PROOF COVER SHEET

Author(s): Elise Lesage, Gorka Navarrete and Wim De Neys

Article title: Evolutionary modules and Bayesian facilitation: The role of general cognitive resources

Article no: PTAR_A_713177

Enclosures: 1) Query sheet
2) Article proofs

Dear Author,

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication.

Please limit changes at this stage to the correction of errors. You should not make insignificant changes, improve prose style, add new material, or delete existing material at this stage. Making a large number of small, non-essential corrections can lead to errors being introduced. We therefore reserve the right not to make such corrections.

For detailed guidance on how to check your proofs, please see
http://journalauthors.tandf.co.uk/production/checkingproofs.asp.

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your name will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1		Elise	Lesage	
2		Gorka	Navarrete	
3		Wim De	Neys	

Queries are marked in the margins of the proofs.

AUTHOR QUERIES

General query: You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. (Please see http://journalauthors.tandf.co.uk/preparation/permission.asp.) Please check that any required acknowledgements have been included to reflect this.

AQ1: Please give departments for affiliations 2 and 3
AQ2: Cosmides & Tooby, 1998: This reference is not listed in the reference list. Please supply a full reference entry.
AQ3: A symbol may have become corrupted; please check throughout
AQ4: Van Lier et al., 2012 submitted: Please update if possible
AQ5: Bonnefon et al., 2012 submitted: Please update if possible
AQ6: Bonnefon et al., 2012 submitted: Please update if possible
AQ7: Van Lier et al., 2012 submitted: Please update if possible

Taylor & Francis
Author Services

How to make corrections to your proofs using Adobe Acrobat

Taylor & Francis now offer you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can edit the proof directly using Adobe Acrobat. This is the simplest and best way for you to ensure that your corrections will be incorporated. If you wish to do this, please follow these instructions:

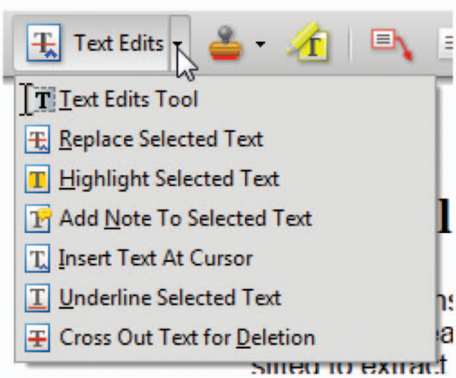
1. Save the file to your hard disk.

2. Check which version of Adobe Acrobat you have on your computer. You can do this by clicking on the "Help" tab, and then "About".
If Adobe Reader is not installed, you can get the latest version free from http://get.adobe.com/reader/.

- If you have Adobe Reader 8 (or a later version), go to "Tools"/ "Comments & Markup"/ "Show Comments & Markup".

- If you have Acrobat Professional 7, go to "Tools"/ "Commenting"/ "Show Commenting Toolbar".

3. Click "Text Edits". You can then select any text and delete it, replace it, or insert new text as you need to. It is also possible to highlight text and add a note or comment.



4. Make sure that you save the file when you close the document before uploading it to CATS using the "Upload File" button on the online correction form. A full list of the comments and edits you have made can be viewed by clicking on the "Comments" tab in the bottom left-hand corner of the PDF.

If you prefer, you can make your corrections using the CATS online correction form.

THINKING & REASONING, 2012, 00 (0), 1–27

Psychology Press
Taylor & Francis Group

Evolutionary modules and Bayesian facilitation: The role of general cognitive resources

Elise Lesage[1], Gorka Navarrete[2], and Wim De Neys[3]

AQ1 [1]School of Psychology, University of Birmingham, UK
[2]?????????????????, Universidad de la Laguna, San Cristóbal de La Luguna, Tenerife, Spain
[3]????????????????, CNRS, LaPsyDE CNRS FRE 3521, Paris Descartes University, Paris, France

Although decisions based on uncertain events are critical in everyday life, people perform remarkably badly when reasoning with probabilistic information. A well-documented example is performance on Bayesian reasoning problems, where people fail to take into account the base-rate. However, framing these problems as frequencies improves performance spectacularly. Popular evolutionary theories have explained this facilitation by positing a specialised module that automatically operates on natural frequencies. Here we test the key prediction from these accounts, namely that the performance of the module functions independently from general-purpose reasoning mechanisms. In three experiments we examined the relationship between cognitive capacity and performance on Bayesian reasoning tasks in various question formats, and experimentally manipulated cognitive resources in a dual task paradigm. Results consistently indicated that performance on classical Bayesian reasoning tasks depends on participants' available general cognitive capacity. Findings challenge the postulation of an automatically operating frequency module.

Keywords: Cognitive capacity; Reasoning; Bayesian facilitation.

Most of the decisions we make in life are based on uncertain information. In contexts ranging from picking a holiday destination that will maximise the odds of good weather to deciding on a cancer treatment that will maximise

Correspondence should be addressed to Elise Lesage, School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK. E-mail: e.lesage@bham.ac.uk

http://www.psypress.com/tar http://dx.doi.org/10.1080/13546783.2012.713177

the chance of survival, people need to consider the consequences of different options based on imperfect predictors. Given the ubiquity of such decisions, one might assume that people are good at weighting the information at hand appropriately.

Until the second half of the twentieth century the idea that human reasoners were good statistical thinkers was indeed taken for granted in the scientific community. The conception of the human mind was heavily influenced by the views of the Enlightenment, in which rationality played a key role. Mathematics, including the calculus of probability, was assumed to be the extension of human reasoning (Laplace, 1902).

In the 1970s Tsversky and Kahneman (1974) proposed that people use a limited number of heuristic rules when reasoning under uncertainty. Generally these heuristics are useful in simplifying information and providing a ball-park answer, but they are susceptible to systematic biases, called "cognitive illusions" (Kahneman & Tversky, 1996). One of these cognitive illusions is base-rate neglect, which refers to a tendency to ignore or grossly underweight the base-rate of a target event when calculating its probability (Tversky & Kahneman, 1982).

Since it was first described by Meehl and Rosen (1955), base-rate neglect has proved a very robust phenomenon that has been replicated in a wide array of populations (Bar-Hillel, 1980; Casscells, Schoenberger, & Grayboys, 1978; Lyon & Slovic, 1976) and under various different conditions (Borgida & Brekke, 1981; Kassin, 1979). Classically, base-rate neglect is investigated with reasoning problems in which a posterior probability has to be calculated according to Bayes' theorem. Eddy (1982) investigated base-rate neglect in the clinical diagnosis of breast cancer and first formulated the "mammography problem". A version of this problem (Gigerenzer & Hoffrage, 1995) goes as follows:

> The probability of breast cancer in the population is 1% for a woman aged 40 who participates in a routine screening. If the woman has breast cancer, the probability is 80% that she will have a positive mammography. If a woman does not have breast cancer, the probability is 9.5% that she will also have a positive mammography. A woman in this age group had a positive mammography in a routine screening.

What is the probability that she actually has breast cancer ? ___ % (1)

The provided information can be written according to Bayes' theorem:

$$P(H|D) = \frac{P(D|H) \times P(H)}{P(D)} \tag{2}$$

where H is the hypothesis that the woman in question has breast cancer, and D is the evidence for this; the positive mammography. To solve this reasoning problem one has to calculate the posterior probability, $P(H|D)$, by combining the probability that H is true with the probability that the test would be positive, given that H is true, $P(D|H)$, divided by the probability of a positive test, $P(D)$.

