anales de psicología
2010, vol. 26, nº 1 (enero), 181-188

ISSN edición impresa: 0212-9728. ISSN edición web (http://revistas.um.es/analesps): 1695-2294

Sets or frequencies? How to explain the facilitation effect in conditional probability problems

Rodrigo Moro* ,and Gustavo Adrián Bodanza

Universidad Nacional del Sur / CONICET – Argentina

Título: ¿Conjuntos o frecuencias? Cómo explicar el efecto facilitador en problemas de probabilidad condicional.
Resumen: Desde los '70s, en el área de psicología cognitiva, el programa Heurísticas y Sesgos ha mostrado que la gente no parece razonar correctamente acerca de problemas bayesianos o de probabilidad condicional. Sin embargo, en los '90s, psicólogos del programa evolucionista descubrieron que, si los mismos problemas eran presentados de un modo diferente, las respuestas mejoraban de manera significativa. Se han ofrecido dos explicaciones para dar cuenta de este efecto facilitador: *la hipótesis de frecuencias naturales* y *la hipótesis de conjuntos anidados*. La evidencia empírica sobre el tema es diversa y no parece apuntar en una sola dirección. En este artículo haremos una revisión de la literatura e intentaremos clarificar el debate resultante en términos de estrategias y técnicas usadas por los investigadores del área. Sostendremos que la evidencia empírica presentada hasta ahora parece favorecer la hipótesis de conjuntos anidados. Sin embargo, también consideramos que hay lugar para objeciones razonables y más trabajo empírico es necesario para dirimir la disputa.
Palabras clave: Psicología cognitiva; programa de heurísticas y sesgos; programa evolucionista; inferencia bayesiana; efecto facilitador; hipótesis de conjuntos anidados; hipótesis de frecuencias naturales.

Abstract: Since the '70s, the Heuristics and Biases Program in Cognitive Psychology has shown that people do not reason correctly about Bayesian or conditional probability problems. In the '90s, however, evolutionary psychologists discovered that if the same problems are presented in a different way, people's performance greatly improves. Two explanations have been offered to account for this facilitation effect: the *natural frequency hypothesis* and the *nested-set hypothesis*. The empirical evidence on this debate is mixed. We review the literature and offer a clarification of the debate in terms of strategies and techniques used by the researchers in the area. We argue that the provided evidence seems to favor the nested-set hypothesis. However, we also argue that there is still room for disagreement and more empirical work is needed to settle the issue.
Key words: Cognitive psychology; heuristic and biases program; evolutionary program; bayesian inference; base-rate neglect; facilitation effect; nested-set hypothesis; natural frequency hypothesis.

Introduction: The facilitation effect

Many debates in Cognitive Psychology are *not* about general issues on the human mind but they are rather narrow in focus. They usually go around providing an account for particular empirical findings, such us, people struggling to solve a given mathematical problem. We will focus on a particular case of this type of debates. Behind this case, however, we will show that there is a more general dispute between rival research programs.

The story begins in the early '70s when Amos Tversky and Daniel Kahnemann founded a research program in Cognitive Psychology called 'Heuristics and Biases' (HBP henceforth). This program found extensive evidence showing that people tend to commit reasoning errors when making judgments under uncertainty (see Gilovich, Griffin, and Kahneman, 2002 for a review). A particular case involves people's tendency to fail when reasoning about conditional probability problems. Here is the most famous example of this type of problem, the medical diagnosis problem:

> Standard version of the medical diagnosis problem:
> If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? (Casscell et. al., 1978, p. 999)

Most people (even physicians!) tend to respond "95%". The correct answer according to the probability calculus is around 2%[1]. Studies show that typically less than 20% of participants get the correct answer (Casscells *et al.*, 1978; Eddy, 1982; Gigerenzer & Hoffrage, 1995; Cosmides and Tooby, 1996; Evans *et al.*, 2000; Sloman *et al.*, 2003). Thus, since the discovery of the phenomenon in the late '70s, it seemed clear that most people (without proper training) were unable to solve this type of problem.

