problems for which the set relations are apparent. The idea being that when the representation of information prevents (the majority of) people from engaging the rule-based system (e.g., when probabilities are used), no amount of incentive will help – most people simply won't "get it." A simple test of this hypothesis would be to compare probability and frequency incentive conditions. Brase et al. did not do this, comparing instead natural frequency conditions with and without additional pictorial representations. One would assume that the pictorial representations enhance nested set relations (target article, sect. 2.5) and increase the likelihood of shift to the rule-based system; hence, incentives would be more effective in the pictorial condition that condition. Brase et al.'s findings were telling: there was a main effect of incentives but this did not interact with the presence or absence of the picture; and indeed there was no main effect of providing a pictorial representation.

Two processes or two kinds of experience? In evaluating the B&S account we believe that it is useful to consider some of the lessons learned from the study of base-rate respect and neglect in category learning. In these studies people learn to discriminate between the exemplars of categories that differ in their frequency of presentation. The question is whether this base-rate information is used appropriately in subsequent categorization decisions, with features from more common categories being given greater weight. The results have been mixed. People can use base-rate information adaptively (Kruschke 1996), ignore it (Gluck & Bower 1988), or show an inverse base-rate effect, giving more weight to features from less frequent categories (Medin & Edelson 1988). Note that the issue of information format does not arise here, as all learning involves assessments of feature frequency. Critically, one does not need to invoke multiple processes to explain these results. Kruschke (1996) has shown that sensitivity and insensitivity to category base-rates can be predicted by a unitary learning model that takes account of the order in which different categories are learned, and allows for shifts of attention to critical features. In brief, people only neglect category-base rates when their attention is drawn to highly distinctive features in the less frequent category. The moral here is that before we resort to dual-process explanations of base-rate respect and neglect, we should first consider explanations based on the way that general learning mechanisms interact with given data structures.

Conclusion. B&S provide a very useful overview of the base-rate-neglect literature and provide convincing arguments for

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

questioning many of the popular accounts of the basic phenomena. The nested sets hypothesis is a sensible and powerful explanatory framework; however, incorporating the hypothesis into the overly vague dual-process model seems unnecessary.

The logic of natural sampling

DOI: 10.1017/S0140525X07001859

David E. Over

Psychology Department, Durham University, Science Laboratories, South Road, Durham City DH1 3LE, United Kingdom.

david.over@durham.ac.uk

Abstract: Barbey & Sloman (B&S) relegate the logical rule of the excluded middle to a footnote. But this logical rule is necessary for natural sampling. Making the rule explicit in a logical tree can make a problem easier to solve. Examples are given of uses of the rule that are non-constructive and not reducible to a domain-specific module.

Barbey & Sloman (B&S) have written a brilliant paper, but they are not explicit about the logic of the elementary set operations they appeal to. As Boolean algebra shows, this logic is the same as elementary propositional logic. The set operations of taking the complement, intersection, and union, which are necessary for natural sampling, are the same Boolean operations (up to isomorphism) as negation, conjunction, and disjunction in propositional logic. In addition to making this general point, I would give more prominence to the logical rule that the authors relegate to the target article's Note 3: the excluded middle. This rule states that all propositions of the form "*p* or not-*p*" are logically true. Its central place in natural sampling cannot be explained by the Swiss-army-knife model of the mind. This point reinforces B&S's criticisms of that model.

In the "mind as Swiss army knife" model, the mind consists of many domain-specific modules for reasoning and decision making. There is, for example, supposed to be a domain-specific module for inferences about cheaters in social arrangements (Cosmides 1989), as well as a separate module for natural sampling. The human mind is not supposed to have a content-independent logical ability to reason about cheaters, natural sampling, or other matters in general. Some supporters of this model even deny that logic is a normative standard for judging rationality, calling it useless "baggage" (Todd & Gigerenzer 1999, p. 365).

Logic is not, however, useless "baggage" for natural sampling problems, which absolutely depend on the logic of the set operations, including the rule of the excluded middle. Problems that appeal to this rule can be hard for people, but can become easier when the relevance of the rule is made explicit. B&S also refer to a difficult logical problem, called THOG, which becomes easier when its logical form is made transparent using logical trees that reveal nested sets (Griggs & Newstead 1982). Consider the following problem (from Levesque 1986) that is simpler to describe than the THOG set up, but which illustrates the same points:

> Jack is looking at Ann but Ann is looking at George. Jack is a cheater but George is not. Is a cheater looking at a non-cheater?
> A) Yes B) No C) Cannot tell

We could call this the "Ann problem." I have modified it to be about cheating, but it could have any content. It is a hard problem for most people. They respond with "Cannot tell," though the correct answer is "Yes "(Toplak & Stanovich 2002). The Ann problem is difficult because the relevance of the excluded middle rule is not transparent. It can, however, be made easier by adding the rule explicitly: that Ann is either a cheater or not a cheater. Logicians would say that the Ann problem requires *non-constructive* inference. The non-constructive step is the application of the excluded middle rule. Thanks to the rule, we can know, from "above," a priori and logically that Ann is either a cheater or not a cheater, although we never learn which she is in the reasoning – that is the non-constructive aspect. We cannot reduce this reasoning to constructive processing from "below" by a domain-specific module (Over & Evans, forthcoming). Such non-constructive inference is the purest example of rule-based thought in a dual-process model, either in the type that B&S endorse or in other types (Evans 2007).

The Ann problem could also be made easier by putting it into a logical tree form, which is so often used to make probability problems easier in natural sampling (Over 2007). The tree would begin with two branches, with one for the possibility of Ann as a cheater and the other for the possibility of Ann as a non-cheater. The tree would reveal that a cheater is looking at a non-cheater in either case. This form of the Ann problem would be closely analogous to a natural sampling problem. Kleiter (1994) used logical trees in his seminal work on natural sampling, and others have followed him in this, but not always with the realization that the trees are purely *logical* constructions (Kleiter realises this, but Zhu & Gigerenzer 2006 do not). Such a tree necessarily depends on the rule of the excluded middle.

It is challenging to think of a context in which knowing whether a cheater is looking at a non-cheater could tell us something important about cheating in the real world. With more space, we could discuss possibilities. In any case, it is clear that actual natural sampling could help us judge how far we could trust someone not to be cheater, depending on the number of times he has, or has not, cooperated with us in the past. B&S specify the conditions under which natural sampling can actually be useful and not merely an abstract exercise. Even then, natural sampling presupposes that we can apply the rule of the excluded middle to construct exhaustive subsets. True, we may find it impossible in a practical sense to apply the rule to a vague term, like "is depressed," rather than a precise one, like "has cancer." But in that case, actual natural sampling cannot be fully carried out. We could not actually complete the task of constructing exhaustive subsets of the depressed and not depressed. Since "is depressed" is vague, it has borderline cases: people we cannot classify as either being depressed or, alternatively, not being depressed. We cannot do without the excluded middle and other logical rules in natural sampling and in other general reasoning about cheating, probability, and many other matters. We cannot adopt a Swiss army model of the mind and call logic useless "baggage," while relying on its rules for natural sampling and constructing logical trees (as in Zhu & Gigerenzer 2006). The alternative is a dual-process theory that gives logic its proper place in our thought, which is limited but still necessary (Evans & Over 1996).

The versatility and generality of nested set operations

DOI: 10.1017/S0140525X07001860

Richard Patterson

Department of Philosophy, Emory University, Atlanta, GA 30306.

rpatt01@emory.edu

Abstract: The target article makes an impressive case that nested set operations (NS) facilitate probability computations by helping make clear the relevant natural frequency partitions; however, NS can also contribute to common errors. That NS constitute a *general* reasoning process is supported by their role in deductive, modal, causal, and other reasoning. That NS are solely a *rule-based* process is problematic.

Barbey & Sloman (B&S) observe that nested set operations (NS) do not eliminate error, and point to the potential facilitating effect of, for example, question form. Still, greater caution is in order:

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

The commonest mistake in the medical probability case – answering with the "hit rate" – can utilize NS (have disease/have disease and test positive/take test); it simply does not recognize which nesting is critical (positive and have disease/test positive, including false positives). So, what predictions concerning accuracy follow simply from people's use of NS?

Looking at the bright side, this situation underlines the generality of NS, in that they can support not only good reasoning, but also the bad and the ugly. Moreover, NS need not be restricted to rule-based reasoning: they can support either complex rule-following processes (as with more difficult syllogisms), or associative processes (e.g., making category connections merely by keeping track of similarity-based groupings). Further, NS can support reasoning at different levels of determinateness: that of simple subset relations, where these, in turn, represent elementary quantitative relations of greater, less, and equal; that of finer, but still coarse, comparisons of size, as with Euler circles drawn more-or-less "to scale"; that of determinate set size, with cardinality represented by numerals, grids, and so on. The medical probability calculation requires the third level; others need only the first or second. In sum, within any domain to which they may apply, NS constitute a highly versatile instrument of thought. In the remaining space I indicate some significant domains in which the role of NS has (apparently) been under-explored.

Deductive reasoning. NS are central to many elementary syllogistic inferences, and this area has been intensively studied (Evans et al. l993). But at least one psychologically and logically important item has not been sufficiently examined. Traditional categorical syllogistic can be modified to produce a system just as powerful as first-order logic (Sommers 1982). Among other things, one may add the Leibnizean principle, "If all As are Bs, then everything that is related in manner R to an A is related in manner R to a B." This allows one to formulate many simple relational inferences not expressible in older syllogistic. For example, "Since all natural disasters are acts of god, all cases of damage caused by natural disasters are cases of damage caused by acts of god." ("The Leibniz Clause"; check your insurance policy.) Combined with the provision that no damage caused by god is insured, the company infers that your earthquake damage is not covered. The gist of this powerful principle is simple and intuitive; it directly exploits elementary NS, and suggests a new angle for psychological exploration of everyday relational inference.

Modal deductive reasoning. Aristotle invented modal syllogistic in part to represent possible forms of human thought; yet it is still largely *terra incognita* for the psychology of reasoning. In this approach, people think in terms of (modally) different ways in which predicates relate to subjects: All As are B; all As are necessarily B; all As are possibly/possibly not Bs. Inferences consist of any mix of modal and non-modal premises and conclusions. One difficult issue would be how people represent the difference between modalities *de dicto* (a statement [*dictum*] is necessarily true, e.g., a law of nature, or a "definitional" truth such as, "All bachelors are unmarried"), versus *de re* (a predicate is necessarily true of some thing [*res*], as in, "Everything in the barnyard is necessarily a chicken"). NS can handle the latter by distinguishing the set of things *possessing* some property from the subset of things *necessarily possessing* that property (and from the coordinate subset of things *contingently* possessing that property). An empirical point of entry would be the question of whether people judge correctly the validity of these two modal syllogisms (the notorious "Two Barbaras"; Patterson l985):

All As are necessarily Bs	All As are Bs
All Bs are Cs (no modality specified)	All Bs are necessarily Cs
Therefore, All As are necessarily Cs	All As are necessarily Cs

One predictable source of error is confusion between *de dicto* and *de re* understandings of the modal propositions. On the former, both Barbaras are invalid; on the latter, only the second is valid. There are literally thousands of modal and mixed modal/non-modal syllogisms awaiting investigation, and the most fundamental of these will be crucial to reasoning involving the modally distinct ways in which sets may be nested.

Causal reasoning. If one thinks of the "causal scenario" of an event E as representing E and its causal effects (a set of events branching into the future, say), NS can naturally capture causal chains: If event A causes B, B causes C, and so forth, the causal scenario of C is nested within that of B, and B's within A's; therefore C's scenario is nested within A's, and causality is transitive. If *prevents* is understood as *causes the non-occurrence of*, prevention simply becomes another link in a causal chain/nesting. NS can also support thought involving "enablers," where these are construed as background or normal conditions in the presence of which something causes an event, by nesting the set of situations in which the causal sort of event occurs within the set of enabling-condition situations. (This is admittedly a hotly contested issue.) Even causal thought via simulation of concrete events will implicitly involve NS – provided, as is often the case, the simulation is construed as representing types of causal events (e.g., physical or psychological forces acting in some manner to produce a given sort of effect).

NS are also involved in simple deontic reasoning, classificatory schemata and inferences of all sorts, essentialist (e.g., biological) schemata in particular, and much more. B&S admirably marshal evidence for the important role of NS in probabilistic reasoning, and their claim of generality for NS is highly suggestive with regard to avenues of future research.

Converging evidence supports fuzzy-trace theory's nested sets hypothesis, but not the frequency hypothesis

DOI: 10.1017/S0140525X07001872

Valerie F. Reyna[a] and Britain Mills[b]

[a]*Departments of Human Development, Psychology, and Cognitive Science, Cornell University, Ithaca, NY 14853;* [b]*Department of Human Development, Cornell University, Ithaca, NY 14853.*

vr53@cornell.edu **bam64@cornell.edu**

http://www.human.cornell.edu/che/HD/reyna/index.cfm

Abstract: Evidence favors the nested sets hypothesis, introduced by fuzzy-trace theory (FTT) in the 1990s to explain "class-inclusion" effects and extended to many tasks, including conjunction fallacy, syllogistic reasoning, and base-rate effects (e.g., Brainerd & Reyna 1990; Reyna 1991; 2004; Reyna & Adam 2003; Reyna & Brainerd 1995). Crucial differences in mechanisms distinguish the FTT and Barbey & Sloman (B&S) accounts, but both contrast with frequency predictions (see Reyna & Brainerd, in press).

Although the evidence adduced by Barbey & Sloman (B&S) clearly supports fuzzy-trace theory (FTT), there are key differences in the mechanisms used to explain base-rate effects. That is, FTT's prediction that "Base-rate neglect is reduced when problems are presented in a format that affords accurate representation in terms of nested sets of individuals" (target article, Abstract) has been confirmed repeatedly. B&S's account "attributes base-rate neglect to associative judgment strategies" (target article, Abstract), but it is questionable whether associative processes are implicated by the data, whereas distinct ideas omitted from their account have been supported through rigorous testing. Therefore, in the remainder, we briefly summarize those differences in mechanisms used to explain the nested sets prediction, note that each assumption in FTT has been separately

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

tested empirically (unlike many competing claims which are philosophical), and spell out some important next steps in theory testing as well as further implications for rationality.

"Diagnosing whether a patient has a disease, predicting whether a defendant is guilty of a crime, and other everyday as well as life-changing decisions" and "people's tendency to neglect base-rates" (target article, sect. 1, para. 1) have been explicitly investigated in the context of FTT (e.g., Reyna & Adam 2003; Reyna & Farley 2006; Reyna et al. 2002; Reyna & Lloyd 2006; Reyna et al. 2001; Reyna et al. 2006). Four factors are essential to our explanation: (1) Confusion about which classes are referred to (present for all ratio concepts, including probability), defaulting to the focal classes in numerators and consequently neglect of denominators; (2) the allure of a salient gist or meaning relation that applies to the focal class and that seems to answer the posed question, but does not; (3) developmental and individual differences in the ability to inhibit this salient gist; and (4) the presence or absence of cues to retrieve known reasoning principles, such as the cardinality principle that subsets cannot be more numerous (or probable) than the sets that include them. Empirical evidence supports each of these theoretical factors and, together, they predict the nested sets findings discussed by B&S.

For example, using these assumptions, we, too, have shown that "Facilitation in Bayesian inference . . . can be attributed to the facilitory effect of prompting use of the sample of category instances presented in the problem to evaluate the two terms of the Bayesian ratio" (target article, sect. 2.3, para. 6), that this effect extends to improving single-event probability judgments, and that the form of the question matters (e.g., Brainerd & Reyna 1990; 1995). Like Girotto and Gonzalez (2001), we also showed that defective partitioning decreased the facilitative effect of making nested sets transparent (Lloyd & Reyna 2001). In addition, Girotto and Gonzalez (in press) provide developmental evidence that echoes earlier developmental findings by Brainerd and Reyna (e.g., 1995), and Sloman et al. (2003), Yamagishi (2003), and Bauer and Johnson-Laird (1993) provide evidence of facilitation using diagrammatic representations that also echoes our earlier findings (e.g., Brainerd & Reyna 1990; Lloyd & Reyna 2001; Reyna et al. 2001), all of which support FTT's nested sets hypothesis.

These findings have been summarized by Reyna and Brainerd (1995) as follows:

> Class-inclusion reasoning, probability judgment, risk assessment, and many other tasks, such as conditional probability, conjunction fallacy, and various deductive reasoning tasks, are subject to what has been called inclusion illusions (Reyna 1991; Reyna & Brainerd 1993; in press). Inclusion illusions occur because part-whole relationships are difficult to process, for children and for adults. (p. 34)
>
> According to Reyna and Brainerd (1993), neglect of base rates is a special case of denominator neglect: Because of the complexity of processing nested classes, subjects focus on numerators (e.g., joint probabilities or relative frequencies of targets, depending on the task). (p. 35)
>
> Processing can be simplified, however, by providing a notational system in which elements of parts and of wholes are distinctly represented. For example, Venn diagrams, used in syllogistic reasoning, represent subsets and more inclusive sets using a system of overlapping circles. Superordinate-set tags can be used to similar effect in class inclusion. (p. 33)

As these examples illustrate, B&S's evidence favoring "nested sets and dual processes," except for that presented in section 2.8, is a replication or extension of earlier findings derived from FTT. Therefore, not surprisingly, their conclusions about the nested sets hypothesis are virtually identical to earlier statements. However, the mechanisms are not identical and the characterization of System 1 (the intuitive system) as "primitive" (sect. 1.2.5, para. 2) is the opposite of what is claimed in FTT (and our claim has received repeated empirical support). We offer a dual-processes account that encompasses the findings presented here, but also has been used to predict additional, counterintuitive findings, some shown in Table 1.