The correct answer is just under 8% (7.77% to be precise). Typically, when people are presented with these problems, a majority ignores the base-rate $P(H)$, and reports the hit rate $P(D|H)$, here 80%. This leads to a systematic overestimation of the sought probability.

The finding that the vast majority of the population does not manage to solve this task is more than a mere statistical curiosity. People's apparent inability to infer the correct conclusions from provided statistical information poses a clear problem for informed decision making in a large number of important real-life situations. Base-rate neglect can mean the difference between life and death in medical contexts, such as when a patient decides to undergo a risky treatment (Fenton & Neil, 2010; Lloyd, 2001), or legal contexts, such as when a jury considers the weight of a lie-detector test. Following these findings, consensus emerged that the human mind does not reason according to the rules of probability (Bar-Hillel, 1980; Kahneman & Tversky, 1973).

An influential explanation for base-rate neglect was proposed by Gigerenzer (1994) and Cosmides and Tooby (1996). These authors argue that the human mind has not evolved to process single-event probabilities, which are used in these problems, because these are not part of a "natural environment" (Gigerenzer, 1994; Cosmides & Tooby, 1996); i.e., an environment without explicit knowledge about the calculus of probability. Systematic biases such as base-rate neglect are then actually artefacts, caused by experimental material that lacks ecological validity. By contrast, presenting the information as frequencies would be a much better match to a natural environment. A core feature of this Ecological Rationality Framework is the idea that reasoning is governed by highly specialised modules, rather than by a flexible general-purpose reasoning apparatus.

The key features of this view borrow from Marr (1982) and Fodor (1983). Fodor (1983) proposes that the mind is composed of different modules for different functions that have evolved largely independently. While Fodor himself did not put forward automaticity or innateness as necessary properties of a module (Coltheart, 1999), other authors have attributed AQ2 these features to cognitive modules (Cosmides & Tooby, 1996, 1998, 2008). This modularity viewpoint opposes classical theories of reasoning that assume reasoning is governed by one flexible single-purpose reasoning system (Cosmides & Tooby, 1994a; Cosmides, Barrett, & Tooby, 2010).

Marr (1982) posits that one can find out the properties of a given mechanism by examining the ecological context in which it has evolved. The fact that the human mind has evolved to function in the manner it currently does, by definition implies a good fit with the environment in which it has evolved. One should then ask oneself what properties a reasoning mechanism should have to have "survived" all this time. This adaptationist approach has been criticised on the grounds that the inherently speculative hypotheses are not falsifiable and that alternative explanations which are not based on natural selection are dismissed (Gould & Lewontin, 1979), but has nevertheless proved very popular. Cosmides and Tooby (1994a, 1994b) propose an evolutionary approach to a wide array of questions in cognitive psychology and have applied these ideas to the subject of base-rate neglect (Brase, Cosmides, & Tooby, 1998; Cosmides & Tooby, 1996). In the Pleistocene, the authors argue, all statistical information would have been acquired through experience. Predictions such as "The probability of a successful hunt is 0.20" would have made no sense. On the other hand there would have been information in the form of frequencies. Tracking naturally occurring sample frequencies, or natural sampling, enables relatively good decision making (Kleiter, 1994). One would have access to memories of hunts both failed and successful, and this information—available as natural frequencies—would have enabled a group to make an informed decision about whether or not to go hunting. Therefore the mind would not have evolved to reason and make decisions based on single-event probabilities, as this kind of information was simply not available. By contrast, natural frequency information has been ubiquitous throughout the course of evolution. Cosmides and Tooby (1996) conclude that a probabilistic reasoning module will compute the correct answer when presented with natural frequencies, but not when presented with single-event probabilities. For instance, a natural frequency variant of the mammography problem goes as follows (Gigerenzer & Hoffrage, 1995):

> 10 out of every 1000 women at age 40 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 40 who got a positive mammography in routine screening.

> How many of these women do you expect to actually have breast cancer? _outof_." (3)

In line with the predictions of the Ecological Rationality Framework, a large body of evidence demonstrates that a question format presenting frequencies instead of probabilities reduces systematic biases such as base-rate neglect. Tversky and Kahneman (1983) and Fiedler (1988) first reported a dramatic decrease in the occurrence of the conjunction fallacy—a violation of the rules of probability similar to base-rate neglect—when the information in the problem was presented as frequencies instead of probabilities. Other studies (Cosmides & Tooby, 1996; Gigerenzer, 1994; Gigerenzer & Hoffrage, 1995) have demonstrated that the presentation in natural frequencies also greatly reduces base-rate neglect in Bayesian word problems. This robust facilitation of Bayesian inference under natural frequency conditions has been named Bayesian facilitation (Sloman & Over, 2003). Not all authors support the specific modularity viewpoint and the strong emphasis on evolutionary theory adopted by Cosmides and Tooby (1996). However, despite differences in approach and background, the Ecological Rationality Framework supporters have in common that they posit that the human mind is capable of reasoning in accordance to Bayes' theorem, but that the algorithm that automatically weights probabilistic information needs the correct input—natural frequencies—in order to do so.

A more recent alternative explanation for the Bayesian facilitation is Nested Sets Theory. The idea that some frequency problems are easier because their nested set structure is transparent was first proposed by Tversky and Kahneman (1982) and has recently regained popularity (e.g., Barbey & Sloman, 2007; Sloman, Over, Slovak, & Stibel, 2003). The central argument behind this theory is that the difference between probabilities and frequencies is not what makes the standard probability format more difficult than the natural frequency format. There is another difference between these two formats that is crucial for Bayesian facilitation. Essential to solving the reasoning problem is the fact that some sets of events are nested. For example, in the classic mammography problem the number of hits is nested within the number of women with cancer and the number of false alarms is nested within the number of healthy women (see Figure 1). When this is clear, the correct answer is relatively easily calculated by dividing the hit rate by the sum of the hit rate and the false-alarm rate. Nested Sets Theory posits that any manipulation of the problem that draws attention to this nesting of events will therefore facilitate reasoning and reduce base-rate neglect.

In the standard probability format—see (1) above—the hit rate and the false-alarm rate are given *relative* to the number of ill and healthy women (i.e., 80% of the 1% women with cancer; 9.6% of the 99% healthy women). The argument is that presenting the subset relative to the set makes it more difficult to see how many events are in a subset and how the sets of events relate. By contrast, in the natural frequency format—see (3) above—the *absolute* number of hits and false alarms is given (8 out of the observed 1000; 95 out of the observed 1000). The hit-rate and false-alarm rate are not normalised with respect to the cancer or health rate. As a result, the relationship between the different sets of events is much clearer (Evans,

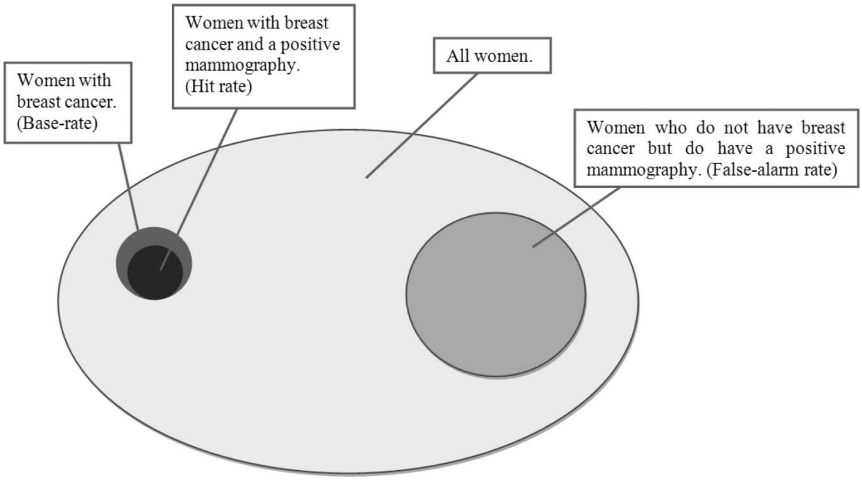


Figure 1. Nested set structure of the mammography problem.