But in the mid '90s, Gerd Gigerenzer and other evolutionary psychologists came along and gave an important turn to the state of the art. Evolutionary psychologists began noticing that a problem like the one presented above has two features: 1) the information is presented in probability format; for example, notice the information about the error rate of the test: it has "a false positive rate of 5%"; and 2) the problem asks a question about a single-event probability, namely, that a given person has the disease. Evolutionary Psychologists showed that if the same problem is framed differently, people's performance greatly improves. More specifically, if the problem presents the information under a specific format called 'natural frequency format', around 50% of participants get the correct answer, in contrast with the 20% success in the probability format. In the case of the medical diagnosis problem, the natural frequency condition would read as follows:

Natural frequency version of the medical diagnosis problem:

* **Dirección para correspondencia** [Correspondence address]:
Rodrigo Moro. Departamento de Humanidades. Universidad Nacional del Sur. 12 de Octubre y San Juan, 5to piso. 8000 – Bahía Blanca (Argentina). E-mail: rmoro@uns.edu.ar

[1] This can been seen as the result of a direct application of Bayes Rule: P (disease/positive) = P (disease and positive) / P (positive) = 0.001 x 1 / 0.001 x 1 + 0.999 x 0.05.

> One out of every 1000 American has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 999 people who are perfectly healthy, 50 of them test positive for the disease. Imagine we have arranged a random sample of 1000 Americans. They were selected by lottery. Those who conducted the lottery had no information about the health status of any of these people. Given the information above, on average, how many people who test positive for disease will *actually* have the disease? ______ out of ______

Notice that the information is presented in a frequentist format (e.g. 1 out of every 1000 American has disease X). But this is not just a frequentist format but, more specifically, it is a *natural* frequentist format. Such a format has two essential features: 1) The relevant frequencies are not normalized (neither in 100 nor in 1000 for example); 2) statistical information is presented as a partition or –as Evolutionary Psychologist like to say– as if it were obtained by *natural sampling*, that is, by updating event frequencies as information is sequentially obtained. It turns out that people perform quite better with natural frequentist versions (see Gigerenzer and Hoffrage, 1995 for a systematic study). This is a robust finding. For more than ten years, though, there has been a heated debate on how to explain this facilitation effect. The goal of this paper is to clarify such a debate offering a review of the main results.

Before we examine competing accounts of the facilitation phenomenon, it is worth mentioning that performance in conditional probability problems seems to be modulated by individual differences. For example, Brase *et al.* (2006) show that motivation and participant-recruitment source have an impact on performance: Paid participants perform significantly better than unpaid participants and top-tier university students do better than students from lower ranked universities. Tubau (2008), in turn, shows that performance is also modulated by individual differences in mathematical skills: Poor-math-level individuals perform significantly better under natural frequency conditions, but this is no so for high-math-level individuals who tend to perform well across conditions. However, Stanovich and West (2000) report empirical results that seem to go in the opposite direction. Students higher in general cognitive ability –as measured by the Scholastic Assessment Test– tend to ignore information on base-rates (i.e. prevalence of the disease in the medical diagnosis problem), which leads to wrong responses.

Two rival explanations: the Natural Frequency Hypothesis *vs.* the Nested-sets Hypothesis

There seems to be agreement about the following point. The natural frequentist version is computationally simpler than the probabilistic version. Gigerenzer and Hoffrage (1995) spell out this point by showing that the equation needed to solve the former version is simpler (i.e. it contains less elements and operations, and uses whole numbers instead of fractions or decimals) than the equation needed to solve the later version. Nonetheless, researchers in the area usually go further than this computational point when trying to account for the facilitation effect at stake.

There are two main proposals, one by members of the Evolutionary Psychology Program (EPP henceforth) and the other by members of the HBP. The natural frequency hypothesis supported by EPP basically says that the *natural frequency format* is the responsible factor for the improvement in people's performance. This format requires that both the information and the question of the problem are given in terms of natural frequencies (rather than in terms of probabilities). The advocates of EPP (e.g. Brase *et al.*, 1998) propose an explanation for why natural frequency formats are successful in promoting facilitation, based on an evolutionary hypothesis. This hypothesis postulates that the acquisition and use of frequentist information help our ancestors to survive; e.g., our ancestors realized that they were successful 5 times out of the 20 times they hunted in the north canyon. As a result, our evolved cognitive architecture favors certain frequentist ways of organizing information, which have the advantage of maintaining information that would be lost if were transformed in a single event probability: important information in "9 out of the 13 trees with red fruit are apple trees" is lost in the expression "69% of the trees with red fruit are apple trees" (for instance, by knowing the approximate number of apples per tree, the first expression favors a rough calculation of the total number of apples that can be obtained, while the second expression does not do so). Moreover, base rates are irrelevant when information is acquired by natural sampling, suggesting that our ancestors' calculations were adjusted to these conditions via repeated experiences of event frequencies (Kleiter, 1994). A complementary proposal from the Evolutionary camp is the individuation hypothesis. In addition to the frequency effect, Brase, Cosmides and Tooby (1998) postulate an individuation effect. They argue that the ability to elaborate adequate probability judgments ultimately depends on the ability to count. And the ability to count depends on the ability to individuate the world, that is, to see the world as composed of individual entities. The individuation hypothesis says that our cognitive mechanisms are better designed for operating over whole-objects rather than over arbitrary parsings of them such as parts or aspects of whole objects. The authors report a series of experiments that seem to support such a hypothesis: whole object problems elicited higher levels of performance than arbitrary parsing problems.