In contrast to ecological and other perspectives on rationality, FTT places unique emphasis on developmental data in informing judgments of rationality. Evidence points to simplified, intuitive, gist-based thinking as a key feature of advanced reasoning that develops with age and expertise, but also accounts for increases in gist-based biases. According to FTT, these cognitive biases are irrational. However, FTT's process model distinguishes types of errors that vary in severity – that is, in degrees of rationality. For example, errors caused by susceptibility to a compelling (but incorrect) gist or by retrieving the wrong reasoning principle are more severe than those caused solely by processing interference from overlapping classes. This theoretical taxonomy maps onto observed trends in development and in the acquisition of expertise (Reyna et al. 2003).

In summary, although the target article's account of base-rate effects is consistent with FTT's (supporting the nested sets hypothesis and refuting an adaptive preference for frequency formats), the proposed dual-processes model differs in important respects. FTT stands in sharp contrast in its characterization of intuition (System 1) regarding rationality. Standard dual-processes models, such as the one advocated by B&S, attribute successful performance on base rate and other inclusion tasks to the operation of an advanced (logical and quantitative) System 2 and the inhibition of a primitive System 1. In contrast, in FTT, intuition is advanced, and focusing

Table 1 (Reyna & Mills). *Examples of counterintuitive findings that differentiate FTT from Barbey & Sloman's dual-processes approach*

1. Intuitive reasoning *increases* with development and with greater expertise (e.g., Reyna & Brainerd 1994; Reyna & Ellis 1994; Reyna & Farley 2006; Reyna & Lloyd 2006).
2. Reducing memory for verbatim problem information in class-inclusion tasks *improves* reasoning performance because it forces greater reliance on memory for gist (e.g., Brainerd & Reyna 1990; Reyna 1991).
3. Increasing working memory load by adding more information about more classes can *improve* class-inclusion reasoning by making class-inclusion relations clearer (e.g., Brainerd & Reyna 1995).
4. Base-rate neglect, disjunction fallacies and similar class-inclusion biases reflect "garden path" effects of usually adaptive reliance on *high-level* semantic gist rather than low-level mindless associations (see Reyna et al. 2003, for examples of cardiologists' judgments of heart attack risk).
5. Independent verbatim and gist memories for frequencies can lead to *contradictory* judgments from the same individuals (e.g., Reyna 1992).
6. Class-inclusion illusions, as exemplified by base-rate neglect, are predicted to persist late in development, and to be *independent of domain knowledge* and expertise (e.g., Adam & Reyna 2005; Lloyd et al. 2001; Reyna 2004; Reyna & Adam 2003; Reyna et al. 2001). Therefore, physicians exhibit base-rate neglect at the same rate as high school students in judging the post-test probability of disease.
7. FTT makes specific predictions about *task variability* (displaying different levels of reasoning in tasks that tap the same underlying competence), accounting for both early precocity and late-persisting illusions in judgments of probability (e.g., Reyna & Brainerd 1994).

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

on quantitative minutia is primitive (lowering class-inclusion performance). This surprising prediction, among others, is supported by converging evidence from multiple tasks.

ACKNOWLEDGMENTS

Preparation of this commentary was supported in part by grants from the National Institutes of Health (MH-061211) and the National Science Foundation (BCS 0553225), and by the National Cancer Institute (Basic and Biobehavioral Research Branch, Division of Cancer Control and Population Sciences).

Varieties of dual-process theory for probabilistic reasoning

DOI: 10.1017/S0140525X07001884

Richard Samuels

Philosophy Department, King's College, Strand, London, WC2R 2LS, United Kingdom.

richard.Samuels@kcl.ac.uk

http://www.kcl.ac.uk/kis/schools/hums/philosophy/staff/r_samuels.html

Abstract: Though Barbey & Sloman (B&S) distinguish various frequentist hypotheses, they opt rapidly for one specific dual-process model of base-rate facilitation. In this commentary, I maintain that there are many distinct but related versions of the dual-process theory, and suggest that there is currently little reason to favor B&S's formulation over the alternatives.

I am in broad agreement with the general position defended in Barbey & Sloman's (B&S's) excellent article. In particular, it is plausible that the data on tasks involving base-rate information are best explained on the following general assumptions:

Dual-process Thesis: Many kinds of reasoning, including those involving base-rate information, depend on the existence and interaction of two (sorts of) cognitive systems: call them System 1 and System 2.

Nested Set Thesis: Set inclusion operations play a central role in base-rate facilitation.

But although B&S painstakingly distinguish various frequentist hypotheses, they opt rapidly for one specific construal of the above pair of commitments. Specifically, they adopt a *rule-utilization hypothesis* in which base-rate facilitation depends on the use by System 2 of set theoretic rules. In what follows, I first highlight that there are many alternative ways of combining the dual-process and nested set theses. I then suggest that there is currently little reason to favor the rule-utilization hypothesis over some of the alternatives.

Varieties of nested set and dual process hypotheses. We can distinguish different versions of the nested set and dual-process theses in terms of how they address the following issues:

Issue A: What kinds of cognitive structure are specialized for the execution of set theoretic operations?'

Issue B: What specific role(s) does System 2 play in base-rate facilitation?

Regarding Issue A: In a manner that mirrors B&S's own discussion of frequentist hypotheses, we can distinguish (at least) three variants of the nested set thesis:

1a. Set Inclusion Mechanism: Base-rate facilitation depends in part on the activity of a specialized mechanism or "module" for elementary set theoretic operations.

2a. Set Inclusion Algorithm: Base-rate facilitation depends in part on the activity of an algorithm for elementary set theoretic operations.

3a. Set Inclusion Rules: Base-rate facilitation depends in part on the deployment of rules for elementary set theoretic operations.

These options are not exhaustive; and nor are they mutually exclusive. But B&S appear to endorse only 3a.

Regarding Issue B: Dual-process theories are ubiquitous in cognitive science, and at least the following functions have been assigned to mechanisms responsible for the sorts of controlled, effortful, and relatively slow processing associated with System 2:

1b. Censorship: System 2 censors the outputs from System 1 processes thereby preventing them from becoming overt responses.

2b. Selection: System 2 selects between the outputs of different System 1 processes.

3b. Inhibition: System 2 inhibits the activity of System 1 processes.

4b. Allocation: System 2 allocates cognitive resources, such as attention and information to System 1 processes.

5b. Rule utilization: System 2 computes solutions to judgmental problems by executing rules of inference.

Notice that of these options only the last requires that System 2 compute solutions to judgmental tasks. In contrast, the other four are broadly executive functions, in the sense that they involve the regulation of cognitive resources, information flow and the flow of control. Again, these options are neither jointly exhaustive nor mutually exclusive. But while B&S endorse 5b – viewing System 2 as a consumer of set theoretic rules – they remain largely neutral on which, if any, of the other functions System 2 might perform.

Is there any reason to endorse the rule-utilization hypothesis? So, there are different versions of both the nested set and dual-process theses; and they can be combined in different ways to produce distinct but related hypotheses about the cognitive systems underlying base-rate facilitation. Perhaps all B&S really want to claim is that some such account is plausibly true. This would already be a substantial and contentious hypothesis. But, as already indicated, another more specific proposal is suggested by much of what they say: a rule-utilization hypothesis that combines claims 3a and 5b. If this is the hypothesis B&S seek to defend, however, then it's far from clear that it is preferable to other ways of combining the dual-process and nested set theses.

First, the data cited in the target article do not settle the matter. To explain these data within a dual-process framework requires that: (1) Set theoretic operations are performed in cases of base-rate facilitation. (This explains the pattern of facilitation.) (2) System 2 is involved when facilitation occurs. (This explains, for example, the effects of incentives, and correlation between intelligence and performance.) But it does not follow that System 2 must itself perform these set theoretic operations. For all the data show, it may instead be that System 2 only plays an executive role while some System 1 mechanism is responsible for performing set theoretic operations.

Suppose, for illustrative purposes, that there exists a mechanism dedicated to set theoretic operations – a "set theory module," if you like (option 1a). Moreover, assume that System 2 performs one or more executive function, such as allocating resources to the set theory mechanism or inhibiting the operation of other mechanisms (options 1b–4b). Such a proposal could accept B&S's contentions that: (1) an associative System 1 process is responsible for base-rate errors, (2) facilitation involves System 2, and (3) it occurs when inputs make set theoretic relations transparent. Thus, System 2 activity could still be invoked to explain the influence of incentives and intelligence; and facilitation could still be explained by reference to the performance of set theoretic operations. But in contrast to the rule-utilization hypothesis, System 2 would play some kind of executive function as opposed to actually computing solutions to judgmental tasks. An account along these lines might offer a possible explanation of the data, while making claims that differ in important respects from those enshrined in the rule-utilization hypothesis.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

But perhaps the rule-utilization hypothesis should be preferred on general theoretical grounds? In particular, one might think that the alternatives are in tension with general assumptions of dual-process theory. Most obviously, one might think the following:

Dual-process theories posit only two reasoning systems. In which case, since System 1 is responsible for errors, System 2 is presumably responsible for successful responses.

System 1 processes are associative, whereas System 2 processes are rule-based. In which case, set-theoretic operations cannot be subserved by System 1 since they are not associative. Therefore, System 2 must be responsible for the execution of set-theoretic operations.

But it would be a mistake to adopt the rule-utilization hypothesis on such grounds. Though some versions of the dual-process theory incorporate these assumptions, they are at best highly contentious and indeed have been the subject of much recent debate. First, there is considerable debate among dual-process theorists over whether "System 1" and "System 2" label individual systems or *kinds* of systems. Indeed, there is a growing consensus amongst researchers that there are *many* System 1 mechanisms (Evans, forthcoming; Stanovich 2004). In which case, a mechanism for set-theoretic operations is wholly consistent with the claims dual-process theorists make about the plurality of reasoning systems.

Similarly, there is no reason to assume at this time that all System 1 processes must be associative and System 2 rule-based. Admittedly, many dual-process theorists, B&S included, appear to make this assumption. (At any rate, B&S adopt the convention of labeling them as such "in an effort to use more expressive labels"; sect. 1.2. 5, para 1.) But once more, these assumptions are highly contentious; and many prominent dual-process theorists are happy to categorize non-associate mechanisms – including a hypothetical frequentist module – as components of System 1 (Evans, forthcoming; Stanovich 2004). At this time, the issue of whether System 1 processes are exclusively associative should, it seems to me, be treated as an open empirical matter.

Of course, it may just be that B&S find the rule-utilization hypothesis more attractive on grounds of parsimony since it avoids any commitment to specialized mechanisms for set-theoretic inference. If so, I sympathize. But, given the current widespread popularity of modularity hypotheses, such considerations are unlikely to bear much weight. Even if B&S are right to advocate the nested set and dual-process theses, much more work is required to adjudicate between the various versions of this general proposal.

The effect of base rate, careful analysis, and the distinction between decisions from experience and from description

DOI: 10.1017/S0140525X07001896

Amos Schurr and Ido Erev

School of Education, The Hebrew University of Jerusalem, Jerusalem, 91905, Israel; Faculty of Industrial Engineering, Technion – Israel Institute of Technology, Haifa, 32000, Israel.

samos@mscc.huji.ac.il **erev@tx.technion.ac.il**
http://ie.technion.ac.il/erev.phtml

Abstract: Barbey & Sloman (B&S) attribute base-rate neglect to associative processes (like retrieval from memory) that fail to adequately represent the set structure of the problem. This commentary notes that associative responses can also lead to base-rate overweighting. We suggest that the difference between the two patterns is related to the distinction between decisions from experience and decisions from description.

Barbey & Sloman's (B&S's) analysis of previous studies of the effect of base rate information demonstrates that in many cases the effect increases when the set structure of the problem is made more transparent. As a result, the participants can perform more complete analysis of the data. For example, the reliance on base rates is enhanced by the following manipulations: Partitioning the data into exhaustive subsets, using diagrammatic representation of all relevant sets, and formulating the question in a way that encourages participants to compute the two terms of the Bayesian ratio first, instead of direct computation of the probability.



Figure 1 (Schurr & Erev). The text used in Experiment 1 of Erev et al. (2007).

The main goal of our commentary is to highlight an interesting set of conditions that lead to the opposite pattern. Under these conditions the presentation of information concerning the set structure reduces the effect of the relevant base rates. For an example, consider the task of reading the text in Figure 1.

This task was studied in Erev et al. (2007) following Bruner and Minturn (1955; cf. Kahneman 2003). Their control condition suggests that the participants are not likely to consider the possibility the central stimulus is a number. About 90% of the participants read the central stimulus as the letter "B" (the remaining 10% read it as the number "13"). Thus, the vast majority behaved as if they overweighted the base rate – that is, the fact that a stimulus that appears between letters is more likely to be a letter than a number. Erev et al. show that manipulations that make the set structure of the problem more transparent (e.g., the presentation of the possible hypotheses) decrease such base-rate effects. In the example given here, the presentation of the possible hypotheses ("B" or "13") increases the proportion of the low base rate responses ("13") from 10% to 50%.

We believe that the difference between the present example and the situations examined by B&S reflects the difference between decisions from description, and decisions from experience (see Hertwig et al. 2004). The main tasks analyzed by B&S involve decisions (or judgment) from description. The decision makers were presented with a description of the task that includes the key factors. Hertwig et al. (2004; see also Erev et al. 2007) show that in decisions from description people deviate from optimal choice in the direction of giving equal weight to all the possibilities. That is, the low base rate categories receive "too much" attention and the objective base rate is neglected.

The example in Figure 1, in contrast, involves decisions from experience. The participants did not receive a description of the possible categories and/or their base rates. Recent research (see review in Erev & Barron 2005) shows a bias toward underweighting of rare events (low base-rate categories) in decisions from experience. People behave "as if" they forget to consider the low base-rate category. That is, in this case forgetting and similar cognitive limitations imply a very strong base-rate effect.

In summary, we propose that it is constructive to distinguish between two ways in which base rates affect human behavior. The first effect is likely to emerge in decisions from description as a product of careful analysis. B&S focus on this effect and note that it can be described as an outcome of the rule-based reasoning. The second effect is likely to emerge in decisions from experience as a product of forgetting and/or neglect of the low base-rate categories. We assert that this effect is rather common, and is likely to be decreased by careful analysis.

ACKNOWLEDGMENT

We would like to thank Uri Leron for useful comments.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Base-rate neglect and coarse probability representation

DOI: 10.1017/S0140525X07001902

Yanlong Sun and Hongbin Wang

School of Health Information Sciences, The University of Texas Health Science Center at Houston, Houston, TX 77030.

Yanlong.Sun@uth.tmc.edu **Hongbin.Wang@uth.tmc.edu**

Abstract: We believe that when assessing the likelihood of uncertain events, statistically unsophisticated people utilize a coarse internal scale that only has a limited number of categories. The success of the nested sets hypothesis may lie in its ability to provide an appropriate set structure of the problem by reducing the computational demands.

The target article by Barbey & Sloman (B&S) challenges the natural-frequency based theories of human judgment and decision making and, instead, supports a nested sets hypothesis. The authors conclude that judgmental errors and biases in Bayesian calculations are attenuated when problems are represented properly through nested sets of individuals. Although in general we agree with the authors' claims, in this commentary we would like to further examine, from a different perspective, why the nested sets hypothesis might provide a more adequate representational account than the natural frequency representations.

We have recently hypothesized that when assessing the likelihood of uncertain events, statistically unsophisticated people utilize a coarse scale that only has a limited number of categories (Sun et al., in press). The essence of the hypothesis is that, without adequate anchors, people's internal representations are coarse and limited, and therefore do not map one-on-one onto the continuously distributed external values, which can be either probabilities (or percentages) or frequencies.

The coarseness hypothesis is based on the large body of behavioral and neuroimaging research on mental presentations of quantity and numbers. Miller (1956) suggested that the number of levels of any variable that can be internalized is not only finite, but also small. In psychometric research, many researchers believe that the number of response alternatives on a scale is quite limited (for a review, see Cox 1980). Recent studies using brain-imaging techniques have provided neurological evidence indicating the existence of a coarse scale for the internal representation of numerical values. Dehaene and colleagues (e.g., Dehaene et al. 1999) suggest that there is a coarse and analog mental number line, which is the foundation of a "number sense" and shared by humans and animals. Particularly, Dehaene et al. show that exact calculations involve linguistic representations of numbers and are controlled by the speech-related areas of the left frontal lobe in the brain. In contrast, approximate calculations are language-independent and rely on visuo-spatial representations of numbers controlled by the left and right parietal lobes. Therefore, it is possible that there are two different calculation processes involved in the Bayesian inference. It might be too early to link this distinction to the dual-process model ("associative" and "rule-based") suggested in the target article, but the theoretical relevance appears to be evident. Particularly for lay persons, who do not have the ability or enough information to carry out exact calculations, it is likely that their intuitive assessment of event likelihoods is a "sense of approximation" based on a coarse internal representation.