Handley, Perham, Over, & Thompson, 2000; Sloman & Over, 2003). Nested Sets Theory argues that if the reasoning problem is construed in a manner that clarifies the set structure this will result in Bayesian facilitation, whether the problem is expressed as single-event probabilities or as frequencies. In support of Nested Sets Theory, manipulations that draw attention to the set structure, by providing diagrams (Sloman et al., 2003; Yamagishi, 2003) or by presenting non-normalised single-event probabilities (Fiedler, Brinkmann, Betsch, & Wild, 2000; Neace, Michaud, Bolling, Deer, & Zevevic, 2008), have been shown to lead to a decrease in systematic biases such as base-rate neglect.**

Barbey and Sloman (2007) incorporate Nested Sets Theory within the influential dual process framework. Dual processing accounts of reasoning and decision making posit that human reasoning is mediated by two distinct systems, often referred to as the heuristic system and the analytic system (Evans, 2006). There is a myriad of dual-process theories that differ slightly in the properties they attribute to both systems, but a relative consensus exists on a number of key characteristics (Evans, 2003, 2008). Analytic processes are conscious, slow, and effortful. The analytic system has a limited capacity and its performance is related to individual differences in fluid intelligence and working memory capacity (Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Stanovich, 1999). On the other hand, the heuristic system is composed of several processes that work fast and automatically, require no effort, and have an infinitely high capacity. This system exhibits the characteristics of the heuristic mind (Tversky & Kahneman, 1974) that provides a fast approximate answer, but is vulnerable to systematic biases. A popular conceptualisation of how these two systems interact is that the heuristic system generates a default answer that ‘‘pops’’ into consciousness, and can be overridden by the analytic system, provided enough executive resources are available (Evans, 2003). Barbey and Sloman (2007) incorporate Nested Sets Theory in this dual process framework to explain Bayesian facilitation. They argue that natural frequency conditions prompt a representation in terms of nested sets that triggers the analytical system. Executive resources are then recruited to calculate the correct answer. A standard probability format obscures the set representation of the problem and therefore does not trigger the analytical system. With the analytical system unable to apply the appropriate rules, the mind is left with the default answer provided by the heuristic system.

To date, the debate between these two explanations of the Bayesian facilitation phenomenon has not been resolved (Barbey & Sloman, 2007; De Neys, 2007; Over, 2007). Here we contrast the predictions of the Ecological Rationality Framework and those of Nested Sets Theory. While Nested Sets Theory offers clear-cut testable hypotheses, the predictions of the Ecological Rationality Framework are more elusive because its proponents differ in the degree to which they view the mind as being composed of independently evolved modules. For example, some authors place less emphasis on evolutionary theory, or do not make strong predictions about the automaticity of the proposed frequency modules (Barrett, Frederick, Haselton, & Kurzban, 2006; Barrett & Kurzban, 2006; Gigerenzer & Hoffrage, 1995). In the present article we derive predictions from a strict interpretation of Ecological Rationality, as proposed by Cosmides and Tooby (2008, p. 66):

> When activated by content from the appropriate domain, these inference engines impose special and privileged representations during the process of situation interpretation, define specialised goals for reasoning tailored to their domain, and make available specialised inferential procedures that allow certain computations to proceed *automatically* or ‘intuitively’ and with enhanced efficiency over what a more general reasoning process could achieve given the same input.

It is this interpretation of Ecological Rationality, including the concept of a specialised, automatically operating, independently evolved module, to which we refer in this article. A core differential prediction of the two explanations concerns the role of available general executive cognitive resources. The Ecological Rationality Framework posits that people perform better under natural frequency conditions because a specialised module automatically processes natural frequencies. As this module by definition functions independently of general-purpose cognitive resources, no relation between cognitive capacity and performance on Bayesian reasoning tasks is predicted; even people with relatively low cognitive resources will be able to reach the correct conclusion when presented with natural frequencies. Nested Sets Theory (Barbey & Sloman, 2007), borrowing from dual process theories of reasoning, posits that formats that make the set structure of the problem explicit will trigger the analytical system. This system will then use executive cognitive resources to compute the correct answer. Thus people perform better on the natural frequency format because they are able to use their analytic capacity. It follows that under conditions that prompt a clear representation of the nested sets, reasoning performance will be related to individual differences in general cognitive capacity: the more resources that are available, the more likely that the correct response will be computed. By contrast, problem formats that obscure the set structure (e.g., formats that normalise the probabilities), will not activate the analytical system. Instead the heuristic system will produce an answer that results from automatic associative processes and is subject to base-rate neglect. Nested Sets Theory thus predicts a positive relation between reasoning performance and general cognitive capacity in conditions that clarify the set structure of the problem (De Neys, 2007). In sum, under facilitating conditions, the Ecological Rationality Framework does not predict a relationship between performance and cognitive capacity, while Nested Sets Theory predicts a positive relationship.

In this study we tested the role of general-purpose cognitive resources in Bayesian facilitation. In three experiments we investigated how cognitive capacity is related to performance on Bayesian reasoning tasks. In the first experiment we tested for the relation between cognitive capacity and Bayesian reasoning performance. Participants were presented with a measure of cognitive capacity and solved classical Bayesian reasoning tasks in one of eight question formats. The question formats were systematically varied according to three features that have been shown to result in Bayesian facilitation in previous studies. Information was either presented in probabilities or frequencies (Cosmides & Tooby, 1996; Gigerenzer, 1994), hit-rate and false-alarm rates were given relative to the superordinate set or in absolute numbers (Barbey & Sloman, 2007; Fiedler et al., 2000; Neace et al., 2008; Yamagishi, 2003), and the total sample size was mentioned or not (Kassin, 1979; Kleiter, 1994). The Ecological Rationality Framework predicts no correlation between performance and executive capacity; by definition there should be no relation between individual differences on a measure of general-purpose cognitive capacity and individual differences in the efficiency of an automatically operating, specialised frequency module. By contrast, Nested Sets Theory predicts that in conditions that facilitate reasoning, the cognitive capacity–performance relationship will be stronger than under conditions that do not facilitate performance, because the limited-resource analytic system would be triggered in the former, but not in the latter.

The second experiment looked at the influence of cognitive capacity from a developmental angle. General cognitive capacity is known to increase throughout adolescence (Gathercole, 1999; Gathercole, Pickering, Ambridge, & Wearing, 2004). Secondary school students from 12 to 19 years old were divided into three age groups (i.e., early, middle, and late adolescents). These three age categories were used as an indirect index of cognitive capacity. Two features of the question format—whether the information was presented as frequencies and whether the presented information was normalised—were manipulated as in the first experiment. Nested Sets Theory predicts a positive correlation between performance and age: as cognitive capacity increases, it becomes more likely the analytical system will manage to compute the correct answer. Specifically, the correlation with age should be more pronounced under conditions that facilitate Bayesian reasoning, as the analytical system is triggered under these circumstances. Modular Ecological Rationality accounts on the other hand would predict that Bayesian facilitation will be relatively stable across secondary school students, as an automatically operating frequency module would not be hindered by the more limited cognitive capacity resources in younger age groups.[1]