It must be noted that while the evolutionary hypothesis tend to justify *why* people will perform better calculations if they are given the information in natural frequency terms rather than in probability terms, it still remains uncertain whether the natural frequency format is the factor that indeed elicits the facilitation effect. This is the central point of disagreement with the HBP. So, the evolutionary explanation of why people perform well with natural frequency formats can be put aside until this is confirmed or disconfirmed.

Members of the HBP, in turn, have proposed the so-called 'nested-set hypothesis' to explain the facilitation effect.[2] The basic idea is that natural frequency versions tend to make transparent the relevant subset relations of the problem. In the medical diagnosis problem, the natural frequency version would make transparent that the set of people with positive tests includes all the sick people and also some of the healthy people (see Figure 1). When people see clearly the set relations involved in this kind of problems, they tend to use correctly base rates and thus, their performance improves. They point out that, according to this view, the success of the frequency effect does not have to do with natural frequency formats per se. They predict that any format whatsoever that makes the relevant set relations clear will aid people's probabilistic judgments. Barbey and Sloman (2007) argue that the dual process theory would provide some theoretical explanation for the nested-set effect. Dual process theory postulates that people use two systems to reason, sometimes called 'System 1' and 'System 2'. System 1 is faster, uses less information, and is based on associative principles. System 2 is slower, uses more information, and is ruled-based. This theory attributes judgmental errors –as typically occurring in conditional probability problems– to cognitive heuristics generated by some associative processes of System 1. In turn, it attributes the use of elementary set operations –and the resulting Bayesian inference in conditional probability problems– to System 2. Barbey and Sloman add that since rule-based inference is cognitively more demanding than associative inference, it is more likely to occur when participants have more time, more incentives, or more external aids to make a judgment.[3] Of course, one can defend the nested-set hypothesis without supporting the dual-process theory. So, regardless the theoretical justification, the key feature of the nested-set hypothesis is, again, the transparency of the set relations of the problem.

The empirical evidence on this debate is mixed. Some studies seem to support the natural frequency hypothesis (Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995; Krämer and Gigerenzer, 2005; Zhu and Gigerenzer, 2006) while others seem to support the nested-set hypothesis (Mellers and McGraw, 1999; Evans *et al.*, 2000; Macchi, 2000; Girotto and Gonzalez, 2001; Yamagishi, 2003; Sloman, Over, Slovak and Stivel, 2003).

We will try to clarify the debate by reviewing the strategies and techniques followed by the researchers in the area. We will then argue that the available evidence seems to favor the nested-set hypothesis. However, we will also present some reasonable objections that can be made against the provided evidence. Thus, we will conclude that issue is still open.

First strategy: Using frequency formats without a clear set structure

In order to decide between our rival hypotheses, one possible strategy is to test natural frequency versions where the nested-sets are not clarified. If people perform well under such frameworks, the natural frequency hypothesis is empirically supported. If not, the nested-set hypothesis receives empirical support. Unfortunately, this possibility is ruled out from the outset. The reason is that one of the main features of *natural* frequency versions is that they present the information in a partitive way, leaving, thus, the set structure of the problem explicitly revealed.

However, some researchers try to follow this strategy anyway (e.g. Evans *et al.*, 2000). According to them, there are some frequency versions that do not clarify the set structure of the problem and, consequently, people's performance is very poor. Nevertheless, these results are not very informative. What these results actually prove is that *non*-natural frequency versions (where frequencies are usually normalized in 100 or 1000) do not produce the facilitation effect. But the very same result is also predicted by the advocates of the natural frequency hypothesis. So, this strategy does not allow us to distinguish between our rival hypotheses.

