

The (In)flexibility of Evolved Frequency Representations for Statistical Reasoning: Cognitive Styles and Brief Prompts Do Not Influence Bayesian Inference

Gary L. Brase

University of Missouri-Columbia

What happens when format manipulations improve Bayesian reasoning? One view is that naturally sampled frequencies help induce a privileged representational system that is relatively specific in its operation. A contrasting view is that naturally sampled frequencies are but one way to induce a more general process of appreciating nested set relationships. This later view implies that fairly brief and immediate interventions (e.g., simple directives) should produce improvement, whereas the former view implies that more extensive interventions and/or more insightful understanding are necessary for improvement. The present research indicates that neither brief and immediate interventions nor pre-existing representational biases or representational flexibility facilitate performance. Some evidence emerged, on the other hand, that frequentist problem interpretation can improve statistical reasoning performance and increase confidence in responses at times. These results support the privileged representational system view.

Keywords: judgment under uncertainty, statistical probability, frequencies, inference, reasoning.

进化频率表征对统计推理（非）灵活性的影响：认知风格和简要提示不影响贝叶斯推理

当形式操纵有助于贝叶斯推理时会有怎样的加工发生呢？一种观点认为自然取样的频率可以激发在其操作中具有相对特异性的特权表征系统。而与之相对的一种观点则认为，自然取样频率只是引发具有嵌套关系的更为普遍的加工的一种方式。比较两种观点，后者预示着只需要使用相当简要和直接的干预（如简单的指示）就能够促进推理的改善，而前者则意味着更为广泛的干预和/或更有洞见的理解才能改善推理。本研究表明，无论是短暂立即的干预，还是预存的表征偏向，抑或是表征的灵活性都不能促进被试的表现。另一方面，也有证据显示，频率论者的问题解释可以改善统计推理表现，而且有时还会增加其反应的信心。这些结果支持了特权表征系统观。

关键词：不确定性判断，统计概率，频率，推测，推理。

分类号： B84-069

Early research on human statistical reasoning was dominated by findings of incompetence. Specifically, research conducted over many decades has revealed that people are generally very bad at Bayesian inference tasks such as the following:

There is a newly discovered disease, Disease X, which has a prevalence rate of 8%. There is a recently developed test to detect whether or not a person has this infection, but this test has a false positive rate of 1/8. Suppose now a patient takes this test. What is the probability this patient actually has the disease, given a positive test result?

Most people, including even doctors and medical students, proved to be largely unable to combine the base rate (8%) with the false positive rate (1/8) and thereby provide a correct posterior probability for a problem such as this (e.g., Casscells, Schoenberger, and Grayboys, 1978); i.e., 8 of the 20 (40%) people

testing positive have the disease. Summaries of these findings included general pronouncements that “man is apparently not a conservative Bayesian: he is not a Bayesian at all.” (Kahneman & Tversky, 1972, p.450). Over the last decade, though, this conclusion has been re-evaluated.

A number of more recent studies have found that performance improves substantially when the task information is given as frequency counts of non-normalized, naturally sampled objects, events, or locations (hereafter, natural frequencies; e.g., Gigerenzer & Hoffrage, 1995). The above problem, for example, becomes:

There is a newly discovered disease, Disease X, which occurs in 8 out of 100 people. There is a recently developed test to detect whether or not a person has this infection, but this test is not perfect. Specifically, 12 out of every 92 perfectly healthy people also get a positive test result. Suppose new population of 100 people are tested for this disease. How many will *actually* have the disease, out of those that test positive? _____ out of _____.

Received 2006-08-30

Correspondence should be addressed to Gary L. Brase, Department of Psychological Sciences, University of Missouri-Columbia, 210 McAlester Hall, Columbia, MO 65211, USA; e-mail: braseg@missouri.edu.