We conducted two experiments to test the coarseness hypothesis (Sun et al., in press). In Experiment 1, participants estimated event probabilities in a free format. The experiment task was to estimate the winning probabilities of poker hands in a one-deck-and-two-player "draw poker" game. Despite the fact that the target probability has an even and nearly continuous distribution of probabilities ranging from zero to close to 100% with relatively small increments, we found that subject responses were highly clustered with approximately 5 clusters each. In Experiment 2, participants made forced comparisons using two different external response scales (a 3-level coarse scale versus a 10-level fine scale). We found that their performance did not measure up to the requirement of the finer scale. These findings indicate that besides the systematic biases, a certain portion of human errors in probabilistic judgment may be due to the low resolution of the internal representations. Further analyses of the experimental data and computer simulations implied that the number of internal categories is about 5.

This coarseness model can be used to account for physicians' overestimation of the probability of cancer, given a positive test result. Specifically, probably to a physician, what matters most is a dichotomy between a positive and a negative test result. No matter how small the probability of a breast cancer can be (due to the low base rate), a positive test result has to be taken seriously. Thus, the test result would naturally serve as an external cue to anchor two distinctive mental states. If the number of options after a test is limited (e.g., two options with one threshold), a large amount of discrepancy manifested as the overestimation of cancer probability could be accounted for by the mismatch between two different scales: one is an approximate evaluation of the seriousness of a situation (e.g., classifying it as "dangerous"), and the other is an exact numerical value in a continuous distribution. Such discrepancy may not be easily made to disappear by simply using a frequency representation, since frequencies, normalized or not, are still continuously distributed.

B&S have demonstrated that diagrammatic representations such as Euler circles, when employed to construct a nested set structure, would facilitate Bayesian inference. This is consistent with our hypothesis for a coarse scale of internal representations. Figure 1 in the target article shows a probability space nested in three levels: all possibilities, number of individuals (or chances of) testing positive, and, finally, the number of individuals (or chances of having) the disease. In this setting, computations are facilitated by external cues. Most important, these cues are represented hierarchically, so that at any moment, human subjects would only need to divide the probability space into a limited number of categories. In effect, when significant facilitation is found using natural frequency representations, it is often the case that hypotheses are preorganized to facilitate a limited number of set operations (such as the binary hypotheses used in Gigerenzer & Hoffrage 1995). Therefore, it seems that the key factor underlying facilitation is a structure with a limited number of sets at each computation phase. The success of the nested sets model may well lie in the "chunking" mechanism (Miller 1956) which reduces the computational demands for human subjects. Other distinctions, such as whether natural frequency is normalized to an arbitrary reference class, do not address this issue. And it would be no surprise that little or less facilitation is found when natural frequencies are used but without transparent nested set structure.

Implications of real-world distributions and the conversation game for studies of human probability judgments

DOI: 10.1017/S0140525X07001914

John C. Thomas

T. J. Watson Research Center, IBM, Yorktown Heights, NY 10598.

jcthomas@us.ibm.com **http://www.truthtable.com**

Abstract: Subjects in experiments use real-life strategies that differ significantly from those assumed by experimenters. First, true

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

randomness is rare in both natural and constructed environments. Second, communication follows conventions which depend on the game-theoretic aspects of situations. Third, in the common rhetorical stance of storytelling, people do *not* tell about the representative but about unusual, exceptional, and rare cases.

In this commentary, I do not directly argue against the findings of Barbey & Sloman (B&S). However, I note that the range of base rate discounting effects under nearly (apparently) identical circumstances is roughly as large as the effect itself. This strongly suggests that important factors are not being addressed. I provide three suggestions for what some of these factors might be.

First, in the real world, people do not typically experience true randomness. In the natural as well as the artificial world, things are typically arranged in a "clumped" fashion. Given that the natural world in which we have evolved as well as the world that we have constructed are both non-random, it would be a priori rather amazing if we humans somehow developed a natural penchant for dealing with frequencies of randomly occurring events. The conservatism shown in probability estimations is consistent with a bias toward believing in the "clumpiness" of distributions.

Not only do artifacts drawn from the natural and artificial worlds tend to "start out" in non-random clumps, but many mechanical and social processes are such that even if an explicit attempt is made to produce randomness, any interruption or incompleteness in that process is likely to result in something that is less than truly random. To give a simple example, if one takes a small canister of black balls, pours them into a larger bin, adds a small canister of white balls on top, and then begins shaking them together, at every point until randomness is achieved, there are likely to be a disproportionate number of white balls on top. Note that this process is very asymmetrical. Therefore, if an experimenter who purports to present a subject with a "random mixture" makes any reasonable kind of error (does not shake long enough, shakes in such a way that layers are not intermixed), the result is that some degree of "clumpiness" will persist.

Second, in the social world, communication is not a mere encoding of what is. More often, there is communicative purpose to communications. Better than an "encoding-decoding" model is a "design-interpretation" model of human-to-human communication (Thomas 1978). How people relate to propositions presented to them is complexly influenced by the inferred motives of those who present the information. It would be astounding if every subject in a psychological experiment simply and naively believed everything an experimenter presented about the purpose and context of the experiment. Even if a subject presumes that they are engaged in a "purely cooperative" effort with the experimenter, conversational postulates will still hold (Grice 1978). These imply that the experimenter only presents data that are necessary and sufficient for the task. Further, what a particular subject views as "real news" depends on what they already know (Clark & Brennan 1991). If an experimenter says that "the earth has one natural satellite," because subjects already know this, they will tend to assume that this is a set-up for a conversation about artificial satellites. When statements are presented about cancer rates, tests, and so on, subjects may evaluate these statements in light of what they already know about these topics. In addition, subjects make some sort of assessment of *why* they are being told. The "real" motive, from the perspective of the experimenter – namely, to determine general characteristics of human cognition – may be a common motive among the experimenter's peers, but it is *not* a motive widely shared in the larger society. Indeed, to many subjects, this may seem to be a cover story for an assessment of their personal capabilities. The assumptions of rhetorical purpose may well interact with the obviousness of the representation. Mood and personality will tend to play a bigger part in the interpretation of an inkblot than in the interpretation of a relatively clear and unambiguous stimulus.

Many of the specific findings reviewed in the target article are understandable from this perspective. For instance, more concrete and specific statements are more believable and more likely to be taken at face value because they are more subject to verification or disproval. University prestige may well make a difference in terms of source credibility, rather than the general intelligence of the subjects. If subjects are paid, there is more chance in our society of legal repercussions for lying or deception, and awareness of this factor also increases source credibility.

Third, a particularly common rhetorical context is storytelling. People deal with stories both in personal interaction and in explicit entertainment contexts such as television, movies, and novels. Stories in these latter contexts are not typically told to communicate about representative situations, but rather, concern the "edges" of human experience, the exceptions, the rare and unusual (McKee 1997). Therefore, it is natural that when one is told a story, one tends to assume purposeful dramatic action. In a movie, if someone goes to the doctor for a diagnostic test for some rare disease, the person is much more likely to have that disease than would be predicted by statistics. Furthermore, writers choose details for rhetorical purpose. For instance, if someone in a soap opera walks into a room and they *look* like a professional football player, the chances are actually high that they *are* a professional football player and not an accountant or salesperson.

These three potentially confounding influences do not mean that better predictability is impossible in this paradigm. There may be an analogy with measuring sensory thresholds. Trying to measure absolute thresholds by asking people whether or not they hear a sound can be very sensitive to expectations, set, and motivation. Asking subjects to specify in which of two intervals a sound appeared is much less sensitive to these social variables. A forced choice paradigm should also work in this context to minimize the potentially confounding effects recounted above.

Why the empirical literature fails to support or disconfirm modular or dual-process models

DOI: 10.1017/S0140525X07001926

David Trafimow

Department of Psychology, MSC 3452, New Mexico State University, Las Cruces, NM 88003-8001.

dtrafimo@nmsu.edu

http://www.psych.nmsu.edu/faculty/trafimow.html

Abstract: Barbey & Sloman (B&S) present five models that account for performance in Bayesian inference tasks, and argue that the data disconfirm four of them but support one model. Contrary to B&S, I argue that the cited data fail to provide strong confirmation or disconfirmation for any of the models.

There is insufficient space here to comment on all of the models for explaining performance on Baysian inference tasks that Barbey & Sloman (B&S) ostensibly disconfirm, so I will focus on what they consider to be the model that makes the strongest claims – the idea that the mind comprises specialized modules (see Barrett & Kurzban [2006] for a recent review). B&S's strategy is to list the prerequisites of the modular model and cite data that contradict them. The listed prerequisites are cognitive impenetrability, informational encapsulation, unique sensitivity to natural frequency formats, transparency of nested set relations, and appeal to evolution, the first three of which are

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

contradicted by some of the cited findings. However, these findings are not a problem for the modular model because researchers who espouse the modular view have long since moved away from these three prerequisites and instead focus on *"how modules process information"* (Barrett & Kurzban 2006, p. 630). Although modules are considered to be domain specific, a domain is not defined as a content domain, but rather, as any way of individuating inputs. It is entirely possible that a module will process information for which it was not originally designed as a by-product if this other information conforms to the properties that determine which inputs are processed.

Let us now consider transparency of nested set relations and appeal to evolution. The former is featured by the model favored by B&S, and so they clearly cannot mean to argue against it, which leaves only the latter as a possible basis for disconfirmation. But few in the scientific world would argue that evolution did not happen and so this is unlikely to be disconfirmed; certainly B&S have not presented any evidence to disconfirm evolution. Consequently, the modular model is not forced to make or not make any of the predictions listed in Table 2 of the target article, and I am compelled to conclude that B&S have failed to disconfirm the modular model (or any of the weaker ones).

The foregoing comments should not be taken as arguments in favor of mental modules. For one thing, the watering down of the concept of modules, which renders it less susceptible to disconfirmation, may have caused the informational content and general utility of the model to also be watered down. In addition, the auxiliary assumptions necessary to make the modular model useful are extremely complicated and these complications may be under-appreciated. As an example, consider an arm as a module. Arms increase the ability to use tools, crawl, fight, balance, climb, and many other abilities. In addition, the arm might be said to comprise features (fingers, elbows, etc.) How would one tease apart the functions for which arms evolved versus those that are mere by-products, especially after taking into account that the features may or may not have evolved for very different reasons? Surely a mind is much more complicated than an arm, and so the potential complications are much more extensive. Perhaps these issues will be solved eventually but my bet is that it will not happen soon. Until this time, the modular model seems unlikely to provide a sound basis for Bayesian theorizing or theorizing in any other area of psychology.

The data cited by B&S also fail to provide much support for the dual-process model they maintain. It is doubtless true that presenting Bayesian problems such that the set structure is more transparent increases performance. But it is not clear why this necessitates a distinction between associative and rule-based processes, a distinction that has not been strongly supported in the literature. In fact, Kruglanski and Dechesne (2006) have provided a compelling argument that these two types of processes are not qualitatively distinguishable from each other; both processes can involve attached truth values, pattern activation, and conditioning. Worse yet, even if the distinction were valid in some cases (and I don't think it is), there is very little evidence that it is valid in the case at hand. B&S seem to argue that when the set structure is not transparent, then people use associative processing; whereas they use a rule when the set structure is more transparent. It could be, however, that when the set structure is not transparent, people use rules but not the best ones. Or, when the set structure is transparent, this transparency may prime more appropriate associations. These alternative possibilities weaken the evidentiary support for the distinction.

B&S provide a section titled, "Empirical summary and conclusions" (sect. 2.10) that illustrates what I consider to be the larger problem with the whole area. Consider the empirical conclusions. First, the helpfulness of frequencies varies across experiments and is correlated with intelligence and motivation. Who would predict that there will be no variance and that intelligence and motivation will be irrelevant to problem solving? Second, partitioning the data so as to make it more apparent what to do facilitates problem solving – another obvious conclusion. Third, frequency judgments are guided by inferential strategies. Again, who would predict that people's memories of large numbers of events will be so perfect as to render inferential processes unnecessary? (To anticipate the authors' Response, modular theorists cannot be forced to predict this.) Fourth, people do not optimally weight and combine event frequencies and use information that they should ignore. Given the trend in both social and cognitive psychological research for the last quarter century or more, documenting the many ways people mess up, this is hardly surprising. Finally, nested set representations are helpful, which is not surprising because they make the nature of the problem more transparent. Trafimow (2003) provided a Bayesian demonstration of the scientific importance of making predictions that are not obvious. Hopefully, future researchers in the area will take this demonstration seriously.

ACKNOWLEDGMENT

I thank Tim Ketelaar for his helpful comments on a previous version of this commentary.

The motivated use and neglect of base rates

DOI: 10.1017/S0140525X07001938

Eric Luis Uhlmann,[a] Victoria L. Brescoll,[a] and David Pizarro[b]

[a]*Department of Psychology, Yale University, New Haven, CT 06520;*
[b]*Department of Psychology, Cornell University, Ithaca, NY 14853.*

eric.uhlmann@yale.edu **victoria.brescoll@yale.edu**
dap54@cornell.edu **http://www.peezer.net/Home.html**

Abstract: Ego-justifying, group-justifying, and system-justifying motivations contribute to base-rate respect. People tend to neglect (and use) base rates when doing so allows them to draw desired conclusions about matters such as their health, the traits of their in-groups, and the fairness of the social system. Such motivations can moderate whether people rely on the rule-based versus associative strategies identified by Barbey & Sloman (B&S).

Barbey & Sloman (B&S) provide a convincing account of the contributions of associative and rule-based cognitive processes to base-rate respect. Absent from their model, however, is a consideration of the effects of psychological motivations on the use of statistical rules. The sorts of motivations known to influence the use of statistical rules fall into three general categories: ego-justifying, group-justifying, and system-justifying (Jost & Banaji 1994).

Ego-justifying neglect of base rates occurs in evaluations of medical diagnoses. For example, Ditto et al. (1998) told participants that they had tested positive for an enzyme (TAA) whose presence was predictive of immunity *or* vulnerability to pancreatic disease. Individuals in the "healthy consequences" condition were told that TAA made it less likely they would get pancreatic disease, whereas individuals in the "unhealthy consequences" condition were informed that TAA increased their chance of getting pancreatic disease. Participants were also told either that the test was highly accurate (1 in 200 failure rate), or relatively inaccurate (1 in 10 failure rate). Participants who were told that their TAA levels put them at risk for pancreatic disease and that the test was relatively inaccurate, perceived the diagnosis as less accurate than participants in the high accuracy condition – a normatively defensible application of the base rate. But participants who were told that their TAA levels reduced the risk of pancreatic disease, and were further informed that the test was inaccurate, were just as likely as participants in the high accuracy conditions to perceive the diagnosis as accurate.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Base-rate neglect can also be driven by bias in favor of one's social group (Ginossar & Trope 1987; Schaller 1992). In one study, the male employees of an organization were described as stronger leaders than the female employees (Schaller 1992). However, there was an additional, much more predictive base rate at work: participants were also told that male employees were dramatically more likely than female employees to be assigned to serve in an executive role. In other words, the males were stronger leaders because more of them were assigned to serve in leadership roles. Making the normatively rational judgment, female participants took into account the base rate of males and females in executive roles and concluded that male and female employees were equally talented leaders. But male participants neglected the assignment of male and female employees into different organizational roles, concluding that male employees were superior leaders. A separate experiment revealed that female participants were likewise biased in favor of their own group. These female participants ignored the base rate of male and female executives when it led them to the (incorrect) conclusion that female employees were superior leaders. In sum, participants neglected a base rate when it allowed them to draw a conclusion favorable to their own gender.

The motivation to uphold the social hierarchy (i.e., system justification) also plays a role in the application of base rates about racial groups to individual group members (McDell et al. 2006; Tetlock et al. 2000). Individuals who are non-prejudiced toward Black Americans make similar estimates of group crime rates among White and Black Americans as prejudiced individuals do. However, only the prejudiced individuals (i.e., those who have a motivation to uphold the social hierarchy) endorse the *use* of base rates to discriminate against an *individual* Black person. Individuals who endorse social hierarchies based on groups competing for power (a so-called "social dominance orientation"; Sidanius & Pratto 1999), are also more likely to endorse the application of base rates to individuals.

These biasing psychological motives likely work through the recruitment of the cognitive processes described by B&S. For example, research has demonstrated that social-psychological motives moderate whether associative or rule-based cognitive processes are employed in the first place. Ditto et al. (1998) presented evidence that people expend little cognitive effort when presented with information that favors a desired conclusion – they quickly accept it with minimal deliberation. Conversely, when presented with *undesired* information (that is, information inconsistent with one of the aforementioned motives), individuals seem especially likely to recruit rule-based, deliberative thinking in an effort to discredit the undesired information.