[2] The nested-set hypothesis was first advanced by Tversky and Kahneman (1983) in the context of the phenomenon of conjunction fallacy (for a presentation of the nested-set hypothesis directly applied to conditional probability problems see Sloman *et al.*, 2003). However, this hypothesis is also defended by independent researchers who do not belong to the Heuristic and Biases program (e.g., Evans *et al.*, 2000, Girotto and Gonzalez, 2001).

[3] A different general account of the way people elaborate conditional probability judgments is given by Fox and Levav (2004) who propose the partition-edit-count model. This model postulates that when evaluating conditional probabilities, people partition the sample space into x elementary possibilities, edit out the possibilities that are to be eliminated given some received information, then count the remaining possibilities and report probabilities as the ratio of the number of focal events to the total number of events. This strategy is perfectly reasonable, but not always it is correctly performed. These authors show how different presentations of the same information can influence participants' partition processes and/or edition processes. More specifically, some presentations make people tend to invoke partitions that are insufficiently refined or fail to edit appropriately the cases that can be eliminated. Notice that this proposal does not entirely fit the ones mentioned above. First, this model, as opposed to the evolutionary proposal, does not require whole objects to work. Second, even if the partition part make it closer to the Heuristics and Biases' proposal, a correct understanding of the set relations may not lead to correct responses since the edition process may independently fail.

Second strategy: Using probability formats with a clear set structure

The main strategy left is, then, to create *probability* versions where the set structure is indeed clarified and see whether such versions elicit the facilitation effect. This is the right strategy to follow since HBP and EPP predict opposite results. HBP predicts that such versions will elicit a performance comparable to natural frequency versions of the problem while EPP predicts no facilitation at all.

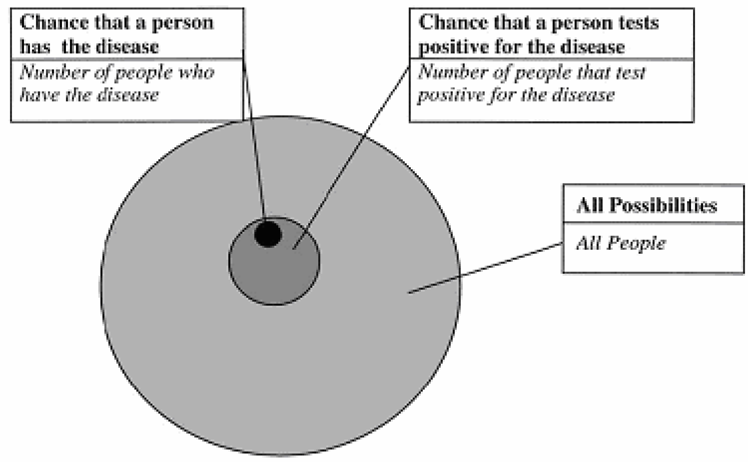

Figure 1: Nested-set relations of the medical diagnosis problem. Bold text appeared in probability conditions. Italicized text in frequency conditions. (Reproduced from Sloman *et al.* (2003, p. 298)

This strategy, however, raises an important methodological problem: how to distinguish probability versions from natural frequency versions. The problem with this distinction is that the concept of probability can be interpreted as suggesting relative frequencies. In fact, the frequentist interpretation is among the most plausible interpretations of probability (von Mises, 1957). Furthermore, it has been noted in the literature that the same numerical expressions – percentages, fractions, even whole numbers- can be legitimately used for both type of versions. In other words, the distinction is vague. While there is some agreement in the literature on classifying some wording as typically frequentist (e.g. "3 out of every 10 cases of A are also cases of B") and some wording as genuinely probabilistic or non-frequentist (e.g. "the chance or probability that the single event A is a case of B is 30%"), some researchers are not particularly careful about this point, thus leading to debates about how to interpret their results. Take, for example, a probability version used by Macchi (2000):

> *360 out of every 1000* students who sit for their High School Diploma fail.
> 75% of the students who are not awarded a Diploma fail the written Italian paper.
> However, 20% of the students who are awarded a Diploma also fail the written Italian paper.
> What percentage of those who fail the written Italian paper are not awarded a Diploma? (Macchi, 2000, p. 24, the italics is ours).