There is no doubt now that such format conversions are capable of improving performance, but a lively debate has continued to today about *why* this is so. The first thing to notice is that the use of natural frequencies, because they include non-normalized subsets, makes Bayesian inference a computationally simpler task. Rather than using the traditional Bayes Theorem to combine single-event probabilities and base-rates, the posterior probability is attainable using whole number addition and division (Gigerenzer & Hoffrage, 1995). Some researchers end their explanation of this improved performance here, claiming that this format induces a perspective that makes the “nested-set relationship cognitively transparent for problem solving” (Sloman, Over, Slovak, and Stibel, 2003, p. 298). By this account, natural frequencies are but one of many ways to facilitate performance:

“The nested-sets hypothesis is the general claim that making nested-set relations transparent will increase the coherence of probability judgment. Making set relations transparent is not the same as inducing a frequency frame, though frequency frames can clarify set relations... Nested-set relations are more general than frequency representations.” (Sloman, et al., 2003, p. 307)

Others, including the original developers of the use of natural frequencies in Bayesian inference tasks, see a greater significance in the natural frequency format itself. By this account, natural frequencies not only simplify the computational demands of the judgment task, but also promotes a representation of the information that meshes with the evolved capability of the human mind – that is, they are a cognitively privileged representational format. This view proposes that the mind is specifically designed to work with frequencies (counts of objects, events, and locations) that are sequentially encountered and categorized as they are received from the environment:

“Evolutionary theory asserts that the design of the mind and its environment evolve in tandem. Assume... 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 humans evolved, the ‘natural’ format was *frequencies* as actually experienced in a series of events, rather than probabilities or percentages. (Gigerenzer and Hoffrage, 1995, p. 686).

Although not excluding other, more general, processes or the existence of subjective probabilities in mental representations, natural frequencies are thus held as particularly potent representations.

Adjudicating between these perspectives has been difficult. The process of making “nested-set relations

transparent” invariably entails creating a non-normalized set structure that is organized exactly as natural frequencies would be organized. In other words, the *natural sampling* part of natural frequencies is extracted out and elevated to be the sole cause of improved performance. The difficulty with this situation is that it is actually rather difficult to describe such nested-set relations without reference to very frequency-like information. For instance, Girotto and Gonzales (2001) (as well as Sloman et al., 2003, and Johnson-Laird, Legrenzi, Girotto, Legrenzi and Caverni, 1999) have claimed that a particular form of probabilities expressed as “chances” can facilitate performance as much as frequencies. Thus, our example problem using this format becomes:

There is a newly discovered disease, Disease *X*. A person has 8 out of 100 chances of having this disease. There is a recently developed test to detect whether or not a person has Disease *X*, but this test is not perfect. Specifically, there are 12 out of 92 chances that a perfectly healthy person will get a positive test result. Suppose person is tested for this disease. Out of 100 chances, how many chances are there that this person will *actually* have the disease, out of those chances that she/he will test positive? _____ out of _____.

The claim that the information in this problem is in the form of probabilities, and not frequencies, is central to the nested-set hypothesis claim that frequencies are superfluous. The validity of this claim, however, depends crucially on peoples’ interpretation of the fundamentally ambiguous “chances” information (independent of if it was labeled as “chances” by experimenters; Brase 2002b; Hoffrage, Gigerenzer, Krauss, and Martignon, 2002) and there is evidence that a frequency interpretation of such chances is a key to good performance (Brase, under review).

Privileged Frequencies Versus Equipotential Nested-sets

One way to potentially distinguish between these two perspectives is to focus on the disputed claim that the human mind evolved to hold frequency information as privileged. This claim is based on the idea that natural selection would have built cognitive mechanisms designed to work with the naturally occurring information available over evolutionary history (i.e., frequencies of objects, events, or locations). Like the old computer science law of GIGO (Garbage In; Garbage Out), natural frequencies most effectively tap into these cognitive mechanisms. Other numerical information formats, although comprehensible to educated modern humans, would less effectively tap into these cognitive mechanisms. Furthermore, because the evolutionary process is not teleological, processes of converting other numerical

information into natural frequencies should not be part of the normally developing cognitive abilities, but rather should require conscious, effortful, and sustained efforts to attain (an analogy can be drawn here with the acquisition of reading and writing, which is effortfully attained as a supplement to spoken language abilities in humans*).

An implication of this perspective is that, given non-frequency information (even when in non-normalized subsets), it should take fairly explicit and concerted effort to train individuals to convert such information to natural frequencies and thereby make it more amenable to understanding and analysis. For example, Sedlmeier (1997, 1999; Sedlmeier & Gigerenzer, 2001) trained people to use natural frequency representations in order to facilitate performance (and found that such training is retained better than other training regimens), but such training typically involved one to two hours of explicit demonstrations and instructions on how to actively and insightfully convert other numerical formats into natural frequencies (also see Ruscio, 2003 and Kurzenhäuser and Hoffrage, 2003 for attempts to develop effective training schemes that last a bit less than one hour).

