Running head: MISSING THE DOG THAT FAILED TO BARK

Missing the dog that failed to bark in the nighttime:

On the overestimation of occurrences over non-occurrences in hypothesis testing

Paolo Cherubini[a], Patrice Rusconi[a], Selena Russo[b], Franca Crippa[a]

[a] University of Milano-Bicocca

[b] University of Trento

Author Note

Paolo Cherubini, Department of Psychology, University of Milano-Bicocca; Patrice Rusconi, Department of Psychology, University of Milano-Bicocca; Selena Russo, Department of Cognitive Sciences and Education, University of Trento; Franca Crippa, Department of Psychology, University of Milano-Bicocca.

The authors wish to thank Katya Tentori for her insightful suggestions, and Jerome Ozkan and Rénee Chow for proof reading the article.

Correspondence concerning this article should be addressed to Patrice Rusconi, Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy. Phone ++39-02-64483736. Fax: ++39-02-64483706. E-mail: patrice.rusconi1@unimib.it.

Abstract

In three studies, we investigated whether and to what extent the evaluation of two mutually exclusive hypotheses is affected by a feature-positive effect, wherein present clues are weighted more than absent clues. Participants ($N = 126$) were presented with abstract problems concerning the most likely provenance of a card that was drawn from one of two decks. We factored the correct response (the hypothesis favored by the consideration of all clues) and the ratio of present-to-absent features in each set of observations. Furthermore, across the studies, we manipulated the presentation format of the features' probabilities by providing the probability distributions of occurrences (Study 1), non-occurrences (Study 3) or both (Study 2). In all studies, both participant preference and accuracy were mostly determined by an over-reliance on present features. Moreover, across participants, both confidence in the responses and the informativeness of the present clues correlated positively with the number of responses given in line with an exclusive consideration of present features. These results were mostly independent of both the rarity of the absent clues and the presentation format. We concluded that the feature-positive effect influences hypothesis evaluation, and we discussed the implications for confirmation bias.

Keywords: the feature-positive effect; hypothesis testing; rarity effect; information sensitivity; confirmation bias.

Missing the dog that failed to bark in the nighttime:

On the overestimation of occurrences over non-occurrences in hypothesis testing

Feature-positive effects refer to the predisposition of human beings and other animals to pay more heed to the occurrences of stimuli rather than to their non-occurrences (e.g., Jenkins & Sainsbury, 1969; Newman, Wolff, & Hearst, 1980). It has been conjectured that feature-positive effects are an adaptation to a typical information pattern, whereby the occurrences of particular features are relatively rare compared to their non-occurrences, and thus they are, from a very general perspective, more informative (Newman et al., 1980). Once consolidated, this tendency to overrate the presence of stimuli may also generalize to those contexts in which the presence of certain stimuli does not necessarily convey more information than their absence. The present study investigates whether and to what extent people overrate the information value of present vs. absent features when they evaluate alternative hypotheses, that is, when they determine which of two mutually exclusive hypotheses is most likely in light of available data. This issue might have relevant practical consequences in professions in which accurate belief revision is critically important, for example, for judges who have to infer a verdict from different clues (e.g., Wells & Lindsay, 1980) or physicians who must formulate a diagnosis. For instance, in a patient with symptoms of hyperthyroidism, the assessment of normal ocular objectivity conveys the same diagnostic value as the reading of the absence of exophthalmos: Both clues should lead a physician towards a diagnosis of a form of non-Basedow thyreopathy (Scandellari, 2005). If physicians systematically underestimate the relevance of absent signs, however, the diagnostic importance of the absence of exophthalmos would be underestimated, resulting in weaker than warranted diagnostic hypotheses.