Clearly, the set structure of the problem is made transparent in this version. Notice, however, that despite of Macchi's classification, this version is closer to a frequentist format than to a probability format. In fact, the first sentence has the typical frequentist wording ("360 out of every 1000 students"). And even if the probability expressions in terms of percentages would make this version non-natural, the presence of a whole number (1000) as the reference class invites to compute all the numbers required for the natural frequentist version of the problem. Given the mixed nature of the format, a good performance (72% of correct answers) does not provide a straightforward interpretation. The facilitation may be caused by the clarification of the set relations or by the cue to use a natural frequentist approach. This criticism has been repeatedly made by members of EPP (Hoffrage *et al.*, 2002, Gigerenzer, 2007). Unfortunately, many studies in the area are susceptible of the very same criticism (e.g., Sloman *et al.*, 2003).

Taking this consideration into account, we next comment on three techniques that apply the general strategy of using probability versions with a clarified set structure. These three techniques are so far the most successful ones in showing important improvements in performance, apparently providing, thus, support for the nested-set hypothesis.

Technique 1: using improved wording

Sloman *et al.* (2003) used different versions of the medical diagnosis problem, some of which have an improved wording that reflects the set structure of the problem. Here is one of these versions:

> Improved probability format version with transparent nested-sets relations.
> The probability that an average American has disease X is 1/1000. A test has been developed to detect if a person has disease X. *If the test is given and the person has the disease, the test comes out positive. But the test can come out positive even if the person is completely healthy.* Specifically, the chance is 50/1000 that someone who is perfectly healthy would test positive for the disease. Consider an average American. Assume you know nothing about the health status of this person. What is the probability that if this person is tested and found to have a positive result, the person would actually have the disease? (Sloman *et al.*, 2003, p. 303, the italic is ours)

Notice that the technique consists of two points: 1) making the problem about an average person, so the statistical information becomes relevant; and 2) stressing out the possibility of positive tests being associated with both sick and healthy people. The problem is that Sloman and colleagues did not find consistent results. They tested several versions of probability formats *without* clarified set relations and several probability versions *with* the improved wording (and also a frequentist version). In some experiments, they found a big improvement in performance (from 20% to 48% of correct answers). However, they tested several versions because they recognized that some of these versions contain ambiguities (e.g. probability versions that may be interpreted as frequentist versions). Now, there is a comparison that Sloman *et al.* (2003) did not make. This is the comparison of the versions when ambiguities are highly reduced or eliminated. This is the most relevant comparison. The versions that contained ambiguities are not reliable because these ambiguities might be source of errors. So, the comparison that really counts is the one with no, or at least less, ambiguous materials. According to their own criteria, when facing the best probability version *without* clear set relations, *39*% of participants gave correct answers. And when facing the best probability version *with* clear set relations, *40*% of people got the correct answer. These percentages are almost identical! So, for the best version of each type, the clarification of the nested-set relation did not seem to bring any improvement.

Given the inconsistency of results with this technique, we cannot confidently pass judgment on the issue at stake. Let us then move on to evidence of the next technique to test our rival hypotheses.

Technique 2: using a *natural* chance format

One of the most effective techniques is the one used by Girotto and Gonzalez (2001). These authors have found a very clever way to express single event probabilities that emulates natural frequency setups. Here is an example of such chance format:

> The applicants for admission to a prestigious university have to pass an entrance examination which involves an oral test and a written test. Here is the information about the results of last year examination.
> An applicant had 5 chances out of 100 of being accepted. 3 of the 5 chances of being accepted were associated with success in the oral test. 7 of the remaining 95 chances of being rejected were associated with success on the oral test. Imagine that Jean is an applicant to the entrance examination. Out of 100 chances, Jean has ___ chances of passing the oral test, ___ of which will be associated with being accepted. (Girotto and Gonzalez, 2001, p. 272-273)

Under this condition, the partitive structure makes transparent the set structure of the problem. This chance format is shown to elicit a similar facilitation effect as the natural frequency version. This result seems to provide empirical support for the nested-set hypothesis.

What is the response from the advocates of the natural frequency hypothesis? The main objection is the suspicion that chances are not but "frequencies in disguise" (Hoffrage *et al.*, 2002). In fact, the structure of natural chance versions emulates the structure of natural frequency versions. Furthermore, the introduction of the problem mentions the results of last year examination. That may induce people to think in frequentist terms. These observations call into question the interpretation of the results. Again, the facilitation effect may be due to a clarification of set relations or to cues leading to think the problem in a frequentist way.