Ego-justification, group-justification, and system-justification motives are difficult to defend as *rational* influences on the use of statistical rules in social judgments. Although a person motivated by racial prejudice may make a "correct" judgment (i.e., a close approximation to the answer Bayes' Theorem would formally provide) when assessing the probability that a member of another race is a criminal, few would argue that this is due to statistical reasoning. Here we can distinguish between the rationality of the *belief* and the rationality of the *process* that led to that belief. Because social motivations easily (and often) lead to error, they make for suspicious guides to truth. Relying on them to achieve a rational belief is like throwing darts to choose stock winners. One may pick the best stocks, but surely it was by accident. Indeed, many people would reject the influence of these biases if they were made aware of them (i.e., such motives fail the test of *subjective rationality*; Pizarro & Uhlmann 2005).

An emphasis on social-psychological motivations may lead not only to a more complete understanding of base-rate neglect, but may also enrich a variety of cognitive theoretical approaches to human judgment. The human mind may possess specific mechanisms (e.g., in-group loyalty) that were adaptive because they aided in the individual's survival in an inherently social environment. Therefore, it may be important to consider such influences when accounting for phenomena that, at first, appear to be non-social in nature. For example, basic cognitive processes such as induction from property clusters contribute to biological explanations for natural kinds (Gelman 2003, Keil 1989). Yet, recent studies demonstrate that system-justifying motives may lead people to endorse biological explanations such that explaining group differences as "natural" helps justify their continued existence (Brescoll & Uhlmann 2007). Thus, applying social-psychological motives to theories of cognitive processes may lead to a more complex, but hopefully also more accurate, portrait of human cognition.

Base-rate respect meets affect neglect

DOI: 10.1017/S0140525X0700194X

Paul Whitney, John M. Hinson, and Allison L. Matthews
Department of Psychology, Washington State University, Pullman, WA 99164-4820.
pwhitney@wsu.edu hinson@wsu.edu almatthews@wsu.edu

Abstract: While improving the theoretical account of base-rate neglect, Barbey & Sloman's (B&S's) target article suffers from *affect neglect* by failing to consider the fundamental role of emotional processes in "real world" decisions. We illustrate how affective influences are fundamental to decision making, and discuss how the dual process model can be a useful framework for understanding hot and cold cognition in reasoning.

In the target article, Barbey & Sloman (B&S) do an admirable job of demonstrating that a dual process model of judgment provides a better account of base-rate neglect than the various alternative accounts. We were struck, however, by a curious dissociation in their article that is representative of research on base-rate neglect in general. The examples provided to illustrate how research on base-rate neglect may be important to "real world" decisions typically involve intrinsically emotional contexts such as cancer diagnosis, pandemic infections, or judgments about the guilt of a defendant. Nevertheless, the target article continues the tradition of neglect of affective factors in reasoning. This neglect is odd considering the recent resurgence of interest in affect in cognitive neuroscience and the increasing evidence that both hot and cold cognition are involved in decision making (e.g., Lee 2006; Sanfey et al. 2006). In fact, one of the most important advantages of the dual-process model of reasoning may be that it provides a coherent framework for understanding sources of affective influences on reasoning. Before we turn to why the dual process model is a useful framework for understanding hot and cold cognition during reasoning, we first briefly review some of the evidence that suggests that affective influences should be integrated into research on reasoning.

As one illustration of the central role of affect in decision making and reasoning, consider the risk-as-feeling hypothesis (Loewenstein 2005; Loewenstein et al. 2001). Loewenstein and colleagues argue that, when in conflict, hot cognitive factors will supersede cold ones in decision making, and that the precedence of hot factors helps to explain some violations of normative decision making in traditional theory. For example, the *certainty effect* (e.g., Kahneman & Tversky 1979) is a commonly observed nonlinearity in the way probabilities are weighted in decision outcomes. Although the difference in a very high probability event and certainty may seem trivial from a cold cognitive perspective, real emotion may be either absent or present in these two cases. A medical patient

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

who is told that the risk of a specific form of cancer is zero, acts and feels very differently from one who is told there is a non-zero probability of the disease. Once the prospect of the disease becomes real, the associated worry and dread become vivid.

In a related area of research, Slovic and colleagues have emphasized the role of an *affect heuristic* in judgment and decision making (e.g., Finucane et al. 2000; Slovic et al. 2004). According to this view, affect is central to reasoning because the information used during decision making often carries affective tags. These affective tags can contribute to quicker and more efficient decision making, but they pose a risk of constraining and biasing decisions, as well. Slovic and colleagues have shown how the affect heuristic is consistent with a wide range of data on reasoning under uncertainty. For example, the affect heuristic provides an explanation for why some medical risks may be overestimated. For instance, the mere knowledge that a chemical could be carcinogenic will lead non-experts to overestimate cancer risks from low-level exposure (Slovic et al. 2005). Because of the emotional impact of cancer, the absence of detailed knowledge of actual risks is ignored. The emotional associations of a disease can have other, unintended consequences. A cancer-screening procedure that reduces risk may, ironically, lead to heightened perceptions of the likelihood of contracting the disease as dread and worry are made more emotionally salient.

A key point made by these and other researchers who have examined affective influences on reasoning is that affect is not a secondary consideration to take up once we have a good "cold" model of reasoning. There are two main reasons to consider affect as fundamental to the study of reasoning. First, affective influences are pervasive. Even in the case of formal logical reasoning with syllogisms, both incidental affective reactions during reasoning and manipulation of affective dimensions of the stimulus material have powerful effects (e.g., Blanchette 2006). Second, affective influences are not simply distractions that perturb the functioning of a cold rational system. Affect is part of the very nature of the reasoning process (e.g., Damasio 1994; Loewenstein 2005). Given that research on hot and cold cognition in reasoning has stressed the integration of multiple kinds of information in explaining choice behavior (e.g., Hinson et al. 2006), modular views of reasoning are theoretic non-starters. In contrast, the dual process framework favored by B&S to explain base-rate neglect can be integrated readily with the data on affective influences on reasoning.

The most obvious way to integrate affective influences into the dual process model is to assign affective influences to the primitive associative judgment system (cf. Epstein 1994). This idea embodies the classic notion that the primary influence of affect in judgment is to cloud our otherwise efficient deliberative cognitive system. However, this view, at the very least, greatly oversimplifies the interaction of hot and cold factors in reasoning. For example, one role for affective processing may be to push the reasoning system either toward or away from deliberative processing. In a positive affective context, the simpler associative processing mode of judgment is more likely to be used, whereas negative context can induce more deliberative processing (see Fiedler 2001). In addition, there is mounting evidence that there are specific affective influences within the deliberative system itself. Particularly relevant to the target article are data on why probability and frequency representation of events often lead to different decision making outcomes (Peters & Slovic 2000). For example, if people are asked to provide a recommendation about release of prisoners, they are far more likely to recommend release if told there is a 10% chance of reoffense, rather than being told than 10 out of 100 will reoffend. This difference in judgment in two identical situations results from the ease with which the frequency representation is associated with the affective consequences of actual people committing real crimes. This example is one of many that suggest that affect neglect is a suboptimal research heuristic.

Adaptive redundancy, denominator neglect, and the base-rate fallacy

DOI: 10.1017/S0140525X07001951

Christopher R. Wolfe

Western College Program, Miami University, Oxford, OH 45056.
WolfeCR@muohio.edu **http://tappan.wcp.muohio.edu/home/**

Abstract: Homo sapiens have evolved a dual-process cognitive architecture that is adaptive but prone to systematic errors. Fuzzy-trace theory predicts that nested or overlapping class-inclusion relations create processing interference, resulting in denominator neglect: behaving as if one ignores marginal denominators in a 2 × 2 table. Ignoring marginal denominators leads to fallacies in base-rate problems and conjunctive and disjunctive probability estimates.

In a pre-scientific era, the Greek philosopher Socrates "demonstrated" that all learning is remembering, by leading an illiterate slave step by step to prove the Pythagorean theorem simply by asking him questions and drawing lines in the sand with a stick (Plato 2006). Of course, such a demonstration reveals more about the mind of Socrates than that of the slave. So it is with contemporary attempts to demonstrate that Homo sapiens are "intuitive Bayesians" (e.g., Gigerenzer & Hoffrage 1995). Researchers can encourage behavior somewhat aligned with Bayes' theorem by providing participants with natural frequencies, posing questions that facilitate Bayesian computation, organizing statistical information around the reference class, or presenting diagrams that highlight the set structure of problems (see Table 2 in the target article). All of these tasks are useful in illuminating our understanding of judgment and decision-making; none of them demonstrate that people are essentially Bayesian.

At its best, evolutionary psychology provides useful constraints on theorizing and more closely aligns brain and behavioral sciences with modern evolutionary biology. At worst, however, claims about the environment of evolutionary adaptation become "just-so stories" conferring scientific legitimacy on the author's initial assumptions rather than producing falsifiable hypotheses. In the case of judgment under uncertainty, it is obvious that our ancestors did not reason with percentages. However, there is no evidence that the mind "naturally" processes frequencies. Indeed, aesthetically, it may be seem more "natural" to imagine our ancestors reasoning about "the chance that this mushroom I just picked is poisonous" rather than "what number out of 40 similar looking mushrooms picked under similar circumstances are poisonous." More to the point, there is very good evidence that at least one contemporary hunter-gatherer culture, the Pirahã people of Brazil, have no words for numbers other than "one, two, and many" and that on numerical cognition tasks, their performance with quantities greater than three is "remarkably poor" (Gordon 2004). If hunter-gatherer peoples of our own time can get by without numeric concepts, why should we assume that proto-humans in the ancestral environment developed hardwired mechanisms that "embody aspects of a calculus of probability" (Cosmides & Tooby 1996, p. 17; quoted in the target article, sect. 1.2.2, para. 3) enabling us to automatically solve story problems in a Bayesian fashion?

A more reasonable assertion is that Homo sapiens have evolved a cognitive architecture characterized by adaptive redundancy. In many areas of reasoning – problem solving, judgment, and decision making – people make use of more

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

than one kind of cognitive process operating on more than one type of cognitive representation. The particular substance of these processes and representations are developed through learning in a cultural context, although the cognitive architecture itself may be part of our biological inheritance. Dual-process models are beginning to characterize the nuts and bolts of this adaptive redundancy in human cognition. The portrait emerging from the research is of a human organism that is generally capable and adaptive (the glass is half full) but also prone to ignoring base rates and other systematic deviations from normative performance (the glass is half empty). Barbey & Sloman's (B&S's) careful review of the literature in the target article clearly suggests that dual process theories best account for the empirical evidence pertaining to base-rate neglect.

B&S highlight the similarities between several dual process theories, asserting that people reason with two systems they label associative and rule-based. They attribute judgmental errors to associative processes and more accurate performance with base rates to rule-based inferences – provided that problems are presented in formats that cue the representation of nested sets underlying Bayesian inference problems. As the authors note, this is the heart of the Tversky and Kahneman (1983) nested set hypothesis. It is here where differences among the dual process theories begin to emerge and where the specific details of Fuzzy-Trace Theory (FTT; Reyna & Brainerd 1995) shed light on intuitive probability judgments.

The dual systems of FTT operate on verbatim and gist representations. FTT asserts that vague impressions are encoded along with precise verbatim information. Individual knowledge items are represented along a continuum such that fuzzy and verbatim memory traces coexist. Gist memory traces are not derived from verbatim representations but are formed in parallel using different mechanisms. The result is the creation of multiple traces in memory. Verbatim and gist traces are functionally independent, and people generally prefer to reason with gist representations for a given task.

FTT predicts that people have difficulty with conditional and joint probabilities because it is hard to reason about nested, hierarchical relationships between items and events. Nested or overlapping class-inclusion relations create processing interference and confusion even in educated thinkers who understand probabilities (Reyna & Brainerd 1995). People prefer to reason with simplified gist representations of problems (the fuzzy-processing preference), and one specific way of simplifying predicted by FTT is denominator neglect.

Denominator neglect consists of behaving as if one is ignoring the marginal denominators in a 2×2 table. Thus, in a 2×2 table the base-rate P(B) is the marginal total of P(B and A) + P(B not A). Ignoring marginal denominators such as P(B) in estimating P(A and B) or P(A given B) can lead to logical fallacies. The FTT principle of denominator neglect allows for a priori and precise predictions about errors of conjunction and disjunction as well as base-rate neglect. We have found that ignoring marginal denominators can lead to systematic errors in problems involving base rates (Wolfe 1995) and conjunctive and disjunctive probability estimates (Wolfe & Reyna, under review).

Denominator neglect also explains conversion errors in conditional probability judgments, that is, confusing P(A given B) with P(B given A) (Wolfe 1995). When problems are presented in a format that affords an accurate representation of nested sets, conjunction and disjunction fallacies, as well as base-rate neglect are generally reduced. Yet, improving performance is one thing, proving that we are intuitive Bayesians is another. The adaptive redundancy that gives us flexibility and cognitive frugality can also lead to serious and systematic errors, a fate shared by Socrates and the slave alike.

Authors' Response

Base-rate respect: From statistical formats to cognitive structures

DOI: 10.1017/S0140525X07001963

Aron K. Barbey[a] and Steven A. Sloman[b]
[a]*Department of Psychology, Emory University, Atlanta, GA 30322;* [b]*Cognitive and Linguistics Science, Brown University, Providence, RI 02912.*
barbeya@nih.ninds.gov **Steven_Sloman@brown.edu**
http://www.cog.brown.edu/~sloman/

Abstract: The commentaries indicate a general agreement that one source of reduction of base-rate neglect involves making structural relations among relevant sets transparent. There is much less agreement, however, that this entails dual systems of reasoning. In this response, we make the case for our perspective on dual systems. We compare and contrast our view to the natural frequency hypothesis as formulated in the commentaries.

R1. Introduction

Updating Koehler's (1996) review of base-rate sensitivity in probability judgment, the target article reviewed a broad range of evidence in support of the nested sets hypothesis. The hypothesis proposes that people's ability to estimate the probability of A, given B, in a way that is consistent with Bayes' theorem depends, in part, on the transparency of the structural relations among the set of events of type A, relative to the set of events of type B. In particular, when the A set is perceived to be nested within the B set, judgments are more coherent than when the relation is not perceived (for an illustration, see Figure 1 of the target article). We contrast this proposal with the idea that facilitation reflects an evolutionary adaptation to process natural frequencies. The responses to our target article revealed a surprising degree of consensus on this issue, demonstrating much agreement that the transparency of structural relations is one important variable in reducing base-rate neglect. We also observed frequent doubt about the value of the dual systems perspective.

Among several insights about the natural frequency hypothesis and nested sets theory was the conclusion that there is more to probability judgment than these approaches address (**Beaman & McCloy**, **Girotto & Gonzalez**, **Griffin, Koehler, & Brenner** [**Griffin et al.**], **Laming**, **Schurr & Erev**, **Sun & Wang**, **Thomas**, **Uhlmann, Brescoll, & Pizarro** [**Uhlmann et al.**], **Whitney, Hinson, & Matthews** [**Whitney et al.**]). Indeed, by framing the nested sets hypothesis within the larger dual process theory of inference, judgment, and decision making, our proposal supports a broader understanding of probability judgment. We agree that the nested sets and dual process theories deserve greater specification and appreciate **Mandel**'s and **Samuels**'s efforts to unpack some of the assumptions of our proposal.

We have organized our response into two general categories: (1) those that address properties of the dual process hypothesis, and (2) those that concern the natural frequency approach.

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

R2. Dual process hypothesis

The dual process hypothesis has proved to be controversial. A number of commentators point out that we don't really argue for the dual systems perspective. We acknowledge that the target article relied primarily on earlier arguments in support of the framework (Evans & Over 1996; Sloman 1996a; Stanovich & West 2000), and we agree with **De Neys** that this framework deserves more careful testing especially with regard to its application to base-rate neglect.

We begin by addressing a common misconception about the hypothesis and reviewing our proposal that the nested sets hypothesis entails dual systems of reasoning. We then address the arguments of commentators who disagreed with us and summarize evidence offered in support of the dual systems framework.

R2.1. Rule-based versus associative ≠ normative versus counter-normative

Evans & Elqayam, **Gaissmaier, Straubinger, & Funder** (**Gaissmaier et al**.), and **Lagnado & Shanks** surprised us by attributing to us a claim that we did not make. We never did and never would align the dual processes of associative and rule-based reasoning with the normative versus non-normative distinction. Indeed, Sloman (1996a) explicitly denies such a parallel and points out that normative rules are only one kind of rule used by the rule-based system. Of course rules can lead to errors and of course associations frequently lead to correct responses; after all, people mostly do pretty well at achieving their goals. The only claim we made in the target article is that base-rate neglect can be remedied when elementary rules of set theory are applied. This is hardly a broad claim about how error prone each system is.

R2.2. On the relation between dual processes and nested sets

Of course, whether or not there are two systems remains an open question and, as **Keren, van Rooij, & Schul** (**Keren et al.**), **Mandel**, **Samuels**, and **Trafimow** point out, the claim is conceptually independent of the nested sets hypothesis. Nonetheless, the dual process hypothesis remains the simplest viable framework for motivating the nested sets hypothesis for several reasons.