The issue of a feature-positive effect in the evaluation stage of hypothesis development is important also from a theoretical standpoint for cognitive psychology, because many scholars argue that *positive testing*, a quite common and spontaneous hypothesis-testing strategy, might result in confirmation bias if combined with feature-positive effects at the hypothesis-evaluation stage (e.g., Klayman, 1995; McKenzie, 2004, 2006). Yet, to the best of our knowledge, no direct empirical evidence has ever corroborated the idea that present clues are rated as more important than absent clues when alternative hypotheses are evaluated. In fact, “It remains to be seen to what degree feature-positive effects occur in hypothesis evaluation” (McKenzie, 2006, p. 587).

Overview of previous literature

Feature-positive effects were described in several domains. The studies of discrimination learning (e.g., Hearst & Wolff, 1989; Jenkins & Sainsbury, 1970; Newman, Wolff, & Hearst, 1980) have shown that the ability to discriminate between two stimuli that differ only by the presence or absence of a feature is acquired more rapidly and correctly when the feature is present on positive trials rather than on negative trials. In the visual perception literature, it has long been demonstrated that there are search asymmetries when the targets are characterized by the presence rather than the absence of a unique feature relative to the distractors. Specifically, it has been shown that visual search is faster when the target-defining feature is present in the target compared to when it is absent (Neisser, 1963; Treisman & Souther, 1985).

Similar effects were observed at increasingly higher levels of cognitive processing. The presence of characteristics is more relevant than their absence in the learning of concepts. The acquisition of a concept is easier for people when they receive positive instances (i.e., information about what the concept is) rather than negative instances (i.e., information of what it is not) (Bourne & Guy, 1968; Hovland & Weiss, 1953; Klayman, 1995; Nahinsky & Slaymaker, 1970). In probability learning, people tend to make their predictions on the basis of the relative frequency of the occurrence of different categories of stimuli, instead of on the basis of the actual probability of each type of stimulus because the latter would require the accurate recall of trials in which the stimulus did not occur (Estes, 1976). In yet another domain, when evaluating two-way contingency tables, people weigh the co-occurrences of stimuli more than the instances in which one or both of the stimuli is absent, a phenomenon labeled *cell weight inequality* (e.g., Beyth-Marom, 1982; Jenkins & Ward, 1965; Kao & Wasserman, 1993; Mandel & Lehman, 1998). Although it might be argued that this tendency is normatively adequate when the stimuli are rare (McKenzie & Mikkelsen, 2007), in other contexts it inflates illusory correlations (e.g., Mandel & Lehman, 1998; Smedslund, 1963).

In hypothesis testing, which is the main focus of interest of the present study, it is well known that, when gathering information for checking whether a hypothesis is true or false, there is a moderate to strong tendency to adopt a positive testing strategy (Baron, Bettie & Hershey, 1988; Cherubini, Rusconi, Russo, Di Bari, & Sacchi, 2010; Klayman, 1995; Klayman & Ha, 1987; Skov & Sherman, 1986; Slowiaczek, Klayman, Sherman, & Skov, 1992; Snyder & Swann, 1978; Wason, 1960). Positive testing, in its current understanding, affects the *gathering*, as opposed to the *evaluation*, of information. It consists of a tendency to preferentially look for the occurrence of features that are more probable when the tested hypothesis is true than when it is false. The occurrence of those features strengthens (namely, inductively confirms) the tested hypothesis, whereas their non-occurrence weakens (i.e., inductively falsifies) it. It is easy to see the possible consequences of a feature-positive effect in the evaluation stage of hypothesis testing, for individuals adopting positivity as a strategy in information-gathering. First, features whose occurrence might verify the hypothesis are searched for; second, if such confirming features indeed occur, they are attended and considered; conversely, if they do not occur, the corresponding falsification of the hypothesis could be neglected or underestimated. The result could be the systematic, improper apportionment of excessive confidence in the truth of the tested hypothesis, namely a *confirmation bias* (Klayman, 1995; McKenzie, 2004, 2007; Nickerson, 1998).