In contrast, the nested-sets hypothesis, which views natural frequencies as only a specific instantiation of a more general capability to appreciate nested set relations, predicts that performance can be improved with much less effort than expended by Sedlmeier and others. Specifically, the nested-sets hypothesis holds that different forms of cognitive representation are completely or nearly equipotential (i.e., representations can be generated and held with about the same level of effort):

“People normally represent a category, e.g., patients who have been tested for a disease, from a perspective that gives access to the category’s internal structure... The frequency frame induces a different perspective by asking participants to think about multiple instances of the category and so about the set or class corresponding to the category.” (Sloman, et al., 2003, p. 298)

If the key to better performance is inducing a representation in terms of instances, then a more fundamental understanding and interpretation of the problem information is not necessary. Simple and explicit indications to take a frequency perspective in viewing a Bayesian reasoning problem should produce immediate facilitation (so long as the problem content is also interpretable as frequencies).

For example, Sloman, et al. (2003) used a manipulation that “increased transparency [of the nested-sets relation] by providing an Euler circle diagram that makes the nested-set relations explicit” (p. 302).

A simple contrast between the frequency hypothesis and the nested-sets hypothesis, then, is to employ a very simple, clear, and brief injunction to represent the task in terms of frequencies and/or instances. This will increase the transparency of any nested-set relations, and so the nested-sets hypothesis predicts improved performance (As a corollary, it should also be possible to depress performance by doing the opposite: providing a similar injunction to represent the task in terms of single-event probabilities and/or a prototypical instance). On the other hand, the frequency hypothesis predicts that this brief injunction will be insufficient to produce the type of improvement that is found with either clear natural frequencies or effortful and sustained training on how to convert other information forms into natural frequencies. In summary, the experiments presented here are designed to assess two contrasting hypotheses: The nested-sets hypothesis predicts that brief instructions (and in Experiments 2a and 2b, individual differences) will significantly influence Bayesian reasoning performance, whereas the frequency hypothesis predicts that these manipulations will not have significant effects on Bayesian reasoning performance.

Experiment 1a

This experiment used a Bayesian reasoning task, presented in a natural sampling (i.e., non-normalized, nested-set relationship) structure, but utilizing the ambiguous “chances” format that is amenable to interpretation as either frequencies or single-event probabilities. Participants received an injunction, however, to adopt one of these representations.

Method

Participants. The participants were 62 undergraduates from a large, Midwestern university. All participants completed the study as partial fulfillment of an introductory psychology class requirement. The average age of participants was 18.4 years.

Materials and Procedure. After reading and signing a separate informed consent form, each participant was given a single sheet that included general task instructions, instructions about how to think about the task information, and a Bayesian reasoning task. The Bayesian reasoning task was a version of the “medical diagnosis problem,” using the numerical format of “chances” that are in a natural sampling framework: There is a newly discovered disease, Disease X, which is transmitted by a bacterial

* This like most analogies, can be both helpful and misleading. The helpful part is in understanding that there is an intuitive and reliably developing ability to acquire and use quantitative information in the human brain. The potentially misleading aspect can come from taking the analogy too far; I do not mean to imply that explicit counting systems are “innate,” nor that large quantities are explicitly represented as specific numbers.

infection. Here is some information about the current research on Disease *X* and efforts to test for the infection that causes it.

A person has 8 chances out of 100 of having the infection. There is a recently developed test to detect whether or not a person has this infection, and every chance of having the infection was associated with a positive reaction from the test. The test is not perfect, however. 12 of the remaining 92 chances of not having the infection (that is, being perfectly healthy) were also associated with a positive reaction from the test. Imagine Michael is tested now. Out of a total of 100 chances, Michael has _____ chance(s) of positive reaction from the test, _____ of which will be associated with actually having the infection.