Actually, there seems to be some empirical evidence to support the suspicion of a frequentist interpretation. This comes from a study by Brase (2008). After giving participants the university admission problem, he made them choose among the following options:

- I thought about the information as a single application with some possibility of having been successful on the oral test and some possibility of having been accepted. [*probability interpretation*]
- I thought about the information as a large number of applications, some of which were successful on the oral test, and some of which were accepted. [*frequency interpretation*]
- Other: I thought about the information as ________. (Brase, 2008, p. 285)

Brase reports that an important percentage of participants (around 30%) selected the frequentist interpretation as their own. Thus, even if most people selected the probabilistic interpretation (around 60%), it is clear that the format is somewhat ambiguous. More importantly, the group that selected the frequentist interpretation performed significantly better than the group that selected the probabilistic interpretation. Thus, the reported facilitation seems to depend partially on the ambiguity of the format at stake. Hence, the main problem with the chance format is whether it can be taken as a genuine probability format or it is rather a frequency format in disguise. Until this point is clarified, Girotto and Gonzalez's results do not seem to provide conclusive evidence in favor of the nested-set hypothesis. Let us consider, then, the last technique that seems to discriminate between our rival hypotheses.

Technique 3: using graphical representations

There is an additional way to create probability versions that reveals the set structure of the problem: to include a graphical representation that show such a structure. This technique was used by Cosmides and Tooby (1996), Sloman *et al.* (2003), and Yamagishi (2003). We think this is actually one of the best ways to reveal the set structure of a problem. However, as in previous conditions, the interpretation of these results is not a straightforward matter. In fact, the same precautions should be taken to avoid suggesting a frequentist reading. This is exactly the problem with Cosmides and Tooby's study because they made participants draw a graphical representation where there was one square per represented individual, clearly suggesting a frequentist reading.

Sloman *et al.* (2003) and Yamagishi (2003), on the other hand, seem to avoid such problem. Sloman and colleagues (experiment 2) use the following version of the medical diagnosis problem:

> Consider a test to detect a disease that a given American has a 1/1000 chance of getting. An individual that does not have the disease has a 5% chance of testing positive. An individual who does have the disease will definitely test positive. What is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? _____ % (Sloman *et al.*, 2003, p. 300)

This version seems genuinely probabilistic. Under the control condition, they gave participants this version alone, and under the key experimental condition, they also included a graphical representation of the situation (see Figure 1). Sloman and colleagues reported a big improvement, from 20% of correct answers *without* the diagram to 48% *with* the diagram. This improvement was almost the same as the one elicited by the frequestist version (51%).

Yamagishi (2003), in turn, reports a similar result. He uses different variations of the following problem:

> A factory manufactures artificial gemstones. Each gemstone has a 1/3 chance that it is blurred, a 1/3 chance that it is cracked, and a 1/3 chance that it contains neither. An inspection machine removes all cracked gemstones, and retains all clear gemstones. However, the machine removes 1/2 of blurred gemstones. What is the chance that a gemstone is blurred after the inspection? (Yamagishi, 2003. p. 99)

In the key experimental condition, Yamagishi included a graphical representation as in Figure 2. Across four experiments, he reported an improvement from an average of 15% of correct answers *without* the diagram to an average of 75% *with* the diagram. In this case, the diagram effect was actually stronger than the natural frequency effect (49%).

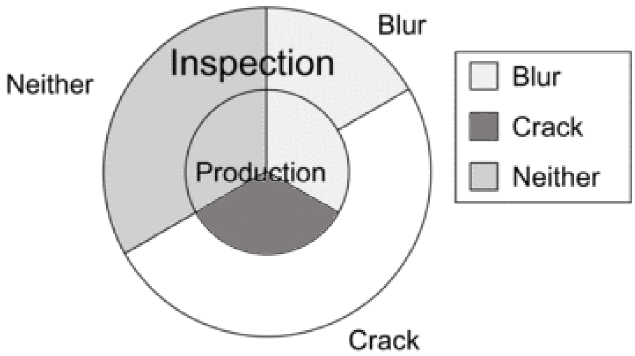

Figure 2: Graphical representation of the gemstone problem. (Reproduced from Yamagishi (2003), p. 98).

These two results are the clearest evidence in favor of the nested-set hypothesis and against the natural frequency hypothesis. None of the advocates of the last hypothesis has provided any response to this challenge. Notice, in passing, that to provide such a response is indeed a difficult task. One may initially think that the advocates of the frequentist camp should try to show that the graphical representations of Sloman's and Yamagishi's studies somehow elicit a frequentist reading. But even if this is so, no graphical condition in such studies would be regarded as *natural* frequency since all the proportions are normalized. Thus, the natural frequency hypothesis would keep predicting no improvement at all under such conditions. As a result, it seems indeed very difficult for the advocates of the natural frequency hypothesis to incorporate the results of the graphical representation technique. We will, however, finish this paper by suggesting some objections to the claim that this is conclusive evidence for the nested-set hypothesis (Sloman *et al.*, 2003; Barbey and Sloman, 2007).