There is little empirical evidence for or against the occurrence of a feature-positive effect in the evaluation stage of hypothesis testing. Fischhoff and Beyth-Marom listed the effect as a typical deviation from a correct Bayesian evaluation of a hypothesis: “In principle, people can ignore the likelihood ratio just as well as the base rate [...]. This may happen, for example, when the datum [...] reports a non-occurrence. A classic example of the latter is Sherlock Holmes’s observation (Doyle, 1974) that his colleague, Inspector Gregory, had not considered the significance of a dog failing to bark when an intruder approached." (Fischhoff & Beyth-Marom, 1983, p. 246). However, the authors did not report empirical evidence in support of the existence and magnitude of such a tendency apart from the anecdotic reference to Arthur Conan Doyle's tale. Screening the relevant literature, we found many references to the possibility that non-occurrences are underestimated in the evaluation stage of hypothesis testing (e.g., Klayman, 1995; McKenzie, 2004, 2006; Nickerson, 1998; Slowiaczek et al., 1992), but the empirical evidence is very scant. For example, in their work on belief revisions, which was mainly focused on how people use answers to questions concerning the presence of features in individuals from a given population, Slowiaczek et al. (1992) provided some evidence of a feature-positive effect, but not consistently across studies. The only empirical investigation that we managed to find which directly and specifically tested the feature-positive effect in hypothesis evaluation is Christensen-Szalansky and Bushyhead's 1981 study on medical diagnosis in a real clinical setting: "This study also examined the physicians' ability to estimate the predictive value of an "absent symptom", since the absence of a symptom also can be helpful in assigning a diagnosis. Past psychological research has suggested that people do not efficiently process the "absence of cues" (Bourne & Guy, 1968; Hovland & Weiss, 1953; Nahinsky & Slaymaker, 1970)." (Christensen-Szalansky & Bushyhead, 1981, p. 931; the studies that the authors mentioned concern feature-positive effects in rule and concept learning, but not in hypothesis evaluation). Actually, Christensen-Szalansky and Bushyhead failed to find a significant underestimation of the diagnostic strength of absent symptoms, but were very cautious about their negative finding: "the realism of the study reduced the experimenters' control of the presence of correlated symptoms. For example, if the absence of symptom X always occurred with the presence of important symptom Y, perhaps physicians' apparent "use" of the absent symptom was simply an artefact due to this correlation. A more controlled experiment is needed to support these results" (p. 934). We did not find any more controlled experiments on this topic in later research.

Basic formal concepts about hypothesis testing

From a logical perspective, inductive hypothesis testing and belief update are mostly viewed (but see Cohen, 1977) as a change in the epistemic probability p that a hypothesis H is true (as opposed to false, corresponding to the probability that its complement, $\neg H$, is true) after acquiring a piece of evidence E, with respect to the probability that H was true before E was acquired. A widespread formal method of belief update is Bayes' rule. A simple formulation of Bayes' rule, in terms of odds and likelihood ratios (Beyth-Marom & Fischhoff, 1983; Fischhoff & Beyth-Marom, 1983), is:

$$p(H \mid E)/p(\neg H \mid E) = p(H)/p(\neg H) \times p(E \mid H)/p(E \mid \neg H)$$

The | symbol stands for a conditional probability (it can be read "given"). Reading from the left, the three terms of the formula are:

(a) the posterior odds: The ratio of the probability that H is true given E to the probability that H is false given E;

(b) the prior odds: The ratio of the probability that H was true before acquiring E to the probability that it was false;

(c) the Bayes factor–that is, the likelihood ratio of E (hereafter, LR): The ratio of the probability of observing E, assuming the truth of H, to the probability of nevertheless observing E if H were false.