Previous work with this format (Brase, under review) has shown that this format is particularly ambiguous with regard to being interpreted by participants as frequencies or single-event probabilities. Preceding this problem was one of two instructions:

- [Frequency version] Some people find it useful to evaluate problems such as these (that talk about “chances”) by thinking about them in terms of frequencies. That is, thinking about the statements as referring to discrete, countable items. For this problem, think about the numbers as frequencies of items. For example, “5 chances out of 10” would be interpreted as “half of the last ten events turned out a certain way”
- [Probability version] Some people find it useful to evaluate problems such as these (that talk about “chances”) by thinking about them in terms of probabilities. That is, thinking about the statements as referring to a single item in some state of uncertainty. For this problem, think about the numbers as probabilities. For example, “5 chances out of 10” would be interpreted as “this event is as likely to occur as it is to not occur”

Results and Discussion

Participants showed no difference in performance between the two instruction conditions. (Probability instruction, $n = 30$: 63.3% vs. Frequency instruction, $n = 32$: 65.6%: $z = 0.19$, $p = .425$, $\eta = .05$; using a difference of proportions test). This result is somewhat difficult to reconcile with the idea that the facilitation of performance is a matter of a relatively simple shift in representation. Instead, facilitation seems to be based on a conceptually deeper and more irrevocable process.

It could be, however, that these results are misleading. Perhaps the participants did not attend to the instructions, or even forgot the instructions once they began working on the problem. It could even be

possible that participants, realizing the different possible interpretations after reading the instructions, disproportionately opted for a frequency representation (or a probabilistic representation in terms of instances of the category) regardless of which interpretation the instructions implored them to take. This would result in very good performance across both conditions. Experiment 1a was therefore repeated, with the addition of two manipulation checks regarding the effectiveness of the numerical representation instructions.

Experiment 1b

This experiment used the same materials as Experiment 1a, but added manipulation check measures after the completion of the Bayesian reasoning task.

Method

Participants. The participants were 61 undergraduates from the same population as in Experiment 1a. The average age of participants was 18.3 years.

Materials and Procedures. After reading and signing a separate informed consent form, each participant was given a single sheet that included the same general instructions, frequency or probability instructions manipulation, and Bayesian reasoning task as in Experiment 1a. After completing the reasoning task, however, participants were instructed to turn over the paper and answer the questions on the other side. The post-test questionnaire asked participants to honestly indicate how much attention they paid to the instructions and how well they kept those instructions in mind during the task. The two items were:

- 1) Without looking back, recall as best you can what the task instructions for the previous task were:
 - a. to think about the numbers as frequencies.
 - b. to think about the numbers as probabilities
- 2) Rate how much you agree with the following statement: I actually was able to, and did, think about the numbers in the task in the way that the instructions said to (followed by 1-6 scale ranging from “totally untrue” to “totally true”)

Results and Discussion

5 of the 61 participants gave incorrect answers on the first manipulation check, and were therefore excluded from further analysis. The remaining participants again showed no difference in performance between the two instruction conditions. (Probability instruction, $n=29$: 51.7% vs. Frequency instruction, $n = 27$: 48.1%: $z = 0.27$, $p = .395$, $\eta = .07$). Because nearly all of the participants demonstrated an awareness of the initial instruction contents, it does not appear that the lack of a difference in performance is due to not reading the

instructions. Perhaps, though, the participants read but did not follow the instructions. The difference in ratings of how well the directions were followed was not significant (Probability instruction, mean = 4.59 vs. Frequency instruction, mean=3.89; $t(54) = 1.94$, $p = .058$). Perhaps a better test, however, is to look specifically at the participants providing the highest ratings on this measure, indicating that it was “somewhat true” or “totally true” that they thought about the numbers as the instructions indicated. Looking at just these participants, performance uniformly shifts up but remains equal across conditions (Probability instruction, $n=16$: 68.8% vs. Frequency instruction, $n = 9$: 66.7%: $z = 0.11$, $p = .458$, $\eta^2 = .04$).

Experiment 2a

An incidental finding from Experiment 1b was that some participants gave relatively low ratings – including disagreement – to the supposition that they had thought about the numbers as the instructions indicated. This result raises the issue of individual differences in representational preferences; biases towards either frequency or probability representation (or towards representation in terms of category instances or category properties). Thus, it is possible that participants could have biases towards, for example, frequency interpretation (or category instances interpretation) *before* encountering this particular Bayesian reasoning problem.