First, dual process theory offers a general framework providing background assumptions to explain the variety of phenomena addressed by the nested sets hypothesis (see Table 2 of the target article). "Inaccurate frequency judgments," for example, result primarily from associative processes (see sect. 2.6 of the target article and **Fantino & Stolarz-Fantino**), whereas the facilitative role of set representations in deductive inference depends primarily on rule-based processes (see sect. 2.9 of the target article). We know of no account of the variety of predictions summarized in Table 2 that does not assume more than one cognitive system. The associative versus rule-based distinction has the advantage of providing a common account of these diverse phenomena and has proven useful for interpreting a variety of judgment effects (Kahneman & Fredericks 2002), especially probability judgment such as the conjunction fallacy (Tversky & Kahneman 1982b; 1983; cf. Sloman 1996b).

Second, in the absence of systematic studies that assess the role of associative and rule-based processes in probability judgment (a point made by **De Neys**), it can be argued that the facilitative effect of nested set representations on Bayesian inference results from (1) different rule-based processes, a possibility raised by **Gigerenzer & Hoffrage**, **Keren et al.**, and **Mandel**, or from (2) multiple associative processes, as **Lagnado & Shanks** argue. These proposals represent logical possibilities but, in the broad form in which they are stated, they have little empirical content. Consider, for example, Lagnado & Shanks's proposal that one associative system results in non-Bayesian responses, whereas another associative system is engaged when people "see" the set inclusion relations illustrated by Euler diagrams and draw the Bayesian solution. Lagnado & Shanks do not specify the associative processes that give rise to normative versus non-normative responses. How do associative processes implement the elementary set operations or whatever operations are responsible for Bayesian responding? We suspect that if the proposal were spelled out, they would end up with a dual process theory that includes associative and rule-based operations.

R2.3. Why rule-based and associative?

Various versions of a two-systems hypothesis have been offered. Our claim (in contrast to **Brainerd**, **Evans & Elqayam**, **Reyna & Mills**, and **Wolfe**) is that Sloman's (1996a) characterization of the associative system is consistent with cases of base-rate neglect (people rely on associations based on statistical regularities embodied by events in experience), and that his characterization of rule-based reasoning is consistent with reasoning during nested sets facilitation (deliberative reasoning about set relations based on rules of combination).

In support of our position, **Evans & Elqayam** point out that there is an association (the one asserted by Kahneman & Tversky [1973] in their original demonstration of the phenomenon) that explains the majority response, namely the association between the hypothesis being evaluated (the presence of breast cancer) and the case data that provide diagnostic information about it (the probability of a positive mammogram given breast cancer). Indeed, this observation supports our claim that associative operations often lead to base-rate neglect. In the context of the Medical Diagnosis problem, this occurs when judgments are based on the association between a positive mammogram and breast cancer, or, in Kahneman and Tversky's terms, when judgments reflect how *representative* a positive mammogram is of breast cancer (see sects. 1.2.5 and 2.3 of the target article).

Of course not *all* responses that neglect base rates are associative in this sense (as **Gigerenzer & Hoffrage** show convincingly). Prior to their assertion that all reasoning on this task is associative, **Lagnado & Shanks** point out that one response that is observed involves a rule: Sometimes people report the complement of the false-alarm rate. That is true. This response involves a subtraction from 1. People find the Medical Diagnosis problem and related problems very difficult, and use a host of different strategies in their struggle to generate a

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

reasonable answer. Many of those strategies involve rules. However, the appeal to a representative outcome does partially account for their response. We actually agree with most of **Evans & Elqayam** and **Macchi & Bagassi**'s description of what goes on when people try to solve the Medical Diagnosis problem. These commentators suggest, however, that the process they describe implies that errors are produced by a *pragmatic* system. We cannot see what explanatory purchase this provides. All systems of reasoning must be sensitive to pragmatics in order for their output to be relevant to reasoners' goals and concerns.

An alternative dual process theory is advocated by **Griffin et al.**, who propose that "The conditions under which . . . [nested set representations] promote base-rate use may be more parsimoniously organized under the broader notion of case-based judgment." We find Griffin et al.'s proposal intriguing but difficult to assess in the absence of detail concerning the cognitive operations that give rise to the "strength of impression of the case at hand," or in the absence of a proposal about how this construct is measured and differentiated from alternative accounts. The case-based theory does appear to be inconsistent with the large body of evidence we review demonstrating Bayesian inference facilitation by virtue of employing samples of category instances that would not seem to strengthen single-case impressions (see sect. 2 of the target article).

Griffin et al. suggest that all forms of base-rate facilitation can be explained in terms of single-case impressions. For instance, they argue that judgments drawn from Euler diagrams depend on case-specific information (see sect. 2.5 of the target article). According to their view, "Diagrams prompting an immediate comparison of the size of circles may allow a low-level perceptual computation to solve the problem." We suspect that facilitation by nested sets takes advantage of visual representations that allow us to see in the world, or in our mind's eye, the relation between relevant sets. But the nested sets hypothesis requires a number of additional steps involving symbol manipulation in order to apply this representation to solving a base-rate problem. First, each set must be labeled; second (as **Patterson** points out), the correct sets must be chosen; and third, a symbolic response (a number) must be generated. Thus, even in the context of diagrammatic representations, Bayesian inference cannot be reduced to "a low-level perceptual computation," without appealing to symbolic operations. Whatever the right theory may be to explain base-rate neglect and respect, these forms of symbol manipulation require an account and Griffin et al.'s proposal does not currently offer one. The case-based theory may explain some instances of facilitation that are outside the scope of the nested sets hypothesis, but it does not seem to be a substitute for it.

Brainerd and **Reyna & Mills** review evidence that supports a dual process theory of judgment, and, in the process, cover some of the history of the ideas that we neglected. These commentators offer the denominator neglect model of inductive inference, a special case of fuzzy trace theory, as an account of base-rate neglect. According to this view, errors in probabilistic inference result from the failure to represent and attend to all of the information present in a nested set relation, specifically the information captured by the denominator of a Bayesian ratio. While we obviously agree with the claim about nested sets, we are less comfortable associating what is and what is not neglected with terms of a mathematical expression. As we do not believe that the mind embodies a mental analogue of Bayes' theorem (see sect. 2.8 of the target article), we also do not believe that judgment errors correspond to neglect of terms of the theorem. Rather, we believe that, in cases of base-rate neglect, people are doing something other than trying to map statistical information from the problem onto a mathematical equation. Specifically, we believe that errors result from a failure to map the problem onto a mental representation of the conceptual relations among sets. According to the nested set hypothesis, representing conceptual relations among sets affords a natural mapping to a correct numerical response. In the case of rule-based processing, this requires several forms of symbol manipulation (e.g., combination rules) that operate from a qualitative representation of structural relations among sets (see sect. 1.2.5 of the target article).

Reyna & Mills distinguish their fuzzy-trace theory dual process model from our dual process account by stating that the former predicts that normative judgment results from associative processes (System 1 operations), whereas counter-normative judgment results from a focus on quantitative details (System 2 operations). This prediction appears to be inconsistent with the large body of evidence demonstrating that (a) under the right conditions, heuristics can produce systematic errors in judgment (for a recent review, see Gilovich et al. 2002), and (b) Bayesian facilitation is sometimes a result of deliberative, rule-based operations (see sect. 2 of the target article).

R2.4. Evidence in favor of dual process theory

Some of the commentaries provide strong arguments in support of our perspective. **Fantino & Stolarz-Fantino** show that base-rate neglect is well-captured by associative principles in the context of trial-by-trial presentations (they also provide additional support for our claim that natural frequency formats are not sufficient to eliminate base-rate neglect). On the flip side, **Over** shows that representation via nested sets is equivalent to the logic of the excluded middle. Taken together, these observations suggest that very different inferential principles apply to (at least some) cases of base-rate neglect and to cases of facilitation via nested set representations. The fact that different inferential principles apply does not entail that different systems of representation and processing apply, but the dual systems hypothesis does offer a simple explanation that is consistent with this and a host of other data (cf. Evans 2003).

Newell & Hayes, who offer several objections to our proposal, also point to results that favor our perspective. We agree that much can be learned from assessments of base-rate usage in the category-learning domain. Those data are well explained by an associative theory that takes account of differential attention to features (Kruschke & Johansen 1999). That is one reason why we appeal to associative processes to explain performance in the absence of additional structural cues. As Newell & Hayes point out, there is nothing in studies of category learning that corresponds to making nested sets

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

transparent. Of course, only if there were would the need arise to invoke the rules of nested set representations.

Patterson points to the generality of nested set representations and their potential role in deductive, modal, deontic, and causal reasoning. His proposal shares with Johnson-Laird's (1983) domain-general theory that these forms of inference are represented in terms of sets of possibilities. The prediction that nested set representations will facilitate deductive inference is supported by evidence reviewed in the target article (see sect. 2.9). Patterson's suggestion of assessing the descriptive validity of the Leibnizean principle (If all As are Bs, then everything that is related in manner R to an A is related in manner R to a B) has already been explored in the context of category induction (Sloman 1998). This research demonstrates that the Leibnizean principle is obeyed only when the category relation is made explicit – further implicating the role of nested set representations in reasoning. Although we agree with Patterson that representing subset relations can facilitate probability judgment and deductive reasoning, we are not optimistic that the nested sets theory will support a general framework for representing modal, deontic, and causal relations (Sloman 2005).

Butterworth and **Sun & Wang** review evidence addressing the cognitive and neural foundations of numeric processing. Sun & Wang provide evidence that the mind embodies a coarse number scale consisting of qualitative categories. The reviewed neuroimaging evidence demonstrates that exact calculations recruit the language system, whereas approximate calculations rely on visuo-spatial representations of numbers mediated by parietal areas. Butterworth suggests that the latter reflects a "classic Fodorian cognitive module," whereas Sun & Wang argue that together these systems may provide the neural foundations for the proposed dual systems theory (cf., Goel 2005). We agree that intuitive probability judgment depends on qualitative representations and find the cited neuroimaging evidence suggestive (for a recent review of the neuroscience literature on reasoning, see Barbey & Barsalou, in press).

Schurr & Erev and **Thomas** address the degree to which the reviewed findings generalize to real-world settings. Schurr & Erev raise an important distinction between decisions from description versus experience. In contrast to the underutilization of base-rates observed in decisions from description (see sects. 1 and 2 of the target article), Schurr & Erev argue that decisions from experience result in base-rate overweighting. Although Schurr & Erev make a convincing case for their proposal, we are not convinced that the cited example involves representing structural relations among sets. It seems rather to involve making alternative interpretations of a stimulus more available. It would be analogous, in the Medical Diagnosis problem, to suggesting that the positive result has a different interpretation. Although Schurr & Erev's proposal may not directly inform the nested sets hypothesis, the distinction they raise is certainly of value in its own right.

Uhlmann et al. and **Whitney et al.** offer important insights that extend the proposed dual process theory to include social-psychological and emotional factors. We appreciate Uhlmann et al.'s suggestion that motivations can moderate whether people rely on rule-based versus associative processes and believe that these factors should be incorporated into any complete theory of judgment. Whitney et al.'s proposal that the dual process theory can be integrated with the literature on affective influences on reasoning offers a worthwhile theoretical challenge.

R3. Natural frequencies or nested sets?

Our review of the natural frequency hypothesis is organized into four subsections. In the first, we attempt to clarify the intent and value of Gigerenzer and Hoffrage (1995). We then review the natural sampling framework, and address the definition of natural frequencies and their proposed equivalence to chance representations of probability.

R3.1. Clarifying the intent and value of Gigerenzer and Hoffrage (1995)

As **Kleiter** makes crystal clear, our intent was to argue that facilitation on base-rate problems often results from clarifying the structural relations among the relevant sets referred to in the problem. **Gigerenzer & Hoffrage** and **Brase** seem to concur, suggesting (happily) that there is wide agreement on this issue. Indeed, Gigerenzer & Hoffrage and **Barton, Mousavi, & Stevens** (**Barton et al.**) argue that this was always the intended meaning of Gigerenzer and Hoffrage (1995) and that they have been repeatedly misunderstood as having suggested that frequencies of any type arising through natural sampling are sufficient for facilitation. In fact, we were very careful to distinguish normalized from non-normalized frequencies, but we (like many others) believed that Gigerenzer and Hoffrage (1995) were trying to say something other than that there are computational advantages to what we have here described as the nested sets hypothesis (Tversky & Kahneman 1983).

On reading **Gigerenzer & Hoffrage**, we find it intriguing that so many researchers are guilty of the identical apparent misinterpretation of Gigerenzer and Hoffrage (1995). It might have to do with passages like the following one from Gigerenzer and Hoffrage (1995).

> Evolutionary theory asserts that the design of the mind and its environment evolve in tandem. Assume—pace Gould—that humans have evolved cognitive algorithms that can perform statistical inferences. These algorithms, however, would not be tuned to probabilities or percentages as input format, as explained before. For what information format were these algorithms designed? We assume that as humans evolved, the "natural" format was *frequencies* as actually experienced in a series of events, rather than probabilities or percentages. (p. 686, emphasis in original)

They also refer to the natural frequency hypothesis as "our evolutionary argument that cognitive algorithms were designed for frequency information acquired through natural sampling" (Gigerenzer & Hoffrage 1995, p. 699). Further quotes from that paper appear in our target article.

Gigerenzer & Hoffrage point out that the evolutionary argument has nothing to do with deriving predictions from the natural frequency hypothesis and here we agree. But it does not seem unreasonable to infer from their own language that these authors put scientific weight on the claim that there exists an evolved frequency-sensitive algorithm. Of course, our review also

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

makes clear – pace **Brase** – that Gigerenzer & Hoffrage's own theorizing has not been entirely consistent over the years (admittedly, neither has ours).

Nevertheless, we do not entirely agree that **Gigerenzer & Hoffrage**'s most recent proposal completely converges with the nested sets hypothesis. According to Gigerenzer & Hoffrage, "the question is how and why reasoning depends on the external representation of information." We believe that the critical question is: How and why does reasoning depend on the *internal* representation of information? Our hypothesis concerns mental representations. The natural frequency hypothesis, even in its new form, is "about the general question of how various external representations facilitate Bayesian computations" (Gigerenzer & Hoffrage's commentary). But the findings we review suggest to us that different external representations (e.g., natural frequencies, chances) map onto the same internal representation.

More specifically, **Gigerenzer & Hoffrage**'s theory is that different textual formats map onto different equations. We don't believe that the mind is composed of equations even in the form of algorithms. Rather, we believe that people invoke different combination rules in a highly context-specific way that depends on techniques they have learned or figured out themselves. The critical mapping process is not from text to mathematical equation, but rather, in the case of rule-based processing, from text to a qualitative representation of structural relations among sets.

Nested set structures do not simplify Bayesian computations themselves; rather they suggest a cognitive representation that affords simple computations. As a result, the nested sets hypothesis cannot be reduced to the equations cited in **Gigerenzer & Hoffrage**'s commentary. Furthermore, the additional predictions cited in their commentary do not bear on the reviewed findings as they suggest. Predictions 2, 3, and 4 are not addressed in our review because they do not distinguish between competing theoretical accounts – nor do they directly bear on the target article's assessment of Gigerenzer and Hoffrage's (1995) "evolutionary argument that cognitive algorithms were designed for frequency information" (p. 699).

R3.2. Does natural sampling support the natural frequency hypothesis?

Natural frequency theorists motivate the evolutionary argument that the mind is designed to process natural frequencies by appealing to the natural sampling framework (Kleiter 1994). As **Kleiter** makes clear in his commentary, however, the natural sampling framework is based on a statistical model that is not consistent with the psychological theory advocated by natural frequency theorists. In particular, the natural sampling framework depends on several assumptions that are rarely satisfied in the natural environment, including complete data, "additive frequencies in hierarchical tree-like sample/subsample structure," and random sampling (see also Kleiter 1994 and sect. 3.1 of the target article). Natural frequency theories that appeal to sequential sampling and evolutionary plausibility have little to do with natural sampling in Kleiter's original sense. Kleiter points out in the commentary that the assumptions of his framework are rarely satisfied in the natural environment and, as a result, the computational advantage of natural sampling has nothing to do with ecological validity.

R3.3. What are natural frequencies?

Barton et al., **Brase**, and **Gigerenzer & Hoffrage** argue that the simple frequencies employed by Brase (2002b)[1] do not represent natural frequencies. These commentators say that single numerical statements (e.g., 1 out of 2) are simple frequencies, whereas natural frequencies necessarily represent the structural relations among the operative sets, or, in their language, the structure of the entire tree diagram. This view is inconsistent with the description of natural frequencies in recent work such as that of Zhu and Gigerenzer (2006), in which the authors talk of "natural frequencies such as 1 out of 2" (p. 15). Moreover, for binary events single numerical statements can satisfy the definition of natural frequencies. Consider, for example, the single numerical statement "I win poker 1 out of 10 nights." This statement directly implies that "I lose poker on 9 out of 10 nights" and therefore represents the size of the reference class (e.g., "10 nights total"), in addition to the relevant subset relations (e.g., "1 night I win" and "9 nights I lose"), and thus the structure of the event tree.