The first objection has to do with the graphical representations in Sloman's and Yamagishi's studies (Figures 1 and 2). Notice that the graphical representations used in both studies give more information than the mere set structure of the problem. The figures also show (Yamagishi) or suggest (Sloman) the *relative proportions* of the sets involved. Notice that this additional feature is *not* required by the nested-set hypothesis as it is usually stated. One can perfectly present the set structure of the problem without suggesting the relative size of each set (see Figure 3). Showing the relative proportions, in turn, may clarify the very goal of the task at hand, which is, after all, to obtain a determined proportion. Thus, it may happen that what produces the facilitation effect is not the clarification of the set structure but rather the clue about proportions. Of course, this is a speculation and should be tested empirically before it can be taken seriously.

Still, it is an open question whether a graphical representation that does *not* suggest proportions can keep eliciting good performance by participants. If this is not case, we would have some negative evidence against the nested-set hypothesis –at least, as it is usually presented.

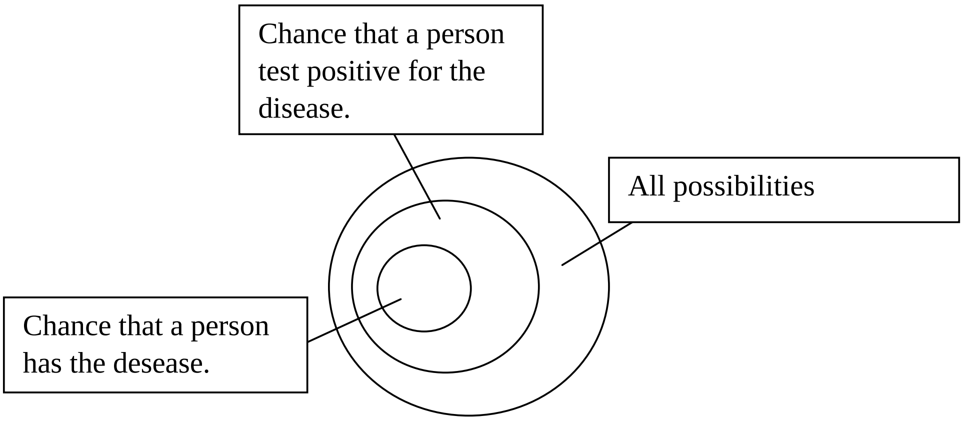

Figure 3: Graphical representation of the medical diagnosis problem that does *not* suggest the relative sizes of the classes involved.

A more general objection can be presented by pointing out the limited evidence offered in favor of the nested-set hypothesis. As argued above, neither the improved wording technique nor the chance technique has provided both consistent and clear results in favor of the nested-set hypothesis. As for the graphical representation technique, the evidence still seems very slim. The improvement in performance was shown in two studies that worked with only one problem each. Furthermore, neither of these problems seems completely adequate. In the first place, the version of the medical diagnosis problem used by Sloman and colleagues is problematic as they themselves admit (Sloman *et al.*, 2003, p. 301).[4] The gemstone problem used by Yamagishi, in turn, can be also questioned since the proportions needed to compute the solution (one half, one third, etc.) are quite simpler than the typical ones used in the literature.[5] Thus, it is an open question whether the phenomenon is robust, that is, whether the facilitation effect still occurs with other versions of the same problems and with other problems as well.

To sum up, although the evidence provided so far seems to favor the nested-set hypothesis, more empirical work is needed to provide a conclusive case for such a hypothesis.