The nested-sets hypothesis, which entails fairly equipotential abilities to represent situations in terms of category instances or in terms of category properties, would fully anticipate both that such individual differences in representational style could exist and that they would influence performance on these tasks. Some participants could be dispositionally biased towards using frequency/instance interpretations in tasks, and those same people could also tend to be better at Bayesian tasks (because they more naturally see the subset relations). If this is the case, then natural frequency presentations (even in the “chances” format) will tend to be associated with better Bayesian reasoning, but it is because this format manages to tap into latent individual differences whereas other formats fail to do the same.

On the other hand, if frequencies are a privileged representational format as the frequency hypothesis holds, then individual differences should have minimal effects on this evolved adaptation. Participants who represent the *task itself* as frequencies should be both more accurate and more confident in their answers, but — somewhat counterintuitively — this effect should occur regardless of individual variations in general representation biases (see Brase, 2002a on the

perception of frequencies as clearer and easier to understand).

Method

Participants. The participants were 105 undergraduates from the same population as Experiments 1a and 1b, with an average age of 19.2 years.

Materials and Procedure. After reading and signing a separate informed consent form, each participant was given a questionnaire that included some brief instructions, a pretest of frequency versus probability interpretation bias, a Bayesian reasoning problem, and follow-up questions on the reverse side of the sheet from the reasoning problem. The pretest consisted of seven numerical statements, representing a diverse range of presentations: a. If the parent has a mutated gene, a child has a 50 percent chance of inheriting it. b. Days such as this have a 30 percent chance of rain. c. The basketball team has 6 out of 10 chances of scoring from the 3-point line. d. 4 out of 5 pro boxers end up with long-term mental impairment. e. 38 year old females have a .10 probability of a positive mammogram test. f. The soccer team has a .43 probability of earning a Champions League spot. g. Prospective college students have 3 out of 4 chances of being admitted.

Each statement was followed by three options for categorizing that statement: Frequency (defined as statistical statements that refer to discrete, countable items), Probability (defined as statistical statements that refer to a single item in some state of uncertainty), and Other. The general instructions and Bayesian reasoning task were the same as in Experiment 1. A question following the Bayesian task (on the other side of the paper) asked: “Which of the following best describes how you thought about the information and reached your answer to this problem?” and provided the options of:

- I thought about the situation in terms of frequencies (discrete, countable events)
- I thought about the situation in terms of probabilities (a single event with an uncertain outcome).
- Other: I thought about the situation as _____

A final question asked participants to rate their “confidence that the answer you provided is in fact the correct answer” on a scale from 1 to 6 (1 being not at all confident and 6 being totally confident).

Results

Is a pre-existing frequency interpretation bias associated with performance?

Responses on the pretest were coded as “1” for frequency and “2” for probability interpretation (“other” responses were deleted). Individual averages were calculated, and the grand mean of these was 1.57. 26% of participants could be considered as showing a

bias towards frequency representation (26/101 participant with scores at or below 1.5). These participants were not significantly better at obtaining the correct Bayesian inference response than those with a bias towards probability interpretation (58.7% vs. 61.5%: $z = 0.25$, $p = .400$, $\eta = .06$). One could argue that the group considered biased towards probability interpretation was conflated with those who should be considered neutral, so the participants were also divided into three groups: Frequentists (scores 1–1.49), Probabilists (scores 1.7–2), and neutrals (scores 1.5–1.69). The results showed that all three groups performed similarly (between 58%–61% correct responses).

Is interpretation on this task associated with performance?

The naturally sampled chances phrasing induced a fairly even split among participants in terms of how it was interpreted. 42.6% interpreted the task information as frequencies, and 47.5% interpreted it as probabilities (the remaining 9.9% of participants selected “other”). This difference in interpretation had a significant effect on task performance: 72.1% of those who used a frequency representation obtained the correct answer, compared to only 50.0% of those who used a probability representation ($z = 2.15$, $p = 0.02$, $\eta = .48$). Interestingly, there was only a weak correlation between scores on the pretest and reported interpretation of the task (excluding those who chose “other” as an interpretation: $r(91) = 0.18$, $p = .04$).

Is interpretation on this task associated with confidence?

The ratings of participants’ confidence in their responses were categorized according to their actual performance (correct/incorrect) and chosen task representation (frequency/probability) and a 2×2 ANOVA was performed. Unsurprisingly, there was a significant effect of having the correct answer ($F(1,87) = 4.544$, $p = 0.036$, $\text{partial } \eta^2 = 0.050$), but there was also an effect of the type of interpretation ($F(1,87) = 12.359$, $p = 0.001$, $\text{partial } \eta^2 = 0.124$). There was no interaction, so regardless of *actual* performance, people who represented the task in terms of frequencies were more confident in their responses.