The clarification offered by **Barton et al.**, **Brase**, and **Gigerenzer & Hoffrage** is helpful of course, for it indicates that the natural frequency theory, like nested sets, concerns the representation of the structural relations among the events. Both positions leave open questions about the conditions of base-rate neglect and respect: When should we expect judgments to be veridical? How much can people represent and what are the computational demands of different problems? We do not believe that ecological considerations or appeals to problem formats provide an answer to these questions. These questions require an analysis of *internal* mental representations, their power and demands, as well as the conditions that elicit them.

Brase raises a further objection concerning the conclusions drawn from Brase et al. (2006). He argues that we "try to infer cognitive abilities and structures from data showing that incentives affect performance." In fact, our conclusion about domain general cognitive processes does not depend on the findings Brase mentions concerning monetary incentives. Our claim is that "a higher proportion of Bayesian responses is observed in experiments that [...] select participants with a higher level of general intelligence ... [which is] consistent with the view that Bayesian reasoning depends on domain general cognitive processes to the degree that intelligence is domain general" (target article, sect. 2.2, para. 5).

R3.4. Are natural frequencies and chances equivalent representations?

By definition, chances refer to the likelihood of a single event (see **Barton et al.**) determined by a distribution of possibilities, not a sample of observations. Consequently, as we have pointed out in Note 4 of the target article, chances are not obtained by "counting occurrences of events as they are encountered and storing the resulting knowledge base for possible use later" (i.e., natural

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

sampling; Brase 2002b, p. 384). In this sense, chances are therefore distinct from natural frequencies. Yet **Gigerenzer & Hoffrage** and **Brase** propose that chances are equivalent to natural frequencies and, as a consequence, that the natural frequency hypothesis predicts that chance representations will facilitate Bayesian inference. We all seem to agree on what facilitates Bayesian inference, but broadening the definition of natural frequencies to include chances appears to undermine the claim that "cognitive algorithms were designed for frequency information, acquired through natural sampling" (Gigerenzer & Hoffrage 1995, p. 686; see target article sect. 1.2.2, para. 3).

We welcome the recent articulation of the natural frequency hypothesis and believe that the current formulation is a roughly accurate characterization of some of the conditions that lead to base-rate respect. By broadening the definition of natural frequencies to include chances, **Gigerenzer & Hoffrage**'s proposal implies that (1) cognitive algorithms were not designed over the course of human evolution to process natural frequencies rather than the likelihood or chance of a single event, (2) the theory no longer appeals to the natural sampling framework (which cannot encode chances), and, finally, (3) the findings are not motivated by the ecological rationality program, which claims that our current environment represents statistical information in the form of natural frequencies "as actually experienced in a series of events" rather than conveying the likelihood or chance of a single event (Gigerenzer & Hoffrage, 1995, p. 686). With these clarifications, we agree with Gigerenzer & Hoffrage and **Brase** that the natural frequency and nested sets hypotheses are hard to distinguish. Of course, as the key ideas have to do with structural relations among sets and not with frequency counts, we believe the term "nested sets" is more adequate descriptively.

R4. Conclusions

There are some real disagreements about how people judge probability. Some theorists believe that the cognitive machinery responsible for judgment is best described as associative, others appeal to simple rules, and others want to focus on the representation of single cases. There are also different views about the value of dual systems as a framework for theorizing. But what we have learned from the commentators' responses to our target article is that there is far more agreement than disagreement about the psychology of judgment and that much of the rhetoric about judgment is programmatic, reflecting pre-theoretic methodological commitments rather than substantive empirical claims. Indeed, almost everyone seems to agree that the empirical record supports the nested sets hypothesis – under one terminological guise or another – suggesting that the transparency of nested sets is one important variable in reducing base-rate neglect.

We wish there were such convergence in opinion about the theoretical prospects of systems for associative and rule-based reasoning, as well. We are still hopeful that the dual systems perspective will gain further support in time.

Beyond advocating a particular theoretical perspective or attempting to resolve a long-standing controversy, we hope that our target article helps propel research past the medical diagnosis task and its relatives, and away from pre-commitments to evolutionary theorizing or any other conceptual framework without solid empirical content. We hope instead to see more assessment of judgment with a focus on the many important questions that remain: What are the cognitive operations that underlie probability judgment across a range of real-world decision contexts and what cognitive, social, and emotional factors mediate the resulting estimates of confidence and probability? What conditions enable people to adapt and reason well in a world of change and uncertainty?

NOTE

1. Our motivation for reviewing the results of Brase (2002b) was to assess the comprehension of statistical formats typically employed in the Bayesian reasoning literature. As we state in the target article (sect. 2.7, para. 2), "Brase (2002b) conducted a series of experiments to evaluate the relative clarity and ease of understanding *a range of statistical formats*" (emphasis added). Our motivation was to assess natural frequencies and percentages in the form employed by research in the Bayesian reasoning literature (see sect. 2.7). Our summary of Brase (2002b) is accurate, pointing to the equivalence in the perceived clarity, impressiveness ratings, and impact on behavior that Brase reports for two formats. **Brase** notes in his commentary that "actual single event probabilities [e.g., 0.33] were *not* understood as well or as clearly as simple frequencies [e.g., 1 in 3] and relative frequencies [e.g., 33%]." That is true. But simple frequencies are normalized (see **Barton et al.**) and absolute frequencies – the "true" natural frequencies according to our reading of Brase – were judged just as unclear as single-event probabilities on average in all experiments but one (as far as we can tell; statistical tests were not reported).

References

[Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.]

Adam, M. B. & Reyna, V. F. (2005) Coherence and correspondence criteria for rationality: Experts' estimation of risks of sexually transmitted infections. *Journal of Behavioral Decision Making* 18(3):169–86. [VFR]

Aitchison, J. & Dunsmore, I. R. (1975) *Statistical prediction analysis*. Cambridge University Press. [GDK]

Antell, S. E. & Keating, D. P. (1983) Perception of numerical invariance in neonates. *Child Development* 54:695–701. [BB]

Ayton, P. & Wright, G. (1994) Subjective probability: What should we believe? In: *Subjective probability*, ed. G. Wright & P. Ayton, pp. 163–83. Wiley. [aAKB]

Barbey, A. K. & Barsalou, L.W. (in press) Reasoning and problem solving: Models. In: *New Encyclopedia of Neuroscience*, ed. L. Squire, T. Albright, F. Bloom, F. Gage & N. Spitzer. Elsevier. [rAKB]

Bar-Hillel, M. (1980) The base-rate fallacy in probability judgments. *Acta Psychologica* 44:211–33. [GG, WDN]

Barrett, H. C. & Kurzban, R. (2006) Modularity in cognition: Framing the debate. *Psychological Review* 113:628–47. [DT]

Bauer, M. I. & Johnson-Laird, P. N. (1993) How diagrams can improve reasoning. *Psychological Science* 4:372–78. [aAKB, VFR]

Birnbaum, M. H. (1983) Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology* 96:85–94. [LM]

Blanchette, I. (2006) The effect of emotion on interpretation and logic in a conditional reasoning task. *Memory and Cognition* 34:1112–25. [PW]

Bonato, M., Fabbri, S., Umiltà, C. & Zorzi, M. (in press) The mental representation of numerical fractions: Real or integer? *Journal of Experimental Psychology: Human Perception and Performance.* [BB]

Bradburn, N. M., Rips, L. J. & Shevell, S. K. (1987) Answering autobiographical questions: The impact of memory and inference on surveys. *Science* 236:157–61. [aAKB]

Brainerd, C. J. (1981) Working memory and the developmental analysis of probability judgment. *Psychological Review* 88:463–502. [VG]

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Brainerd, C. J. & Reyna, V. F. (1988) Generic resources, reconstructive processing, and children's mental arithmetic. *Developmental Psychology* 24:324–34. [CJB]
(1990) Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review* 10:365–403. [CJB, VFR]
(1995) Autosuggestibility in memory development. *Cognitive Psychology* 28:65–101. [CJB, VFR]
Brannon, E. M. & Terrace, H. S. (1998) Ordering of the numerosities 1 to 9 by monkeys. *Science* 282:746–49. [BB]
(2000) Representation of the numerosities 1–9 by rhesus macaques (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes* 26(1):31–49. [BB]
Brase, G. L. (2002a) Ecological and evolutionary validity: Comments on Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni's (1999) mental-model theory of extensional reasoning. *Psychological Review* 109:722–28. [aAKB, GLB]
(2002b) Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making* 15:381–401. [AB, arAKB, GLB]
(2007) The (in)flexibility of evolved frequency representations for statistical reasoning: Cognitive styles and brief prompts do not influence Bayesian inference. *Acta Psychologica Sinica* 39(3):398–405. [AB]
Brase, G. L. & Barbey, A. K. (2006) Mental representations of statistical information. In: *Advances in psychology research*, *vol. 41*, ed. A. Columbus, pp. 91–113. Nova Science. [GLB]
Brase, G. L., Cosmides, L. & Tooby, J. (1998) Individuation, counting, and statistical inference: The role of frequency and whole object representations in judgment under uncertainty. *Journal of Experimental Psychology: General* 127:3–21. [GLB]
Brase, G. L., Fiddick, L. & Harries, C. (2006) Participant recruitment methods and statistical reasoning performance. *The Quarterly Journal of Experimental Psychology* 59:965–76. [arAKB, GLB, WDN, DAL, DL, BN]
Brenner, L., Griffin, D. & Koehler, D. J. (2005) Modeling patterns of probability calibration with Random Support Theory: Diagnosing case-based judgment. *Organizational Behavior and Human Decision Processes* 97:64–81. [DG]
(2006) Case-based biases in asset pricing. Paper presented at the Society for Judgment and Decision Making Conference, Houston, TX, November 18–20, 2006. [DG]
Brescoll, V. L. & Uhlmann, E. L. (2007) System-justifying motivations underlie biological attributions for gender differences. Unpublished manuscript, Yale University. [ELU]
Bright, G. W., Behr, M. J., Post, T. R. & Wachsmuth, I. (1988) Identifying fractions on number lines. *Journal for Research in Mathematics Education* 19(3):215–32. [BB]
Bruner, J. S. & Minturn, A. L. (1955) Perceptual identification and perceptual organization. *Journal of General Psychology* 53:21–28. [AS]
Butterworth, B. (1999) *The mathematical brain*. Macmillan. [BB]
(2001) What seems natural? *Science* 292:853–54. [BB]
Calvillo, D. P. & Revlin, R. (2005) The role of similarity in deductive categorical inference. *Psychonomic Bulletin and Review* 12:938–44. [aAKB]
Cantlon, J. F., Brannon, E. M., Carter, E. J. & Pelphrey, K. A. (2006) Functional imaging of numerical processing in adults and 4-year-old children. *Public Library of Science Biology* 4(5):844–54. [BB]
Casscells, W., Schoenberger, A. & Graboys, T. B. (1978) Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine* 299:999–1000. [aAKB, WDN]
Castelli, F., Glaser, D. E. & Butterworth, B. (2006) Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *Proceedings of the National Academy of Science* 103(12):4693–98. [BB]
Chase, W. G. & Simon, H. A. (1973) Perception in chess. *Cognitive Psychology* 4:55–81. [WG]
Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT Press. [GLB]
Clark, A. & Thornton, C. (1997) Trading spaces: Computation, representation, and the limits of uninformed learning. *Behavioral and Brain Sciences* 20(1):57–92. [GK]
Clark, H. H. & Brennan, S. E. (1991) Grounding in communication. In: *Perspectives on socially shared cognition*, ed. L. B. Resnick, J. M. Levine & S. D. Teasley. American Psychological Association. [JCT]
Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187–276. [DEO]
Cosmides, L. & Tooby, J. (1994) Better than rational: Evolutionary psychology and the invisible hand. *American Economic Review* 84:327–32. [DG]
(1996) Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58:1–73. [AB, aAKB, BB, JStBTE, CRW]
(2003) Evolutionary psychology: Theoretical foundations. In: *Encyclopedia of cognitive science*, ed. L. Nadel, pp. 54–64. Macmillan. [GLB]
Cox, E. P., III. (1980) The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research* 17:407–22. [YS]
Crespi, L. P. (1942) Quantitative variation of incentive and performance in the white rat. *American Journal of Psychology* 55:467–517. [GLB]
(1944) Amount of reinforcement and level of performance. *Psychological Review* 51:341–57. [GLB]
Damasio, A. R. (1994) *Descartes' error: Emotion, reasoning, and the human brain*. G. P. Putnam. [PW]
Dawes, R. M. (1988) *Rational choice in an uncertain world*. Harcourt Brace Jovanovich. [LM]
Dehaene, S., Spelke, E., Pinel, P., Stanescu, R. & Tsivkin, S. (1999) Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science* 284(5416):970–74. [BB, YS]
De Neys, W. (2006a) Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology* 59:1070–100. [WDN]
(2006b) Dual processing in reasoning: Two systems but one reasoner. *Psychological Science* 17:428–33. [WDN]
Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M. & Lockhart, L. K. (1998) Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology* 75:53–69. [ELU]
Duchaine, B., Cosmides, L. & Tooby, J. (2001) Evolutionary psychology and the brain. *Current Opinion in Neurobiology* 11:225–30. [GLB]
Eddy, D. M. (1982) Probabilistic reasoning in clinical medicine: Problems and opportunities. In: *Judgment under uncertainty: Heuristics and biases*, ed. D. Kahneman, P. Slovic & A. Tversky, pp. 249–67. Cambridge University Press. [aAKB, DAL]
Edwards, W. (1968) Conservatism in human information processing. In: *Formal representation of human judgment*, ed. B. Kleinmutz, pp. 17–52. Wiley. [WG]
Elqayam, S. (2007) Normative rationality and the is-ought fallacy. In: *Proceedings of Second European Cognitive Science Society Conference* (*Delphi*, *Greece*, *May 23–27*, *2007*), ed. S. Vosniadou, D. Kayser & A. Protopapas, pp. 294–99. Erlbaum. [JStBTE]
Epstein, S. (1994) Integration of the cognitive and psychodynamic unconscious. *American Psychologist* 49:709–24. [PW]
Erev, I. & Barron, G. (2005) On adaptation, maximization and reinforcement learning among cognitive strategies. *Psychological Review* 112(4):912–31. [AS]
Erev, I., Shimonowitch, D., Schurr, A. & Hertwig, R. (2007) Base rates: How to make the intuitive mind appreciate or neglect them. In: *Intuition in judgment and decision making*, ed. H. Plessner, C. Betsch & T. Betsch, pp. 135–48. Erlbaum. [AS]
Ermer, E., Cosmides, L. & Tooby, J. (2007) Functional specialization and the adaptationist program. In: *The evolution of mind: Fundamental questions and controversies*, ed. S. Gangstead & J. Simpson. Guilford Press. [GLB]
Estes, W. K., Campbell, J. A., Hatsopoulos, N. & Hurwitz, J. B. (1989) Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:556–71. [aAKB]
Evans, J. (forthcoming) How many dual process theories do we need? One, two or many? In: *In two minds: Dual processes and beyond*, J. Evans & K. Frankish. Oxford University Press. [RS]
Evans, J. St. B. T. (2003) In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences* 7:454–59. [rAKB]
(2006) The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review* 13:378–95. [JStBTE]
(2007) *Hypothetical thinking: Dual process in reasoning and judgement*. Psychology Press. [JStBTE, DEO]
Evans, J. St. B. T., Handley, S. J., Over, D. E. & Perham, N. (2002) Background beliefs in Bayesian inference. *Memory & Cognition* 30:179–90. [aAKB, JStBTE, DAL]
Evans, J. St. B. T., Handley, S. J., Perham, N., Over, D. E. & Thompson, V. A. (2000) Frequency versus probability formats in statistical word problems. *Cognition* 77:197–213. [aAKB, JStBTE, GG]
Evans, J. St. B. T., Newstead, S. E. & Byrne, R. M. J. (1993) *Human reasoning: The psychology of deduction*, Ch. 7. Erlbaum. [RP]
Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and reasoning*. Psychology Press. [arAKB, GK, DEO]
Fabbri, S., Tang, J., Butterworth, B. & Zorzi, M. (submitted) The mental representation of fractions: Interaction between numerosity and proportion processing with non-symbolic stimuli. [BB]
Fantino, E., Kanevsky, I. G. & Charlton, S. (2005) Teaching pigeons to commit base-rate neglect. *Psychological Science* 16:820–25. [EF]
Fiedler, K. (2000) Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review* 107:659–76. [EF]
(2001) Affective influence on social information processing. In: *Handbook of affect and social cognition*, ed. J. P. Forgas, pp. 163–85. Erlbaum. [PW]

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Finucane, M. L., Alhakami, A., Slovic, P. & Johnson, S. M. (2000) The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making* 13:1–17. [PW]

Fodor, J. A. (1983) *Modularity of mind.* MIT Press. [aAKB, BB, GLB]