The LR is a measure of the strength of confirmation (or falsification) conveyed by E. It conveys an immediate and direct description of the impact of evidence on the revision of the initial belief. If it is 1, E does not change the probability of H, and thus it is uninformative. If LR > 1, E confirms H, by increasing its probability. If LR < 1, E falsifies H, by decreasing its probability (or, correspondingly, it confirms $\neg H$ by the magnitude of 1/LR). Given the prior probability of H and its posterior probability following the receipt of E (computed by Bayes' rule), it is possible to formally estimate the information value of E in terms of Shannon's (1948) entropy–that is, in bits. The entropy of a discrete random variable X with possible values $\{x_1, \ldots, x_n\}$, each one being assigned the probability $p(x_i)$, is expressed as follows:

$$E_n(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

[whenever the *n* possible values are equally probable, this equation reduces to $log_2(n)$]. The information gain associated with a body of evidence *E*, namely ΔI_E is the difference between initial entropy and entropy after *E* has been taken into account:

$$\Delta I_E = \sum_{i=1}^{n} p(x_i)\log_2 p(x_i) - \sum_{i=1}^{n} p(x_i \mid E)\log_2 p(x_i \mid E)$$

The ΔI_E is a convenient quantitative measure for estimating the amount of information conveyed by a set of clues. However, several alternative (Bayesian) models of the utility value of clues have been proposed (e.g., Crupi, Tentori, & Gonzalez, 2007; Nelson, 2005, 2008). In particular, recent experimental work by Nelson, McKenzie, Cottrell, and Sejnowski (2010) has shown that probability gain predicted human information search better than other measures of the value of information (but see Nelson 2005 for data showing that information gain and Kullback-Leibler distance were slightly better predictors than probability gain and impact). According to probability gain the information value of the presence of a feature (e.g., the “yes” answer to a dichotomous question) is computed as:

$$\max[p(H \mid D), p(\neg H \mid D)] - \max[p(H), p(\neg H)]$$

that is, probability gain is the difference between the probability of the hypothesis favored by the evidence and the probability of the hypothesis that is most likely a priori.

In tasks in which the prior probabilities of the hypotheses are equiprobable, probability gain and impact lead to the same values of information of a datum while information gain makes the same predictions of the Kullback-Leibler distance (e.g., Nelson 2005, 2008; Nelson et al., 2010).

From a formal standpoint, it does not matter whether information is conveyed by the presence of an attribute in a situation or by its absence. A highly likely occurrence shifts the belief towards a hypothesis exactly as the non-occurrence of a highly unlikely event, and vice versa. Accordingly, testing whether the occurrence of features affects belief revisions more so than their absence is equivalent to testing whether people, with regard to their spontaneous belief revisions, are biased by a formally irrelevant aspect of the situation.

Overview

In three paper-and-pencil experiments, we investigated whether and to what extent people overestimate the importance of present features in contrast to absent ones when evaluating which of two alternative hypotheses provides a better account for a set of observations. The three experiments shared the same design, materials and procedure, except for the manipulation of whether the probability distributions of occurrences (Study 1), non-occurrences (Study 3) or both (Study 2) were presented to participants. In all the studies we also explored whether the ratio of present-to-absent features in each set of observations can affect the tendency to overestimate the importance of present features. Finally, we explored correlationally whether the informational strength of the set of present or absent features, computed as information gain and probability gain, can affect that tendency. Since participants came from the same pool and the three experiments were identical except for the presentation format of the probabilistic information, we will describe the methods and the results as if they were a single experiment.

Method

Participants

A total of 126 participants took part in the three studies. Forty-two graduate and undergraduate students of the University of Milano-Bicocca volunteered to participate in each study (first study: 18 females, 24 males; mean age = 22.7 years, range: 20-29 years; mean education = 16, *SD* = 1.7; second study: 23 females, 19 males; mean age = 22.6 years, range: 19-32 years; mean education = 15.8, *SD* = 1.5; third study: 21 females, 21 males, mean age = 22.2 years, range: 20-27 years; mean education = 15.7, *SD* = 1.7). No participants took part in more than one of the studies.