Table 1 Means and standard deviations (in parentheses) of confidence ratings for Experiment 2a, sorted by frequency or probability interpretations of the task and by correct or incorrect answers.

	Correct Answer	Incorrect Answer	
Frequency Interpretation	4.24 (1.32)	3.71 (1.39)	4.09 (1.35)
Probability Interpretation	3.29 (1.56)	2.54 (1.14)	2.92 (1.40)
	3.83 (1.49)	2.93 (1.33)	

Experiment 2b

Although the results of Experiment 2a seem to support the frequency hypothesis predictions, there is the same concern as with Experiment 1a with the possibility that a lack of significant differences could have been due to an ineffective manipulation. Of particular concern is that one could argue the relevant individual difference is not bias toward one or another type of representation, but rather differences in ability to flexibly and adaptively shift across different representational frameworks. Thus, the nested-sets hypothesis would predict that people who are particularly adept at shifting from one framework (e.g., thinking in terms of category properties) to another (e.g., thinking in terms of category instances) will be at a distinct advantage when it comes to Bayesian reasoning tasks. With this in mind, Experiment 2a was repeated, but this time with a different measure of cognitive representation style.

Method

Participants, Materials, and Procedures. The participants were 94 undergraduates from the same population as Experiment 2a, with an average age of 18.4 years. After reading and signing a separate informed consent form, each participant was given a questionnaire that included some brief instructions, a Bayesian reasoning problem, follow-up questions on the reverse side of the sheet from the reasoning problem (same as Experiment 2a), and then The Actively Open-Minded (AOM) Thinking Scale (Sá, Kelley, Ho, & Stanovich, 2005; Sá, West, & Stanovich, 1999; Stanovich & West, 1997, in press). The AOM Thinking Scale has been developed and refined over several years, and the current version consists of 41 statements rated on 6-point Likert scales (e.g., “People should always take into consideration evidence that goes against their beliefs”). This scale measures openness to belief change and cognitive flexibility; exactly what is of interest here.

Results

The naturally sampled chances phrasing again split participants in terms of how it was interpreted. 32.7% interpreted the task information as frequencies, and 58.4% interpreted it as probabilities (the remaining 8.8% of participants selected “other”). This difference in interpretation had an effect similar to that found in Experiment 2a (i.e., frequency representation aided performance), although here it was not statistically significant (64.9% correct using a frequency representation, compared to 53.0% using a probability representation; $z = 1.171$, $p = 0.122$, $\eta = .24$). High scorers on the AOM Thinking scale did not do significantly better than other participants (using a median split, 60.7% of High AOM obtained the

correct answer, compared to 53.6% of Low AOM participants; $z = 0.759$, $p = 0.225$, $\eta^2 = .14$. Similar results are obtained when comparing just the top and bottom quartiles; 69% versus 63%).

The ratings of participants' confidence in their responses were categorized and analyzed as in Experiment 2a. Once again there was a significant effect of having the correct answer ($F(1,99) = 30.30$, $p < 0.001$, partial $\eta^2 = 0.23$), but instead of the main effect of type of interpretation ($F(1,99) = 1.50$, $p = 0.22$, partial $\eta^2 = 0.015$) there was instead a weak interaction ($F(1,99) = 4.25$, $p = 0.04$, partial $\eta^2 = 0.04$). The means in Table 2 reveal that there was a slight boost in confidence for a correct answer within a frequency interpretation, but not a similar boost for an incorrect answer.

Table 2 Means and standard deviations (in parentheses) of confidence ratings for Experiment 2b, sorted by frequency or probability interpretations of the task and by correct or incorrect answers.

	Correct Answer	Incorrect Answer	
Frequency Interpretation	4.67 (1.37)	2.54 (1.33)	3.92 (1.69)
Probability Interpretation	3.74 (1.42)	2.77 (1.18)	3.29 (1.39)
	4.12 (1.46)	2.70 (1.21)	

General Discussion

For both practical and theoretical reasons, it is important to establish the boundary conditions for manipulations that improve statistical reasoning. For practical purposes, it does not appear that people can be pre-selected for a "frequency perception bias" in order to improve statistical reasoning capability (e.g., in employee selection), nor does it appear that immediate, concise, and explicit injunctions to think about task information as frequencies is sufficient to improve performance.