Fox, C. & Levav, J. (2004) Partition-edit-count: Naïve extensional reasoning in judgment of conditional probability. *Journal of Experimental Psychology: General* 133:626–42. [aAKB]

Funder, D. C. (1995) Stereotypes, base rates, and the fundamental attribution mistake: A content-based approach to judgmental accuracy. In: *Stereotype accuracy: Toward appreciating group differences*, ed. L. Jussim, Y-T. Lee & C. McCauley, pp. 141–56. American Psychological Association. [WG]

(1996) Base rates, stereotypes, and judgmental accuracy. *Behavioral and Brain Sciences* 19:22–23. [WG]

Gaissmaier, W., Schooler, L. J. & Rieskamp, J. (2006) Simple predictions fueled by capacity limitations: When are they successful? *Journal of Experimental Psychology: Learning, Memory and Cognition* 32:966–82. [WG]

Gelman, S. A. (2003) *The essential child: Origins of essentialism in everyday thought.* Oxford University Press. [ELU]

Gigerenzer, G. (1993) The superego, the ego, and the id in statistical reasoning. In: *G. A Handbook of Data Analysis in the Behavioral Sciences*, ed. G. Keren & G. Lewis, pp. 331–39. Erlbaum. [aAKB]

(1996) The psychology of good judgment: Frequency formats and simple algorithms. *Medical Decision Making* 16:273–80. [aAKB]

(2002) *Calculated risks: How to know when numbers deceive you*. Simon & Schuster. (UK version: *Reckoning with risk: Learning to live with uncertainty.* Penguin). [GG]

(2004) Fast and frugal heuristics: The tools of bounded rationality. In: *Blackwell handbook of judgment and decision making*, ed. D. Koehler & N. Harvey, pp. 62–88. Blackwell. [GG]

(2006) Ecological rationality: Center for Adaptive Behavior and Cognition summary of research area II. Retrieved October 1, 2006, from the Center for Adaptive Behavior and Cognition Web site: http://www.mpib-berlin.mpg.de/en/forschung/abc/forschungsfelder/feld2.htm [aAKB]

(2007) *Gut feelings: The intelligence of the unconscious*. Viking Press. [GG]

Gigerenzer, G., Hell, W. & Blank, H. (1988) Presentation and content: The use of base-rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance* 14:513–25. [aAKB, DG]

Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B. & Katsikopoulos, K. (2005) "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis* 25:623–29. [GG]

Gigerenzer, G. & Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102:684–704. [AB, arAKB, BB, CPB, GG, VG, WG, DL, DRM, YS, CRW]

(1999) Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review* 106:425–30. [GLB, GG, VG]

Gigerenzer, G., Hoffrage, U. & Ebert, A. (1998) AIDS counseling for low-risk clients. *AIDS Care* 10:197–211. [aAKB]

Gigerenzer, G. & Regier, T. P. (1996) How do we tell an association from a rule? *Psychological Bulletin* 119:23–26. [AB, aAKB, GG]

Gigerenzer, G. & Selten, R., eds. (2001) *Bounded rationality: The adaptive toolbox.* MIT Press. [aAKB]

Gigerenzer, G., Todd, P. & the ABC research group (1999) *Simple heuristics that make us smart.* Oxford University Press. [aAKB, WG]

Gilovich, T., Griffin, D. & Kahneman, D., eds. (2002) *Heuristics and biases: The psychology of intuitive judgment.* Cambridge University Press. [arAKB]

Ginossar, Z. & Trope, Y. (1987) Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology* 52:464–74. [ELU]

Girotto, V. & Gonzalez, M. (2001) Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition* 78:247–76. [AB, aAKB, DG, DRM, LM, VFR]

(2002) Chances and frequencies in probabilistic reasoning: Rejoinder to Hoffrage, Gigerenzer, Krauss, and Martignon. *Cognition* 84:353–59. [aAKB]

(in press) Children's understanding of posterior probability. *Cognition*. DOI:10.1016/j.cognition.2007.02.005. [aAKB, VG, VFR]

Gluck, M. A. & Bower, G. H. (1988) From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General* 117:227–47. [aAKB, DG, DAL, BN]

Goel, V. (2005) Cognitive neuroscience of deductive reasoning. In: *Cambridge Handbook of Thinking and Reasoning*, ed. K. Holyoak & R. Morrison. Cambridge University Press. [rAKB]

Goodie, A. S. & Fantino, E. (1995) An experientially derived base-rate error in humans. *Psychological Science* 6:101–106. [EF]

(1996) Learning to commit or avoid the base-rate error. *Nature* 380:247–49. [EF]

(1999) What does and does not alleviate base-rate neglect under direct experience. *Journal of Behavioral Decision Making* 12:307–35. [EF]

Gordon, P. (2004) Numerical cognition without words: Evidence from Amazonia. *Science* 306:496–99. [CRW]

Green, D. M. (1960) Psychoacoustics and detection theory. *Journal of the Acoustical Society of America* 32:1189–1203. [DL]

Green, D. M. & Swets, J. A. (1966) *Signal detection theory and psychophysics*. Wiley. [DL]

Grice, H. P. (1975) Logic and conversation. In: *Syntax and semantics: Vol. 3. Speech acts*, ed. P. Cole & J. L. Morgan. Academic Press. [LM]

(1978) Some further notes on logic and conversation. In: *Syntax and semantics. Vol. 3: Speech acts*, ed. P. Cole, pp. 225–42. Academic Press. [JCT]

Griffin, D. & Buehler, R. (1999) Frequency, probability, and prediction: Easy solutions to cognitive illusions? *Cognitive Psychology* 38:48–78. [aAKB, DAL]

Griffin, D. & Tversky, A. (1992) The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24:411–35. [DG]

Griggs, R. A. & Newstead, S. (1982) The role of problem structure in a deductive reasoning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:297–307. [aAKB, DEO]

Grossen, B. & Carnine, D. (1990) Diagramming a logic strategy: Effects on difficult problem types and transfer. *Learning Disability Quarterly* 13:168–82. [aAKB]

Hammerton, M. (1973) A case of radical probability estimation. *Journal of Experimental Psychology* 101:252–54. [aAKB]

Hartl, J. A. & Fantino, E. (1996) Choice as a function of reinforcement ratios in delayed matching to sample. *Journal of the Experimental Analysis of Behavior* 66:11–27. [EF]

Hartnett, P. & Gelman, R. (1998) Early understandings of numbers: Paths or barriers to the construction of new understandings? *Learning and Instruction* 8(4):341–74. [BB]

Hauser, M., MacNeilage, P. & Ware, M. (1996) Numerical representations in primates. *Proceedings of the National Academy of Sciences USA* 93:1514–17. [BB]

Hertwig, R., Barron, G., Weber, E. U. & Erev, I. (2004) Decisions from experience and the weighting of rare events. *Psychological Science* 15(8):534–539. [AS]

Hinson, J. M., Whitney, P., Holben, H. & Wirick, A. K. (2006) Affective biasing of choices in gambling task decision making. *Cognitive, Affective, and Behavioral Neuroscience* 6:190–200. [PW]

Hoffrage, U. & Gigerenzer, G. (1998) Using natural frequencies to improve diagnostic inferences. *Academic Medicine* 73:538–40. [GG]

Hoffrage, U., Gigerenzer, G., Krauss, S. & Martignon, L. (2002) Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition* 84:343–52. [aAKB, GLB]

Hoffrage, U., Lindsey, S., Hertwig, R. & Gigerenzer, G. (2000) Communicating statistical information. *Science* 290:2261–62. [GG]

Hooper, P. M. (2007) Exact beta distributions for query probabilities in Bayesian networks. Technical Report. Department of Mathematical and Statistical Sciences, University of Alberta, Canada. [GDK]

Howson, C. & Urbach, P. (2006) *Scientific Reasoning: The Bayesian Method*, 3rd edition. Open Court. [DAL]

Jenkins, H. M. & Ward, W. C. (1965) Judgment of contingency between responses and outcomes. *Psychological Monographs* 79, No. 1 (Whole No. 594). [EF]

Johansen, M. K., Fouquet, N. & Shanks, D. R. (in press) Paradoxical effects of base rates and representation in category learning. *Memory & Cognition*. [DAL]

Johnson-Laird, P. N. (1983) *Mental models: Towards a cognitive science of language, inference, and consciousness.* Harvard University Press. [rAKB]

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S. & Caverni, J.-P. (1999) Naïve probability: A mental model theory of extensional reasoning. *Psychological Review* 106:62–88. [aAKB, GG, DRM]

Jost, J. T. & Banaji, M. R. (1994) The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology* 33:1–27. [ELU]

Juslin, P., Wennerholm, P. & Winman, A. (2001) High level reasoning and base-rate use: Do we need cue-competition to explain the inverse base-rate effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:849–71. [DAL]

Kahneman, D. (2003) Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review* 93(5):1449–75. [AS]

Kahneman, D. & Frederick, S. (2002) Representativeness revisited: Attribute substitution in intuitive judgment. In: *Heuristics and biases: The psychology of intuitive judgment*, ed. T. Gilovich, D. Griffin & D. Kahneman, pp. 49–81. Cambridge University Press. [arAKB, JStBTE, GK]

(2005) A model of heuristic judgment. In: *The Cambridge Handbook of Thinking and Reasoning*, ed. K. J. Holyoak & R. G. Morris, pp. 267–93. Cambridge University Press. [aAKB]

Kahneman, D. & Lovallo, D. (1993) Timid theories and bold forecasts: A cognitive perspective on risk taking. *Management Science* 39:17–31. [GG]

Kahneman, D. & Tversky, A. (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3:430–54. [GG]

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

(1973) On the psychology of prediction. *Psychological Review* 80:237–51. [arAKB, JStBTE, DG]
(1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–91. [PW]
(1996) On the reality of cognitive illusions. *Psychological Review* 103:582–91. [aAKB]
Keil, F. C. (1989) *Concepts, kinds, and cognitive development*. MIT Press. [ELU]
Keren, G. & Schul, Y. (under review) Two is not always better than one: A critical evaluation of two-systems theories. [GK]
Keren, K. & Thijs, L. J. (1996) The base-rate controversy: Is the glass half-full or half empty? *Behavioral and Brain Sciences* 19:26. [aAKB]
Kleiter, G. D. (1994) Natural sampling: Rationality without base-rates. In: *Contributions to mathematical psychology, psychometrics, and methodology*, ed. G. H. Fischer & D. Laming, pp. 375–88. Springer-Verlag. [arAKB, GDK, DEO]
(1996) Propagating imprecise probabilities in Bayesian networks. *Artificial Intelligence* 88:143–61. [GDK]
Kleiter, G. D. & Kardinal, M. (1995) A Bayesian approach to imprecision in belief nets. In: *Symposia Gaussiana: Proceedings of the Second Gauss Symposium, Conference B: Statistical Sciences*, ed. V. Mammitzsch & H. Schneeweiß, pp. 91–105. De Gruyter. [GDK]
Kleiter, G. D., Krebs, M., Doherty, M. E., Gavaran, H., Chadwick, R. & Brake, G. B. (1997) Do subjects understand base-rates? *Organizational Behavior and Human Decision Processes* 72:25–61. [aAKB, GDK]
Koehler, J. J. (1996) The base-rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences* 19:1–53. [arAKB, CPB, GK, DRM]
Kruglanski, A. W. & Dechesne, M. (2006) Are associative and propositional processes qualitatively distinct? Comment on Gawronski and Bodenhausen (2006). *Psychological Bulletin* 132:736–39. [DT]
Kruschke, J. K. (1996) Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22:3–26. [BN]
(2001) The inverse base-rate effect is not explained by eliminative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:1385–400. [DAL]
Kruschke, J. K. & Johansen, M. K. (1999) A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 25:1083–119. [rAKB]
Kurzenhauser, S. & Hoffrage, U. (2002) Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher* 24:516–21. [aAKB]
Lakatos, I. (1977) *The methodology of scientific research programmes: Philosophical papers, vol. 1*. Cambridge University Press. [GK]
Laming, D. (1968) *Information theory of choice-reaction times*. Academic Press. [DL]
(2001) Statistical information, uncertainty, and Bayes' theorem: Some applications in experimental psychology. In: *Symbolic and quantitative approaches to reasoning with uncertainty. Lecture Notes in Artificial Intelligence, vol. 2143*, ed. S. Benferhat & P. Besnard, pp. 635–46. Springer-Verlag. [DL]
(2004) *Human judgment: The eye of the beholder*. Thomson Learning. [DL]
Larkin, J. H. & Simon, H. A. (1987) Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11:65–99. [WG, GK]
Lee, D. (2006) Neural basis of quasi-rational decision making. *Current Opinion in Neurobiology* 16:191–98. [PW]
Lemer, C., Dehaene, S., Spelke, E. & Cohen, L. (2003) Approximate quantities and exact number words: dissociable systems. *Neuropsychologia* 41(14):1942–58. [BB]
Levesque, H. J. (1986) Making believers of computers. *Artificial Intelligence* 30:81–108. [DEO]
Levinson, S. C. (1995) Interactional biases in human thinking. In: *Social intelligence and interaction*, ed. E. Goody. Cambridge University Press. [LM]
(2000) *Presumptive meanings*. MIT Press. [LM]
Lewin, K. (1931) The conflict between Aristotelian and Galileian modes of thought in contemporary psychology. *Journal of Genetic Psychology* 5:141–77. [DRM]
Lewis, C. & Keren, G. (1999) On the difficulties underlying Bayesian reasoning: Comment on Gigerenzer and Hoffrage. *Psychological Review* 106:411–16. [GG]
Lichtenstein, S. & Slovic, P. (1973) Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology* 101:16–20. [DL]
Lindley, D. V. (1985) *Making decisions*. Wiley. [LM]
Lindsey, S., Hertwig, R. & Gigerenzer, G. (2003) Communicating statistical DNA evidence. *Jurimetrics* 43:147–63. [aAKB]
Linton, M. (1975) Memory for real-world events. In: *Explorations in cognition*, ed. D. A. Norman & D. E. Rumelhart, pp. 376–404. Freedman Press. [aAKB]
(1982) Transformations of memory in everyday life. In: *Memory observed*, ed. U. Neisser, pp. 77–91. Freedman Press. [aAKB]
Lloyd, F. J. & Reyna, V. F. (2001) A web exercise in evidence-based medicine using cognitive theory. *Journal of General Internal Medicine* 16(2):94–99. [VFR]
Lloyd, F., Reyna, V. F. & Whalen, P. (2001) Accuracy and ambiguity in counseling patients about genetic risk. *Archives of Internal Medicine* 161:2411–13. [VFR]
Loewenstein, G. (2005) Hot-cold empathy gaps in medical decision making. *Health Psychology* 24:S49–S56. [PW]
Loewenstein, G. F., Weber, E. U., Hsee, C. K. & Welch, N. (2001) Risk as feelings. *Psychological Bulletin* 127:267–86. [PW]
Macchi, L. (1995) Pragmatic aspects of the base rate fallacy. *The Quarterly Journal of Experimental Psychology* 48 A(1):188–207. [LM]
(2000) Partitive formulation of information in probabilistic problems: Beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes* 82:217–36. [aAKB, GG, LM]
(2003) The partitive conditional probability. In: *Thinking: Psychological perspectives on reasoning, judgment and decision making*, ed. D. Hardman & L. Macchi, pp. 165–87. Wiley. [LM]
Macchi, L., Bagassi, M. & Ciociola, P. B. (2007) The pragmatic approach versus the dual-process theories on probabilistic reasoning. (Unpublished manuscript.) [LM]
Macchi, L. & Mosconi, G. (1998) Computational features vs frequentist phrasing in the base-rate fallacy. *Swiss Journal of Psychology* 57:79–85. [GG, LM]
Mack, N. (1995) Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research Mathematics Education* 26:422–41. [BB]
Mandel, D. R. (in press) Violations of coherence in subjective probability: A representational and assessment processes account. *Cognition*. doi:10.1016/j.cognition.2007.01.001. [DRM]
Marr, D. (1982) *Vision: A computational investigation into the human representation and processing visual information*. W. H. Freeman. [GK]
Martignon, L., Vitouch, O., Takezawa, M. & Forster, M. R. (2003) Naive and yet enlightened: From natural frequencies to fast and frugal decision trees. In: *Thinking: Psychological perspectives on reasoning, judgment and decision making*, ed. D. Hardman & L. Macchi, pp. 189–211. Wiley. [GG]
Matsuzawa, T. (1985) Use of numbers by a chimpanzee. *Nature* 315:57–59. [BB]
McCloy, R., Beaman, C. P., Morgan, B. & Speed, R. (2007) Training conditional and cumulative risk judgments: The role of frequencies, problem-structure and einstellung. *Applied Cognitive Psychology* 21:325–44. [CPB]
McCloy, R. Byrne, R. M. J. & Johnson-Laird, P. N. (submitted) Understanding cumulative risk. [CPB]
McDell, J., Uhlmann, E. L., Omoregie, H. & Banaji, M. R. (2006) The psychological correlates of Bayesian racism. Poster presented at the Society of Personality and Social Psychology meeting, Palm Springs, CA. [ELU]
McKee, R. (1997) *Story: Substance, structure, style and the principles of screenwriting*. Harper-Collins. [JCT]
Medin, D. L. & Edelson, S. M. (1988) Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General* 117:68–85. [DAL, BN]
Mellers, B. & McGraw, A. P. (1999) How to improve Bayesian reasoning: Comments on Gigerenzer & Hoffrage (1995) *Psychological Review* 106:417–24. [aAKB]
Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81–97. [YS]
Monaghan, P. & Stenning, K. (1998) Effects of representational modality and thinking style on learning to solve reasoning problems. In: *Proceedings of the 20th Annual Meeting of the Cognitive Science Society*, ed. M. A. Gernsbacher & S. J. Derry, pp. 716–21. Erlbaum. [aAKB]
Morton, J. (1968) A singular lack of incidental learning. *Nature* 215:203–204. [CPB]
Newstead, S. E. (1989) Interpretational errors in syllogistic reasoning. *Journal of Memory and Language* 28:78–91. [aAKB, BN]
Newstead, S. E., Handley, S. J., Harley, C., Wright, H. & Farelly, D. (2004) Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology* 57A:33–60. [JStBTE]
Ni, Y. & Zhou, Y. (2005) Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist* 40:27–52. [BB]
Nieder, A. (2005) Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience* 6(3):1–14. [BB]
Nisbett, R. E. & Wilson, T. D. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3):231–59. [GK]
Nosofsky, R. M., Kruschke, J. K. & McKinley, S. C. (1992) Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:211–33. [aAKB]