The first two experiments looked at the correlation between Bayesian reasoning performance and general cognitive capacity. However, this is not sufficient to infer a causal role for general-purpose cognitive resources in performance on Bayesian reasoning tasks. Therefore in Experiment 3 executive resources were directly burdened using a dual task paradigm. Introducing a secondary task is an effective way to experimentally assess whether a process depends on general purpose cognitive resources (e.g., see De Neys, 2006a; De Neys & Schaeken, 2007; De Neys, Schaeken, & d'Ydewalle, 2005; De Neys & Verschueren, 2006; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Sloman, 1996). In this third experiment focus will be on the classical natural frequency format—see (3)—which has been show to elicit Bayesian facilitation most strongly and most reliably. According to the Ecological Rationality Framework the correct Bayesian answer is computed automatically by a specialised module when people are presented with natural frequencies. It follows that performance should not be hindered by a secondary task that burdens available general-purpose executive resources. Nested Sets Theory posits that when information is presented in a natural frequency format, the limited-capacity analytical system will kick in and recruit general-purpose executive resources to compute the correct answer. If the cognitive burden imposed by the

[1]Assuming the participant is old enough to read and understand the written material, of course. As our youngest participants were 12-year-old high school students, one can be reasonably confident that this condition was met.

secondary task leaves insufficient capacity to correctly perform the task, performance should drop under dual task conditions.

EXPERIMENT 1

Method

Participants. A total of 363 first-year psychology students from the University of Leuven, Belgium, participated in exchange for course credit.

Materials and design. The design was a 2 × 2 × 2 factorial design, in which three features of the question format were independently manipulated. The factors were (1) Frequency (information was presented as frequencies or as probabilities), (2) Absoluteness (information was presented either in an absolute, non-normalised manner, or relative to the superordinate set, i.e., normalised manner) and (3) Total Sample (the total sample size was either explicitly given or not). Participants were randomly assigned to one of these eight conditions.

Each participant was asked to solve two Bayesian reasoning problems. The first was an adapted version of the breast cancer problem (Eddy, 1982), the second, the lung disease problem, was an equivalent medical diagnosis problem of the same format. First, some background information about the reasoning problem was provided. For the mammography problem the background information was the following:

> Try to solve the following reasoning problem. Mammography is a screening method that allows us to detect very small lumps in breasts. This method is used to help diagnose breast cancer early. A positive mammography means that a lump was found. A negative mammography means no lump was detected.

Then the reasoning problem was presented in one of the eight formats. For instance, the format with absolute (i.e., non-normalised) probabilities went as follows (see the supplementary materials for a complete overview of all presented formats of the two problems):

> The study contains data from a large number of women. 99% of the women did not have breast cancer and 1% had breast cancer. Of the women without breast cancer, 10% have a positive mammography and 90% had a negative mammography. Of the women with breast cancer, 80% had a positive mammography and 20% had a negative mammography. What is the probability of breast cancer, if a woman has a positive mammogram?

Procedure. Students were tested in groups of 20 to 30 participants. Each participant was asked to solve two reasoning problems of the same question format. We used the Cognitive Reflection Test (CRT; Frederick, 2005) as a proxy to measure individual differences in general executive cognitive capacity. The CRT is a three-question cognitive ability test which measures the "ability or disposition to resist reporting the response that first comes to mind" (Frederick, 2005), or the tendency for "miserly processing" (Toplak, West, & Stanovich, 2011). The CRT is a potent predictor of a wide range of intelligence and executive function tests, as well as performance on a variety of reasoning tasks (Toplak et al., 2011). The test shows remarkably good correlations with classic cognitive abilities tests (e.g., $r = .43$ with the Wonderlic Personnel Test and $r = .46$ with the ACT).

The following is an example of a CRT item:

> In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days.

For all three questions, the correct solution requires the inhibition of impulsive erroneous answers. It is this response override requirement that heavily taxes general executive resources (Frederick, 2005).

The reasoning problems and the CRT were solved in a booklet (see supplementary materials). Participants went sequentially through the booklet and were not allowed to go back to a previous problem.

Results and discussion

Participants who had not answered the two problems or had not filled out the CRT were excluded from analysis ($N = 22$). The analyses were conducted on the data of the 341 remaining participants. Answers were regarded correct if the deviation from the correct answer was under 1%. The mean number of correct responses across the two problems was taken as the reasoning performance index and dependent variable.

A first set of analyses was aimed at verifying that effects classically found with these question formats were present in the data. The effect of the Frequency, Absoluteness, and Total Sample manipulations on performance
AQ3 was assessed by means of χ^2 tests for independence. Both the Frequency manipulation (17% vs 47% correct answers), $\chi^2(2) = 52.6$, $p < .001$, and the Absoluteness manipulation (21% vs 43% correct answers), $\chi^2(2) = 22$, $p < .001$, resulted in a significant increase of correct responses (see Figure 2). Information regarding the total sample was not associated with a significant increase in performance (35% vs 30% correct answers), $\chi^2(2) = 2.2$, $p = .338$. These results are shown in Table 1. Taken together, these findings imply that expressing the information in a frequency format is not a necessary condition for Bayesian facilitation, but it is a sufficient condition. This result is

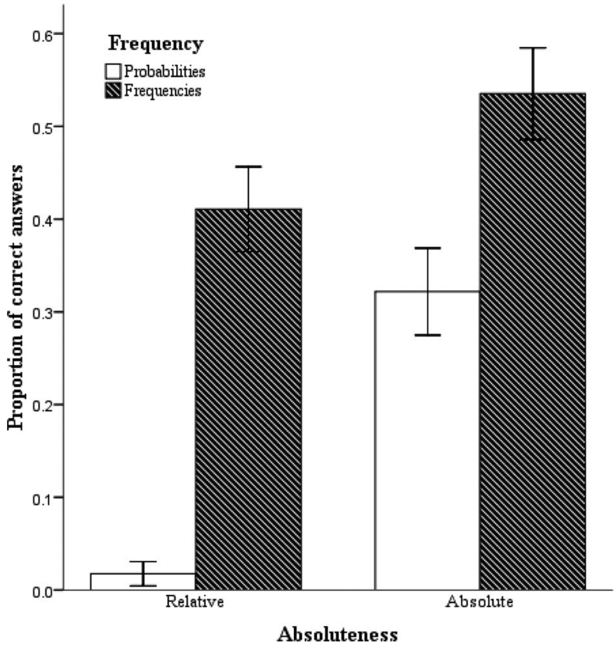


Figure 2. Effect of Frequency and Absoluteness on reasoning performance in Experiment 1. Error bars denote ± 1 standard error of the mean.

TABLE 1
Proportion of correct answers and number of participants in each condition in Experiment 1

Condition			*Proportion correct*	*Standard error*	*N*
Total Sample	Probabilities	Relative	0%	0%	42
		Absolute	33%	7%	44
	Frequencies	Relative	43%	7%	41
		Absolute	42%	7%	44
No Total Sample	Probabilities	Relative	4%	3%	43
		Absolute	31%	7%	43
	Frequencies	Relative	40%	6%	43
		Absolute	66%	7%	41

congruent with the studies by Yamagishi (2003) and Neace et al. (2008), which showed that although a clarified set structure was enough to improve performance, frequency format had an added facilitating effect.