The lack of differential facilitation from a simple injunction to think in terms of frequencies of instances, as well as the lack of inhibition from a similar injunction to think in terms of probabilities of a single case (Experiments 1a and 1b), are failures for the idea that the underlying cognitive process is a simple shift in representation between general "inside views" and "outside views" (i.e., representations in terms of relations among category features/properties versus in terms of sets or classes corresponding to a category; Sloman, et al., 2003). Experiments 2a and 2b established that neither a general bias towards representing quantitative statements as frequencies nor a greater cognitive flexibility bear any systematic relationship to reasoning task performance, again arguing against a very general form of cognitive representational distinctions. Interestingly, Stanovich and West (1998) found that "discrepancies between descriptive and normative models on noncausal base-

rate problems are not as easily explained by recourse to cognitive capacity." (p. 174) That is, unlike a range of other positive relations between reasoning task performance and cognitive ability, Bayesian reasoning task performance—as in this study—was not positively related to general cognitive abilities.

Experiments 2a and 2b also demonstrated that there is genuine disagreement across participants in the interpretation of numbers expressed as chances (i.e., as to whether they are frequencies or probabilities). The specific interpretation of the quantities within a statistical reasoning task itself was related to performance, with participants who used a frequency representation performing better than those using a single-event probability representation (significantly better in Experiment 2a). Participants using a frequency representation were also more confident in their responses, which is consistent with previous findings by Brase (2002a) on the clarity and ease of understanding for different numerical formats. Together, these results support the position that natural frequencies constitute a privileged representational format that both facilitates statistical reasoning and that supports an intuitive satisfaction with that reasoning.

Sloman and colleagues (2003) argued that the effects of natural frequencies were "frequency illusions," in that nested set relations were key to performance rather than natural frequencies. The apparent metaphor with visual illusions is interesting, because although such illusions are errors in perception they are not arbitrary errors. The systematic nature of how visual illusions take advantage of artificial or highly unusual stimuli provide insights into the normally well-functioning system of visual perception. In truth, what appears to exist is not actually a frequency illusion, but an *illusion of frequency*; a set of circumstances in which presentations that are in some ways (from a strict theoretical position) not entirely or necessarily frequencies can nevertheless be represented in the mind as frequencies and thereby produce improved performance. This can be seen in the effects on performance depending on how participants understand the ambiguous nested sets of "chances" (Hoffrage, et al., 2002; Brase, under review), and this can also be seen in the effects of different pictorial representations on statistical reasoning (Sedlmeier, 1997, 1999; Brase, in prep).

References

- Brase, G.L. (2002a). Which Statistical Formats Facilitate what Decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*, 15, 381-401.
- Brase, G.L. (2002b). Ecological and evolutionary validity: comments on Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni's (1999) mental model theory of extensional reasoning. *Psychological Review*, 109, 722-728.

- Brase, G.L. (under review). Interpreting ambiguous statistical information as frequencies can facilitate Bayesian reasoning.
- Brase, G.L. (in prep.) The role of pictorial representations and numerical representations in statistical reasoning.
- Casscells, W., Schoenberger, A., & Grayboys, T. (1978) Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999-1000.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.
- Giroto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition*, 78, 247-276.
- Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition*, 84, 343-352.
- Johnson-Laird, P. N., Legrenzi, P., Giroto, V., Legrenzi, M. S. & Caverni, J-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, 106, 62-88.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.
- Kurzenhäuser, S. & Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher*, 24, 516-521
- Ruscio, J. (2003). Comparing bayes's theorem to frequency-based approaches to teaching bayesian reasoning. *Teaching of Psychology*, 30, 325-328.
- Sá, W., Kelley, C., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences*, 38, 1149-1161.
- Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, 91, 497-510.
- Sedlmeier, P. (1997). BasicBayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments and Computers*, Vol 29, 328-336.
- Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.
- Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 130, 380-400.
- Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior & Human Decision Processes*, 91, 296-309.
- Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, 127, 161-188.
- Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, 89, 342-357.
- Stanovich, K. E., & West, R. F. (in press). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*.