Conclusion

Conditional probability problems are hard. Even in the most helpful conditions, people struggle to get the right answer. However, it has been shown that when problems are presented under a natural frequency format, people's performance improves. Gerd Gigerenzer and other evolutionary psychologists argue that the use of frequency formats is, thus, the factor that causes the facilitation effect. The advocates of the nested set hypothesis disagree and postulate one of the features of the natural frequency format as the responsible factor, namely, the clarification of the relevant set

[4] A first problem has to do with the fact that the information is given of an individual "getting" a disease, without specifying a time period over which the disease might be "gotten", so that whether the base rate information applies to the event at hand is questionable. A second problem has to do with the fact that this version of the problem does not mention that the person was selected at random. So, participants may assume a different prior probability for the hypothesis of the disease. It may be the case, for example, that participants assume that the patient at stake has already certain symptoms that make the doctor recommend the testing. If this is the case, one should assign a probability for the disease hypothesis higher than the base rate (1/1000). We actually discover an additional problem with this version of the problem. It has to do with the applicability of the base of the disease. On the one hand, the information about the prevalence of the disease (base rate) is given for *Americans*. On the other hand, the question of the problem is about the chance that *a person* (not explicitly an American) selected is sick. Given that the class of Americans is a subset of the class of all people, a quite different picture of the set relations may be obtained. As stated, the problem says nothing about the prevalence of the disease among non-American people. Thus, it is not clear whether the information about the base rate should be even considered by participants.

[5] Notice also that the diagram used by Yamagishi also suggests a way to compute the answer. For the numerator, it is clear that the proportion of blurred gemstones among the approved by the inspection is one half of one third. Regarding the denominator, it is clear that it is composed by the previous amount plus one third of the gemstones corresponding to the completely clear ones. Of course, nothing in the nested-set hypothesis is said about facilitating the computation of the answer. Thus, although a very interesting finding, it is difficult to interpret the result. The facilitation effect could be due either to the clarification of the set structure of the problem or to the clue about how to calculate the normative response.

relations of the problem. We argue that the right strategy to discriminate between both hypotheses was to use genuine probability problems with a clarified set structure and see whether these conditions elicit or not a performance comparable to the natural frequency effect. We review the literature and argue that there is at least one technique –the one that uses graphical representations- that seems to provoke a performance as good as the one elicited by natural formats, giving, thus, a stronger support for the nested set hypothesis. However, we do not think the last word about the issue has been said. We mentioned some reasonable objections to the claim that there is a conclusive case for the nested-set hypothesis. Thus, more empirical work is needed to settle the issue.

Acknowledgments.- We want to thank two anonymous reviewers for helpful comments and suggestions. This research was supported by ANPCYT (PICT 693 and PIDRI - PRH 2007) and SeCyT-UNS (PGI 24/ZI29 and PGI 24/ZI38).

References

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

Brase, G. L. (2008). Frequency interpretation of ambiguous statistical information facilitates Bayesian reasoning. *Psychonomic Bulletin & Review*, *15*, 284-289.

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

Brase, G. L., Fiddick, L., & Harries, C. (2006). Participant recruitment methods and statistical reasoning performance. *The Quarterly Journal of Experimental Psychology*, *59*(5), 965−976.

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

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

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In D. Kahneman, P. Slovik, & A. Tverky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 249-267). Cambridge: Cambridge University Press.

Evans, J. S., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, *77*, 197-213.

Fox, C.R. & Levav, J. (2004). Partition–edit–count: naive extensional reasoning in judgment of conditional probability. *Journal of Experimental Psychology: General*, *133*, 626-642.

Gigerenzer, G. (2007). The role of representation in Bayesian reasoning: correcting common misconceptions. *Behavioral and Brain Sciences, 30*, 264-267.

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

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment.* Cambridge: Cambridge University Press.

Girotto, 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.

Kleiter, G. (1994). Natural sampling: Rationality without base rates. In G. H. Fischer & D. Laming (Eds.), *Contributions tomathematical psychology, psychometrics, and methodology* (pp. 375-388). New York: Springer-Verlag.

Krämer, W., & Gigerenzer, G. (2005). How to confuse with statistics or: the use and misuse of conditional probabilities. *Statistical Science*, *20*, 223-230.

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes*, *82*, 217-236.

Mellers, B., & McGraw, P. (1999). How to improve Bayesian reasoning: comment on Gigerenzer and Hoffrage (1995). *Psychological Review*, *106*, 417-424.

Mises, R. v. (1957). *Probability, Statistics and Truth.* London, Allen and Unwin.

Sloman, S., Over, D., Slovak, L., & Stivel, J. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*, 296-309.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate?. *Behavioral and Brain Sciences, 23*, 645–726.

Tubau, E. (2008). Enhancing probabilistic reasoning: the role of causal graphs, statistical format and numerical skills. *Learning and Individual Differences, 18*, 187-196.

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.

(Artículo recibido: 31-3-2009; revisado: 20-5-09; aceptado: 16-11-2009)