Our central question concerned the association between cognitive capacity, as measured by CRT, and the reasoning performance index. Nested Sets Theory predicts that such an association will be present, and that it will be stronger in conditions where Bayesian facilitation is observed. By contrast, the Ecological Rationality Framework does not predict that cognitive capacity is associated with the performance. Correlation analyses revealed a significant association (Spearman's rho = 0.254, $p < .001$) between participants' CRT scores and Bayesian reasoning performance. This relationship between reasoning performance and our general cognitive capacity measure argues in favour of a general-purpose reasoning mechanism that is recruited in Bayesian reasoning, and against an independently evolved module that automatically operates on frequency information. Interestingly, this association was systematically stronger in conditions that elicited more Bayesian facilitation (see Table 2). In the natural frequency condition (absolute frequencies) and the standard probability condition (relative probabilities) the association between general cognitive ability and Bayesian reasoning performance was respectively the strongest (Spearman's rho = 0.48, $p < .001$) and the weakest (Spearman's rho = 0.17, $p = .13$). The more pronounced link between reasoning performance and CRT scores in conditions that elicit Bayesian facilitation supports a Nested Sets Theory view whereby facilitatory formats allow for cognitive resources to be recruited.

EXPERIMENT 2

Method

Participants. A total of 214 secondary school students (age range 12–19 years) participated on a voluntary basis. Participants were divided into three age groups; young (grade 7/8, mean age 13.3 years), middle (grade 9/10; mean age 15.2 years), and late adolescents (grade 11/12, mean age 17.2 years). Such partition is common in developmental research (see for example Choi, Lotto, Lewis, Hoover, & Stelmachowicz, 2008; Moor et al., 2012; van den Bos, Westenberg, van Dijk, & Crone, 2010).

Materials and design. The same Bayesian reasoning problems as in Experiment 1 were used (see supplementary materials). A 2 × 2 × 3 factorial design was employed. The factors Frequency and Absoluteness were manipulated in the same manner as in Experiment 1. Since it was not found

TABLE 2
Correlation between CRT score and performance in Experiment 1

Condition		*Proportion Correct*	*Standard Error*	*Spearman's Rho*	*p-value*
Probabilities	Relative	2%	1%	0.17	0.13
	Absolute	32%	5%	0.23	< 0.05
Frequencies	Relative	41%	5%	0.37	< 0.005
	Absolute	54%	5%	0.48	< 0.001

to affect performance in Experiment 1, the factor Total Sample was no longer manipulated in Experiment 2; total sample size was not presented in any of the question formats. Participants were randomly assigned to one of these four remaining conditions. The third factor in the design was age group (see above).

Procedure. Participants were tested in large groups of 56 to 92 students, and were seated two seats apart so as not to be able to read each other's responses. Booklets with the two reasoning problems were filled in. Participants went sequentially through the booklet and were not allowed to go back to a previous problem.

Results and discussion

Participants who had not answered the two problems were excluded from analysis ($N = 24$). Analyses were conducted on the remaining 190 participants. As in the previous experiment, an answer was deemed correct if it did not deviate from the correct answer more than 1%. The average number of correct answers across the two problems was again used as the measure of reasoning performance and as the dependent variable. The influence of question format was assessed by means of χ^2 tests for independence as in the first experiment. As before, this was to ensure previous findings could be replicated. Non-parametric correlation analyses were performed to look at the relation between age category and performance, our central question.

Overall, participants performed better when information was presented in frequencies (28% vs 47% correct answers); $\chi^2(2) = 12.4$, $p < .01$, and when the set structure was made explicit (22% vs. 52% correct answers, $\chi^2(2) = 29.2$, $p < .001$). These results, illustrated in Table 3 and Figure 3, replicate the main effects for Frequency and Absoluteness observed in Experiment 1. As can be seen from Figure 4, these trends were present in each age category. Across age groups there is an overall increase in correct responses when the information is presented in absolute format compared to relative format. Similarly, there seems to be facilitation in the frequency conditions compared to the probability conditions.

TABLE 3
Proportion of correct answers and number of participants per condition in Experiment 2

Condition		*Proportion correct*	*Standard error*	*N*
Probabilities	Relative	12%	3%	50
	Absolute	45%	5%	47
Frequencies	Relative	42%	3%	42
	Absolute	59%	6%	51

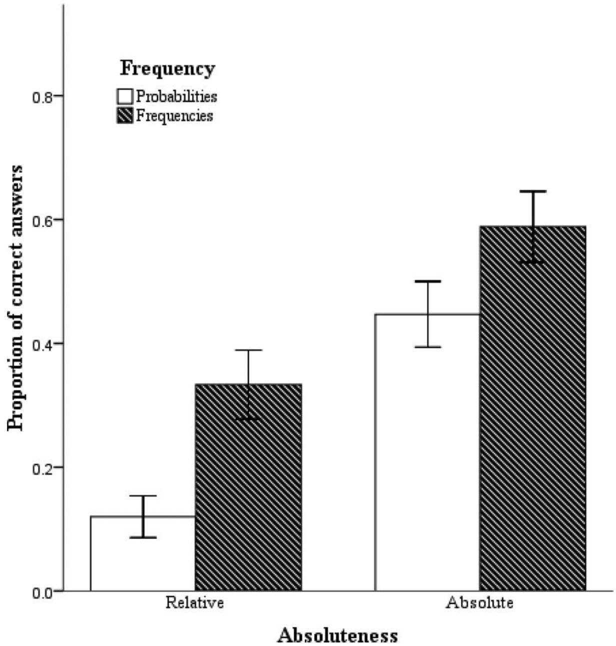


Figure 3. Overall effect of Frequency and Absoluteness on reasoning performance in Experiment 2. Error bars denote $\pm$ 1 standard error of the mean.

The central question in this experiment was whether age is related to performance in Bayesian word problems. An automatically operating frequency module should not be hindered by the more limited cognitive capacity resources in younger age groups, whereas the efficiency of the analytic system should benefit from the increased executive cognitive capacity in the older age groups. Consistent with the predictions of Nested Sets Theory, age category was correlated with performance (Spearman's rho = 0.24, $p < .01$).

More specifically, when this correlation analysis was done separately for problems that presented absolute (non-normalised) and relative numbers, the correlation was much stronger in cases where the set structure of the problem was emphasised (i.e., absolute number condition, Spearman's rho = 0.34, $p < .001$), whereas it did not reach significance when this structure was obscured (i.e., relative numbers, Spearman's rho = 0.14, $p = .174$). This differential pattern was less clear for the frequency factor, however (i.e., frequency condition, Spearman's rho = 0.26, $p < .05$; probability condition, Spearman's rho = 0.24, $p < .05$). Nevertheless the observed format-related trend difference is of particular importance in this experiment because participants in different age groups differ in more

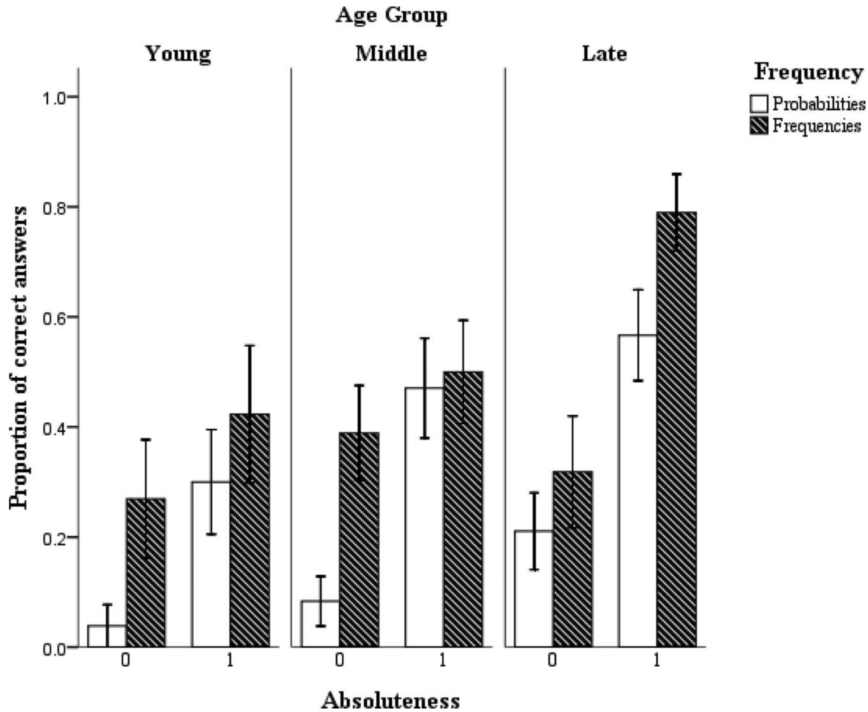


Figure 4. Effects of Frequency and Absoluteness in each age category. Error bars denote ± standard error of the mean.

respects than merely their cognitive capacity. A non-specific relation between performance and age group could easily be ascribed to, for example, experience-related improvements in solving maths problems or general educational confounds. By contrast, such an interpretation could not explain why performance would remain poor in problems that obscure the set structure, and improve with age in problems that clarify set structure.