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Oaksford, M. & Chater, N. (2007) *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press. [GDK]
Over, D. E. (2000a) Ecological rationality and its heuristics. *Thinking and Reasoning* 6:182–92. [aAKB]
(2000b) Ecological issues: A reply to Todd, Fiddick, & Krauss. *Thinking and Reasoning* 6:385–88. [aAKB]
(2003) From massive modularity to meta-representation: The evolution of higher cognition. In: *Evolution and the psychology of thinking: The debate*, ed. D. E. Over, pp. 121–44. Psychology Press. [aAKB]
(2007) Content-independent conditional inference. In: *Integrating the mind: Domain general versus domain specific processes in higher cognition*, ed. M. J. Roberts, pp. 83–103. Psychology Press. [aAKB]
Over, D. E. & Evans, J. St. B. T. (forthcoming) Conditionals and non-constructive reasoning. In: *The psychology of conditionals*, ed. M. Oaksford. Oxford University Press. [Based on talks available at London Reasoning Workshop site: http://www.bbk.ac.uk/psyc/staff/academic/moaksford/londonreasoningworkshop] [DEO]
Over, D. E. & Green, D. W. (2001) Contingency, causation, and adaptive inference. *Psychological Review* 108:682–84. [aAKB]
Patterson, R. (l985) *Aristotle's modal logic*. Cambridge University Press. [RP]
Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993) *The adaptive decision maker*. Cambridge University Press. [GG]
Peters, E. & Slovic, P. (2000) The springs of action: Affective and analytic information processing in choice. *Personality and Social Psychology Bulletin* 26:1465–75. [PW]
Pfeifer, N. & Kleiter, G. D. (2005) Towards a mental probability logic. *Psychologica Belgica* 45:71–99. [GDK]
PHLS Communicable Disease Surveillance Centre. (1995–1996) *Communicable Disease Report*, Monthly edition. London: Communicable Disease Surveillance Centre, Public Health Laboratory Service. [DL]
Piazza, M., Izard, V., Pinel, P., Le Bihan, D. & Dehaene, S. (2004) Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44:547–55. [BB]
Piazza, M., Mechelli, A., Butterworth, B. & Price, C. (2002) Are subitizing and counting: implemented as separate or functionally overlapping processes? *NeuroImage* 15(2):435–46. [BB]
Piazza, M., Mechelli, A., Price, C. J. & Butterworth, B. (2006) Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research* 1106:177–88. [BB]
Piazza, M., Pinel, P., Le Bihan, D. & Dehaene, S. (2007) A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* 53(2):293–305. [BB]
Pillemer, E. D., Rhinehart, E. D. & White, S. H. (1986) Memories of life transitions: The first year in college. *Human Learning: Journal of Practical Research and Applications* 5:109–23. [aAKB]
Pinker, S. (1990) A theory of graph comprehension. In: *Artificial intelligence and the future of testing*, ed. R. Freedle, pp. 73–126. Erlbaum. [WG]
Pizarro, D. A. & Uhlmann, E. L. (2005) Do normative standards advance our understanding of moral judgment? [Commentary]. *Behavioral and Brain Sciences* 28:558–59. [ELU]
Plato (2006) *Meno*. In: *Plato's Meno*, ed. D. Scott. Cambridge University Press. [CRW]
Ramsey, F. P. (1964) Truth and probability. In: *Studies in subjective probability*, ed. H. E. Kyburg, Jr. & E. Smokler, pp. 61–92. Wiley. [aAKB]
Reber, A. S. (1993) *Implicit learning and tacit knowledge*. Oxford University Press. [JStBTE]
Reyna, V. F. (1991) Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review* 11:317–36. [aAKB, CJB, VFR]
(1992) Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In: *Development of long-term retention*, ed. M. L. Howe, C. J. Brainerd & V. F. Reyna, pp. 103–27. Springer-Verlag. [VFR]
(2004) How people make decisions that involve risk. A dual-processes approach. *Current Directions in Psychological Science* 13:60–66. [VFR]
Reyna, V. F. & Adam, M. B. (2003) Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. *Risk Analysis* 23:325–42. [VFR]
Reyna, V. F. & Brainerd, C. J. (1990) Fuzzy processing in transitivity development. *Annals of Operations Research* 23:37–63. [CJB]
(1992) A fuzzy-trace theory of reasoning and remembering: Paradoxes, patterns, and parallelism. In: *From learning processes to cognitive processes: Essays in honor of William K. Estes*, ed. A. Healy, S. Kosslyn & R. Shiffrin, pp. 235–59. Erlbaum. [aAKB]
(1993) Fuzzy memory and mathematics in the classroom. In: *Memory in everyday life*, ed. G. M. Davies & R. H. Logie, pp. 91–119. North Holland Press. [CJB, VFR]
(1994) The origins of probability judgment: A review of data and theories. In: *Subjective probability*, ed. G. Wright & P. Ayton, pp. 239–72. Wiley. [aAKB, CJB, VFR]
(1995) Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences* 7:1–75. [aAKB, CJB, VFR, CRW]
(in press) Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*. [VFR]
Reyna, V. F. & Ellis, S. C. (1994) Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science* 5:275–79. [VFR]
Reyna, V. F. & Farley, F. (2006) Risk and rationality in adolescent decision-making: Implications for theory, practice, and public policy. *Psychological Science in the Public Interest* 7(1):1–44. [VFR]
Reyna, V. F., Holliday, R. & Marche, T. (2002) Explaining the development of false memories. *Developmental Review* 22:436–89. [VFR]
Reyna, V. F. & Lloyd, F. (2006) Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. *Journal of Experimental Psychology: Applied* 12:179–95. [VFR]
Reyna, V. F., Lloyd, F. J. & Brainerd, C. J. (2003) Memory, development, and rationality: An integrative theory of judgment and decision-making. In: *Emerging perspectives on judgment and decision research*, ed. S. Schneider & J. Shanteau, pp. 201–45. Cambridge University Press. [VFR]
Reyna, V. F., Lloyd, F. & Whalen, P. (2001) Genetic testing and medical decision making. *Archives of Internal Medicine* 161:2406–408. [VFR]
Reyna, V. F., Mills, B. A., Estrada, S. M. & Brainerd, C. J. (2006) False memory in children: Data, theory, and legal implications. In: *The handbook of eyewitness psychology: Memory for events*, ed. M. P. Toglia, J. D. Read, D. F. Ross & R. C. L. Lindsay, pp. 473–510. Erlbaum. [VFR]
Rieskamp, J. & Otto, P. E. (2006) SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General* 135:207–36. [GG]
Sanfey, A. G., Loewenstein, G., McClure, S. M. & Cohen, J. D. (2006) Neuroeconomics: Cross currents in research on decision-making. *Trends in Cognitive Sciences* 10:108–16. [PW]
Savage L. J. (1954) *The foundations of statistics*. Wiley. [aAKB]
Schaller, M. (1992) In-group favoritism and statistical reasoning in social inference: Implications for formation and maintenance of group stereotypes. *Journal of Personality and Social Psychology* 63:61–74. [ELU]
Schwartz, N. & Sudman, S. (1994) *Autobiographical memory and the validity of retrospective reports*. Springer Verlag. [aAKB]
Schwartz, N. & Wanke, M. (2002) Experimental and contextual heuristics in frequency judgment: Ease of recall and response scales. In: *Etc. Frequency processing and cognition*, ed. P. Sedlmeier & T. Betsch, pp. 89–108. Oxford University Press. [aAKB]
Sedlmeier, P. & Betsch T. (2002) *Etc. Frequency processing and cognition*. Oxford University Press. [aAKB]
Sedlmeier, P. & Gigerenzer, G. (2001) Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General* 130:380–400. [aAKB]
Sedlmeier, P., Hertwig, R. & Gigerenzer, G. (1998) Are judgments of the positional frequencies of letters systematically biased due to availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:754–70. [CPB]
Sidanius, J. & Pratto, F. (1999) *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press. [ELU]
Sloman, S. A. (1996a) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [arAKB, JStBTE, GG, GK]
(1996b) The probative value of simultaneous contradictory belief: Reply to Gigerenzer and Regier (1996). *Psychological Bulletin* 119:27–30. [GG, rAKB]
(1998) Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology* 35:1–33. [arAKB]
(2005) *Causal models: How we think about the world and its alternatives*. Oxford University Press. [arAKB]
Sloman, S. A., Lombrozo, T. & Malt, B. C. (in press) Mild ontology and domain-specific categorization. In: *Integrating the mind*, M. J. Roberts. Psychology Press. [aAKB]
Sloman, S. A. & Over, D. E. (2003) Probability judgment from the inside and out. In: *Evolution and the psychology of thinking: The debate*, ed. D. E. Over, pp. 145–70. Psychology Press. [aAKB]
Sloman, S. A., Over, D. E., Slovak, L. & Stibel, J. M. (2003) Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes* 91:296–309. [aAKB, CPB, VFR]
Slovic, P., Finucane, M. L., Peters, E. & MacGregor, D. G. (2004) Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24:311–22. [PW]
Slovic, P., Peters, E., Finucane, M. L. & MacGregor, D. G. (2005) Affect, risk, and decision making. *Health Psychology* 24:S35–S40. [PW]
Smith, C. L., Solomon, G. E. A. & Carey, S. (2005) Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology* 51:101–40. [BB]

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press

Smith, E. R. & DeCoster, J. (2000) Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review* 4:108–31. [JStBTE]
Sommers, F. (l982) *The logic of natural language*. Oxford University Press. [RP]
Stafylidou, S. & Vosniadou, S. (2004) The development of student's understanding of the numerical value of fractions. *Learning and Instruction* 14:508–18. [BB]
Stanovich, K. E. (1999) *Who is rational? Studies of individual differences in reasoning*. Erlbaum. [aAKB, JStBTE]
(2004) *The robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press. [RS]
(2006) Is it time for a tri-process theory. Distinguishing the reflective and the algorithmic mind. In: *In two minds: Dual process theories of reasoning and rationality*, ed. J. St. B. T. Evans & K. Frankish. Oxford University Press. [JStBTE]
Stanovich, K. E. & West, R. F. (1998a) Individual differences in rational thought. *Journal of Experimental Psychology: General* 127:161–88. [aAKB, WDN]
(1998b) Who uses base rates and P(D/ ¬ H)? An analysis of individual differences. *Memory and Cognition* 28:161–79. [JStBTE]
(2000) Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences* 23:645–726. [arAKB, GK, DRM, LM, WDN]
Starkey, P. & Cooper, R. G., Jr. (1980) Perception of numbers by human infants. *Science* 210:1033–35. [BB]
Stolarz-Fantino, S. & Fantino, E. (1990) Cognition and behavior analysis: A review of Rachlin's *Judgment, Decision, & Choice*. *Journal of the Experimental Analysis of Behavior* 54:317–22. [EF]
(1995) The experimental analysis of reasoning: A review of Gilovich's *How We Know What Isn't So*. *Journal of the Experimental Analysis of Behavior* 64:111–16. [EF]
Stolarz-Fantino, S., Fantino, E. & Van Borst, N. (2006) Use of base rates and case cue information in making likelihood estimates. *Memory and Cognition* 34:603–18. [EF]
Sun, R., Slusarz, P. & Terry, C. (2005) The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review* 112:159–92. [JStBTE]
Sun, Y., Wang, H., Zhang, J. & Smith, J. W. (in press) Probabilistic judgment on a coarser scale. *Cognitive Systems Research*. [YS]
Tanner, T. A., Rauk, J. A. & Atkinson, R. C. (1970) Signal recognition as influenced by information feedback. *Journal of Mathematical Psychology* 7:259–274. [DL]
Tanner, W. P., Jr., Swets, J. A. & Green, D.M. (1956) Some general properties of the hearing mechanism. University of Michigan: Electronic Defense Group, Technical Report 30. [DL]
Tetlock, P. E., Kristel, O. V.;Elson, S. B., Green, M. C. & Lerner, J. S. (2000) The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology* 78:853–70. [ELU]
Thomas, E. A. C. & Legge, D. (1970) Probability matching as a basis for detection and recognition decisions. *Psychological Review* 77:65–72. [DL]
Thomas, J. C. (1978) A design-interpretation analysis of natural English. *International Journal of Man-Machine Studies* 10:651–68. [JCT]
Todd, P. M. & Gigenenzer, G. (1999) What we have learned (so far). In: *Simple heuristics that make us smart*, ed. G. Gigerenzer, P. M. Todd, & the ABC Research Group, pp. 357–65. Oxford University Press. [DEO]
Tooby, J. & Cosmides, L. (2005) Conceptual foundations of evolutionary psychology. In: *The handbook of evolutionary psychology*, ed. D. M. Buss, pp. 5–67. Wiley. [GLB]
Tooby, J., Cosmides, L. & Barrett, H. C. (2005) Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In: *The innate mind: Structure and content*, ed. P. Carruthers, S. Laurence & S. Stich. Oxford University Press. [GLB]
Toplak, M. E. & Stanovich, K. E. (2002) The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical reasoning skill. *Journal of Educational Psychology* 94:197–209. [DEO]
Trafimow, D. (2003) Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review* 110:526–35. [DT]
Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–31. [aAKB, BB]
(1980) Causal schemata in judgments under uncertainty. In: *Progress in social psychology*, ed. M. Fishbein, pp. 49–72. Erlbaum. [DG]
(1982a) Evidential impact of base rates. In: *Judgment under uncertainty: Heuristics and biases*, ed. D. Kahneman, P. Slovic & A. Tversky, pp. 153–60. Cambridge University Press. [EF]
(1982b) Judgement under uncertainty: Heuristics and biases. In: *Judgement under uncertainty: Heuristics and biases*, ed. D. Kahneman, P. Slovic & A. Tversky, pp. 3–20. Cambridge University Press. [rAKB]
(1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90:293–315. [arAKB, CJB, CRW]
Villejoubert, G. & Mandel, D. R. (2002) The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. *Memory & Cognition* 30:171–78. [DRM]
Wagenaar, W. A. (1986) My memory: A study of autobiographical memory over six years. *Cognitive Psychology* 18:225–52. [aAKB]
(1988) *Paradoxes of gambling behaviour*. Erlbaum. [DL]
Ward, W. C. & Jenkins, H. M. (1965) The display of information and the judgment of contingency. *Canadian Journal of Psychology/Review of Canadian Psychology* 19:231–41. [EF]
Winer, G. A. (1980) Class-inclusion reasoning in children: A review of the empirical literature. *Child Development* 51:309–28. [CJB]
Wolfe, C. R. (1995) Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory account. *Journal of Behavioral Decision Making* 8:85–108. [CRW]
Wolfe, C. R. & Reyna, V. F. (under review) The effects of analogy and transparency of set relations on probability judgment: Estimating conjunctions, disjunctions, and base rates. [CRW]
Wynn, K., Bloom, P. & Chiang, W. C. (2002) Enumeration of collective entities by 5-month-old infants. *Cognition* 83(3):B55–B62. [BB]
Yamagishi, K. (2003) Facilitating normative judgments of conditional probability: Frequency or nested sets? *Experimental Psychology* 50:97–106. [aAKB, VFR]
Zacks, R. T. & Hasher, L. (2002) Frequency processing: A twenty-five year perspective. In: *Etc. Frequency processing and cognition*, ed. P. Sedlmeier & T. Betsch, pp. 21–36. Oxford University Press. [aAKB]
Zhu, L. & Gigerenzer, G. (2006) Are children intuitive Bayesians? *Cognition* 77:197–213. [GG]
(2006) Children can solve Bayesian problems: The role of representation in mental computation. *Cognition* 98:287–308. [DEO, rAKB]

https://doi.org/10.1017/S0140525X07001860 Published online by Cambridge University Press