EXPERIMENT 3

Method

Participants. Participants were 179 first-year psychology students, who took part in the experiment in exchange for course credit.

Materials and design. Participants solved reasoning problems in the natural frequency format, which was shown to facilitate Bayesian reasoning the most in the previous experiments. The availability of executive resources was experimentally manipulated in a dual task paradigm. A secondary task was introduced to reduce available cognitive resources. The factor Load distinguished between the dual task condition and the control condition. In the Load condition, participants were presented with a complex dot pattern (i.e., a ''three-piece'' pattern based on Bethell-Fox & Shepard, 1988, and the work of Verschueren, Schaeken, & d'Ydewalle, 2004; see Figure 5) which they were instructed to keep in mind while solving the reasoning problems. Storage of these complex dot patterns has been shown to efficiently tap executive cognitive resources (e.g., De Neys 2006a; Miyake et al., 2001). No pattern was presented in the No Load condition. Participants were randomly assigned to one of the two conditions. All participants were also presented with the CRT to assess individual differences in executive capacity that might affect the impact of the load task.

Procedure. Participants were tested in groups of 19 to 32. As in the previous experiments two problems of the same format were presented, after which the CRT was completed. Problems were presented on a PC, using E-prime software.

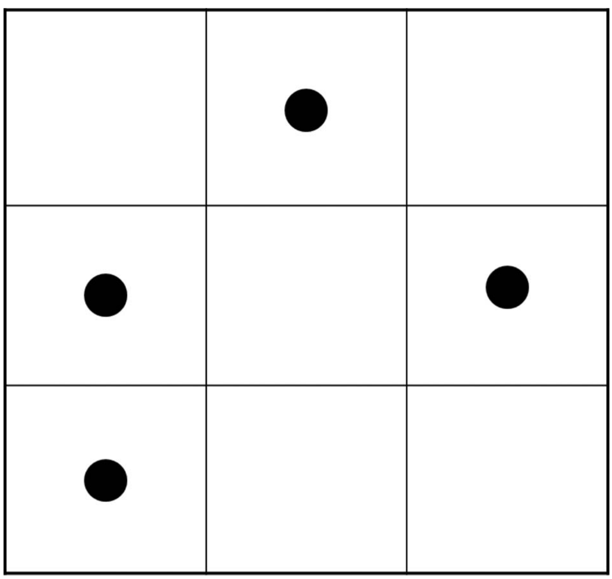


Figure 5. Sample dot pattern. Participants in the Load condition were presented with such a pattern for 850 ms and kept this pattern in mind while solving the subsequently presented reasoning problem.

The same reasoning problems as in the previous experiments were used. First, background information was presented. Participants read this at their own pace and pressed the enter key to move on to the next screen. Second, a dot pattern was presented for 850 ms, after which the reasoning problem was presented. When participants had entered an answer, a grid in which they could reproduce the dot pattern appeared. This was repeated for the second problem. The procedure in the No Load condition was identical, except no dot pattern was presented or reproduced. Participants were asked to follow the instructions on the screen, solve the reasoning problems and, if applicable, reproduce the dot pattern as indicated by the program. They were instructed to solve the three CRT problems when the computer-based part of the experiment was over.

Results and discussion

Data from participants who had not completed both problems or did not fill out the CRT ($N = 8$) were excluded from analysis.

In order for the load manipulation to be effective, the participants must make an effort to maintain the dot pattern in working memory and reproduce it correctly (De Neys, 2006b). To ensure this, a performance criterion was set to the pattern reproduction. On average, 3.35 of the 4 dots in the pattern were correctly reproduced. Participants who performed worse than 2 standard deviations below the mean (fewer than two dots correctly reproduced) were assumed not to be performing the task and were excluded from analysis ($N = 7$). The data of the remaining 164 participants were analysed.

As we used the same secondary task for each participant, the task is likely to have a different impact on participants with different cognitive capacities. The cognitive burden can be expected to be relatively high for low-capacity participants, and relatively low for high-capacity participants. We therefore performed a median split on our measure of cognitive capacity, the CRT score, and looked at the effects of Load in each group separately.

The impact of question format was assessed separately for the Low Capacity (CRT score lower than 2) and the High Capacity (CRT score of 2 or higher) groups by means of χ^2 tests for independence, using the same dependent variable as in the two previous experiments: the mean number of correct answers across the two problems. Participants with a lower cognitive capacity performed significantly worse than participants with a higher capacity (56% vs 83%, $\chi^2(2) = 22.22$, $p < 0.001$), replicating the results of Experiment 1. As Figure 6 shows, the dual task affected performance differently in the two groups. In the Low Capacity group performance was significantly worse in the Load condition than in the No Load condition (44% vs 62%), $\chi^2(2) = 6.37$, $p < .05$, whereas the High Capacity group did

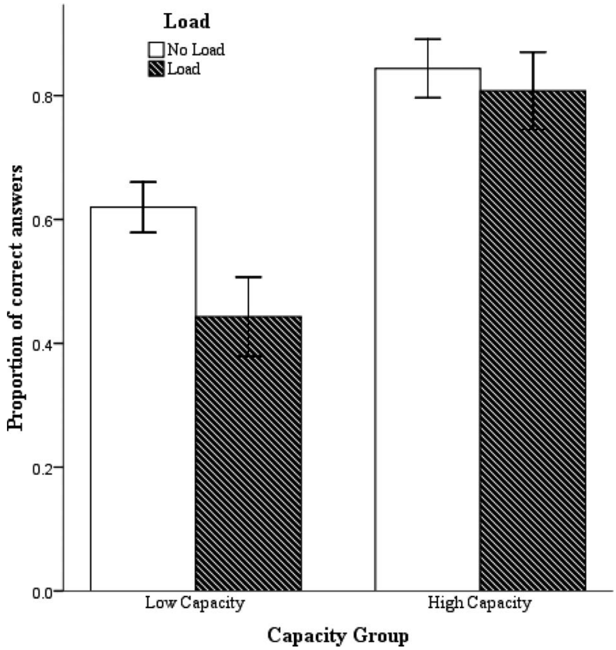


Figure 6. Effects of Capacity and Load on the mean number of correct answers in Experiment 3.

not show this performance decrease (81% vs. 84%), $\chi^2(2) = 0.62$, $p > .50$. Hence, in line with predictions of the Nested Sets framework, findings demonstrate that the processing of natural frequency formats depends on available general cognitive resources. These findings contradict the idea that natural frequencies are automatically processed by an isolated module that functions independently from general-purpose cognitive capacity.

GENERAL DISCUSSION

The tendency for people to ignore the base-rate in Bayesian reasoning problems has been extensively documented (Gigerenzer & Hoffrage, 1995; Tversky & Kahneman, 1982). However, relatively minor adjustments to the traditional question format, expressing the information as natural frequencies, result in a dramatic improvement in reasoning performance (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). Two influential theoretical frameworks have attempted to explain this Bayesian facilitation phenomenon. The first, the Ecological Rationality Framework (Cosmides & Tooby, 1996), argues that humans have evolved a module that automatically processes natural frequencies (Cosmides & Tooby, 2008). This module is said to function independently from general-purpose cognitive processes. The rival theory, Nested Sets Theory (Barbey & Sloman, 2007), posits that a clearer formulation of the information clarifies the nested set structure of the problem. This clear representation will trigger analytic processes, which recruit limited, general-purpose executive resources to calculate the correct answer.

The two explanatory frameworks make opposite predictions regarding the role of general cognitive capacity. The Ecological Rationality Framework predicts performance to be independent from general cognitive resources, specifically when the proposed frequency algorithm receives the correct input, namely natural frequencies. Nested Sets Theory on the other hand, predicts a positive association between cognitive resources and reasoning performance, specifically in conditions that promote Bayesian facilitation, because then the analytical system is recruited (De Neys, 2007).

Here we have tackled this key differential prediction in three experiments. The first experiment looked at the correlation between a measure of general cognitive capacity and performance in different question formats, which each elicit a different degree of Bayesian facilitation. It was found that this correlation was systematically stronger in conditions where performance on the reasoning task was better. The relation between executive resources and performance tended to be stronger for the conditions which facilitated reasoning. This is inconsistent with the natural frequentist claim that an encapsulated module automatically processes frequencies, as the effectiveness of this module would not be related to individual differences in cognitive capacity. If anything, an Ecological Rationality interpretation would be consistent with a stronger performance-capacity correlation in the standard probability format. A very high-capacity minority might be able to calculate the correct answer in the ‘‘unnatural’’, difficult conditions without the automatic frequency algorithm, whereas under natural frequency conditions, when this algorithm receives the correct input, it should work equally well for high-capacity and low-capacity individuals alike. Our results do however conform to the predictions of Nested Sets Theory. In conditions that elicit the best performance overall—the conditions in which the limited-capacity analytic system is recruited—performance is more strongly related to cognitive capacity than in the condition which elicits the worst performance.

The second experiment looked at the cognitive capacity–reasoning relationship from a developmental angle, studying three groups of adolescents in a cross-sectional design. We observed that reasoning performance was positively correlated with age group. Interestingly, this age trend was clearest when the information in the reasoning problem was presented in a format that clarified the set structure. By contrast, when the hit rate and false-alarm rate were normalised, older participants did not reason better than younger participants. This pattern of results fits well with

Nested Sets Theory, as one would expect a stronger age-performance relation under conditions where a clarified set structure triggers the capacity-dependent analytical system. Note that in line with these developmental findings with adolescents, Zhu and Gigerenzer (2006) also found an age-related increase in Bayesian reasoning performance in a sample of elementary school children when problems were presented in a format that facilitates reasoning. This underscores the point that although Bayesian facilitation might be observed at fairly young ages, the extent of this facilitation may show changes over the course of development.

In Experiment 3 available executive capacity was experimentally manipulated in a dual task paradigm. Participants retained a dot configuration in short-term memory while solving the reasoning problems in the natural frequency format, which had been shown to elicit Bayesian facilitation most reliably in the literature (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) and in our previous experiments. According to the Ecological Rationality Framework, an experimental reduction of executive resources would have no effect under natural frequency conditions. When presented with the correct input, the frequency algorithm automatically outputs the correct answer. According to a modular view of the mind, the secondary task would be executed by separate, independent algorithms (Cosmides & Tooby, 1994b). By contrast, Nested Sets Theory predicts that the non-normalised natural frequency format will trigger the analytic system. Since the efficiency of the analytic system is related to individual differences in executive capacity, imposing a cognitive load will leave insufficient resources in relatively low-capacity individuals to complete the reasoning task successfully. Thus Nested Sets Theory predicts a drop in reasoning performance under dual task load, especially for low-capacity individuals. Consistent with the Nested Sets view, results did indeed show that performance in the lower-capacity group suffered from the cognitive burden in the dual-task condition, whereas reasoning performance in higher-capacity participants was not affected.

One might note that the finding that higher-capacity participants did not suffer from the dual task load suggests that Bayesian reasoning with natural frequency formats is quite effortless for these participants. Therefore a strong proponent of the Ecological Rationality Framework could be tempted to argue that at least for high-capacity participants, one can still postulate the existence of an automatically operating natural frequency module. However, it should be clear that the idea that the efficiency of a module depends on the available general-purpose cognitive resources, undermines the very core of the strict modularity concept. Moreover one has to keep in mind that the present study was run with a population of university students. This implies that the present ‘‘lower-capacity’’ participants will still be among the more gifted individuals in the population at large. Hence, even if an Ecological Rationality theorist were willing to make the capacity-dependent-module assumption, it would only explain Bayesian facilitation in a very small, gifted subgroup of the general population.

It is worth noting that the CRT used here to probe cognitive ability is a test which captures intelligence and executive functioning, but also motivational elements and thinking style (Stanovich & West, 2008; Toplak et al., 2011). When discussing cognitive ability in the context of the present experiments, this may then relate to individual differences in these thinking dispositions as well as cognitive ability in the strict sense.

With respect to the theoretical implications of our findings it is important to reiterate that we focused on the strict modular view proposed by Cosmides and Tooby (1996, 2008) that specifies that modules operate automatically. In general, it can be noted that evolutionary psychologists have weakened the strict automaticity claim. That is, is has been argued that not all modules necessarily operate automatically, and that the efficacy of some modules can depend on general cognitive resources (Barrett et al., 2006; Barrett & Kurzban, 2006). Clearly, under such weakened modularity claims, the present findings do not directly argue against a frequency module per se. However, we do believe that such a weak modularity view loses much of its a priori appeal. In our view the strength of the module idea is that it facilitates performance by freeing up central processing resources. The present findings indicate that, in sharp contrast with this basic idea, the frequency format actually cues people to start allocating *more* central processing resources to the problem at hand. Although it might not be completely impossible to integrate this central processing recruitment with a weak modular account, the findings are directly predicted by the Nested Set account.

For completeness, we would like to stress that we have no issue with the idea of strict modularity per se. Indeed, note that by adopting a related experimental approach our group found clear evidence for Cosmides and Tooby's postulation of an automatically operating module in the context of AQ4 social exchange reasoning (Van Lier, Revlin, De Neys, 2012; see also AQ5 Bonnefon, Hopfensitz, & De Neys, 2012, for evidence with respect to the strict modular nature of trustworthiness detection). Our point is therefore not that modules cannot operate automatically but rather that the characteristics of postulated modular process need to be verified experimentally. We believe it makes perfect sense for some modular processes to operate completely automatically. Our point here is that frequency computations are not among these.

The present results also have practical implications for a wide range of real-life contexts. Every day, life-and-death decisions are made on the basis of conditional probabilities, in context as diverse as the court room, the doctor's office, and the war room. With this in mind, the goal is to determine how information can be presented in such a way that people are capable of making truly informed decisions. Efforts have been made to apply findings from the Bayesian reasoning literature. Presenting natural frequencies instead of single-event probabilities has been shown to improve reasoning in a medical decision making context, for example (e.g., Carling et al., 2009; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). From our data, however, it appears not everybody will benefit equally from a reformulation that facilitates reasoning. Specifically, groups with a lower cognitive capacity may not benefit as much from facilitating manipulations. Considering the important real-life situations in which base-rate neglect plays a role, this is not a trivial issue. Other studies have provided indirect evidence that manipulations aimed at facilitating Bayesian reasoning are more effective in groups with larger average executive capacity. Chapman and Liu (2009) found that natural frequency manipulations are more beneficial for people that are high in numeracy, a construct that partly, though not entirely, overlaps with general cognitive ability. These findings lend credence to the present observations and further question the claim that natural frequencies are processed automatically. From an applied point of view, one might note that rather than assuming an optimal effect when natural frequencies are used, based on near-perfect performance in a population of high-functioning young adults (Cosmides & Tooby, 1996), additional facilitating measures, such as the use of diagrams (Yamagishi, 2003) can be considered to help decision-making in real-life contexts in the general population (Garcia-Retamero & Galesic, 2010).

In sum, in the present paper we put to the test a crucial property of theories of cognition as proposed by Cosmides and Tooby (1996, 2008), namely the independence of specialised frequency algorithms from general-purpose cognitive processes. In each of our experiments, available general-purpose cognitive resources were consistently found to be related to performance in Bayesian reasoning problems. Taken together, the findings lend credence to Nested Sets Theory and contradict the popular claim that natural frequencies are automatically processed by an isolated module that functions independently from general-purpose cognitive capacity.

Manuscript received 24 January 2012
Revised manuscript received 4 July 2012
First published online day/month/year

REFERENCES

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*, 241–254.

Barrett, H. C., Frederick, D. A., Haselton, M. G., & Kurzban, R. (2006). Can manipulations of cognitive load be used to test evolutionary hypotheses. *Journal of Personality and Social Psychology*, *91*, 513–518.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological review*, *113*(3), 628–647.

Barr-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*, 211–233.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*: *14*, 12–23.

Bonnefon, J. F., Hofensitz, A., & De Neys, W. (2012). *Evidence for the modularity of trustworthiness detection in an economic game*. Manuscript submitted for publication.

AQ6

Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuals, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, *127*, 3–21.

Borgida, E., & Brekke, N. (1981). The base-rate fallacy in attribution and prediction. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 3). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Carling, C. L. L., Kristoffersen, D. T., Montori, V. M., Herrin, J., Schunemann, H. J., Treweek, S., et al. . (2009). The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: A randomized trial. *PLOS Medicine*, *6*(8), 10.1371.

Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, *299*, 999–1000.

Chapman, G. B., & Liu J. J. (2009). Numeracy, frequency, and Bayesian reasoning. *Judgment and Decision Making*, *4*(1), 34–40.

Choi, S., Lotto, A., Lewis, D., Hoover, B., & Stelmachowicz, P. (2008). Attentional modulation of word recognition by children in a dual-task paradigm. *Journal of Speech, Language and Hearing Research*, *51*(4), 1042–1054.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P.C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence*, *32*(3), 277–296.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*(3), 115–120.

Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Adaptive specialisations, social exchange and the evolution of human intelligence. *Proceedings of the National Academy of Sciences*, *107*(2), 9007–9014.

Cosmides, L., & Tooby, J. (1994a). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, *50*(1–3), 41–77.

Cosmides, L., & Tooby, J. (1994b). Better than rational: Evolutionary psychology and the invisible hand. *The American Economic Review*, *84*(2), 327–332.

Cosmides, L., & Tooby, J. (1996). Are humans goods intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Cosmides, L., & Tooby, J. (2008). Can a general deontic logic capture the facts of human moral reasoning? How the mind interprets social exchange rules and detects cheaters. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 53–119) Cambridge, MA: MIT Press.

De Neys, W. (2006a). Dual processing in reasoning; two systems but one reasoner. *Psychological Science*, *17*(5), 428–433.

De Neys, W. (2006b). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-tasks considerations. *The Quarterly Journal of Experimental Psychology 59*(6), 1070–1100.

De Neys, W. (2007). Nested sets and base-rate neglect: two types of reasoning? *Behavioral and Brain Sciences*, *30*, 260–261.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*, 128–133.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*, 349–381.

De Neys, W., & Verschueren, N. (2006). Working memory capacity and a notorious brain teaser: The case of the Monty Hall Dilemma. *Experimental Psychology*, *53*, 123–131.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge, UK: Cambridge University Press.

Evans, J. St. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *10*(7), 454–459.

Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, *13*(3), 378–395.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, *59*, 255–278.

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.

Fenton, N., & Neil, M. (2010). Comparing risks of alternative medical diagnosis using Bayesian arguments. *Journal of Biomedical Informatics*, *43*, 485–495.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, *50*, 123–129.

Fiedler, K., Brinkmann, B., Betsch, T., & Wild, B. (2000). A sampling approach to biases in conditional probability judgments: Beyond base rate neglect and statistical format. *Journal of Experimental Psychology: General*, *129*(3), 399-418.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT press.

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

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science and Medicine*, *70*, 1019–1025.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *11*(3), 410–419.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–162). Chichester, UK: Wiley.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Academy of London*, *205*(1161), 581–598.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Medicine: Communicating statistical information. *Science*, *290*(5500), 2261–2262.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251.

Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*(3), 283–291.

Kassin, S. M. (1979). Base rates and prediction: The role of sample size. *Personality and Social Psychology Bulletin*, *5*, 210–213.

Kleiter, G. D. (1994). Natural sampling – Rationality without base rates. In G. H. J. Fischer & N. Lambell (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 375–388). New York: Springer.

Laplace, P. S. (1902). *A philosophical essay on probabilities*. New York: Drummond.

Lloyd, A. J. (2001). Communicating and understanding risk: The extent of patients' understanding of the risk of treatments. *Quality in Health Care*, *10*(Supplement1), i14–i18.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, *40*, 287–298.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficacy of psychometric signs, patterns or cutting scores. *Psychological Bulletin*, *52*, 194–216.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent variable analysis. Journal of Experimental *Psychology*: General, *130*, 621–640.

Moor, B. G., Güroğlu, B., Op de Macks, Z. A., Rombouts, S. A., Van der Molen, M. W., & Crone E. A. (2012). Social exclusion and punishment of excluders: neural correlates and developmental trajectories. *Neuroimage*, *59*(1), 708–717.

Neace, W. P., Michaud, S., Bolling, L., Deer, K., & Zecevic, L. (2008). Frequency formats, probability formats, or problem structure? A test of the nested-sets hypothesis in an extensional task. *Judgment and Decision Making*, *3*(2), 140–152.

Over, D. E. (2007). The logic of natural sampling. *Behavioral and Brain Sciences*, *30*(3), 277.

Sloman, S., & Over, D. E. (2003). Probability judgment from the inside out. In D. E. Over (Ed.), *Evolution and the psychology of thinking: The debate*. Hove, UK: Psychology Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organisational Behavior and Human Decision Processes*, *91*, 296–309.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory and Cognition*, *39*, 1275–1289.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *4157*, *185,* 1124–1131.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge, UK: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

van den Bos, W., Westenberg, M., van Dijk, E., & Crone, E.A. (2010). Development of trust and reciprocity in adolescence. *Cognitive Development*, *25*, 90–102.

Van Lier, J., Revlin, R., & De Neys, W. (2012). *Detecting cheaters without thinking: Testing the*
AQ7 *automaticity of the cheater detection module*. Manuscript submitted for publication.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2004). Everyday conditional reasoning with working memory preload. In *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Yamagishi, K. (2003). Facilitating normative judgments of conditional probability: Frequency or nested sets? *Experimental Psychology*, *50*, 97–106.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*, 287–308.