Materials and procedure

Participants received a questionnaire comprising a cover page on which some personal data (e.g., age, gender, and years of education) were collected, written instructions and 18 judgmental problems.

The instructions told participants that each judgmental task concerned two card decks—deck 1 and deck 2 (denoting two competing hypotheses). Each deck was composed of 100 cards. Each card within each deck had between zero and five letters printed on its face (between zero and four letters for the six problems in which the ratio of present and absent features was 2:2, see the Design section below). The letters were chosen from a set composed of: B, C, D, F, G (G was omitted in the problems in which the maximum number of letters was four). The instructions stated that the number of cards showing a letter was unrelated to the number of cards showing any other letter. In other words, the probability of occurrence/non-occurrence of a letter was class-conditionally independent of the probabilities of occurrence/non-occurrence of the other letters (i.e., class-conditional independence of the features). In each of the 18 problems participants were presented with a different table illustrating the number of cards (out of the 100 within each deck) on which each letter was printed. Presentation format of the numbers in the tables changed across the three studies. Specifically, participants were provided with the probabilities of the occurrence (Study 1), the occurrence and non-occurrence (Study 2), or the non-occurrence (Study 3) of each letter in each deck. In other words, the table entries reported the frequencies corresponding to the likelihoods of the features (i.e., the letters) under the two hypotheses, that is: p (E | H) and p (E | ¬H) in Study 1; p (E | H), p (E | ¬H), p (¬E | H) and p (¬E | ¬H) in Study 2; p (¬E | H) and p (¬E | ¬H) in Study 3. The likelihoods shown in Table 1 are the probabilities corresponding to the frequencies that participants received in Study 1, while in Study 2 their complements were shown in addition to them, and in Study 3 only their complements were presented.

The instructions told participants to imagine that the experimenter drew a card from a randomly selected deck without disclosing from which deck it was drawn. The random selection was meant to convey to participants the information that the prior probability of each hypothesis (i.e., each deck) was *p* = .5. Although they were not told from which deck the card was drawn, participants learnt the content of the drawn card which was described (by stating which letters, either two or three, were printed on it and which letters were absent from it) and pictorially shown to them. In each problem only one card was drawn and its content changed across the 18 problems together with the frequencies of each letter within each deck (the likelihoods corresponding to the frequencies that participants received in the 18 problems are shown in Table 1).

Participants were asked to determine from which deck the card was most likely drawn by checking one of three boxes labeled “deck 1”, “deck 2”, or “equiprobable” for each problem. Upon completion of each problem, participants were asked to express their confidence in the correctness of their answers on a 7-point scale (1 = *not confident*, 7 = *very confident*). The order of presentation for the three alternative conclusions in each problem (i.e., “equiprobable”, “deck 1” or “deck 2”) was fully balanced across participants so that six versions of the questionnaire were created in each of the three studies. Participants were individually approached in libraries and study rooms at the University of Milano-Bicocca. They were asked to participate in a study on the hypothesis-testing process, and those who accepted were given the experimental booklet. In order to familiarize participants with the task, the second and the third pages of the booklet provided the instructions and a sample problem with detailed explanations about the task and its requirements (the original Italian instructions, together with an English translation, are reported in the Appendix).

Design

The quantitative parameters and formal properties denoting each one of the problems that we used are provided in Tables 1-2. In all problems, the two subsets of present and absent features (i.e., letters) pointed in opposite directions: Namely, if the present features taken alone supported the choice of deck 1, then the absent features supported the selection of deck 2, and vice versa (see Table 2). The 18 problems were devised according to a 3 × 3 fully within-subjects design (with two different problem versions in each cell), factoring the type of the correct response and the ratio of present-to-absent features. The correct response–namely the hypothesis most probable if taking into account all of the features, including present and absent ones, according to Bayes' theorem–could match either the hypothesis suggested by the present features alone (labeled "presence-consistent" problems), the hypothesis suggested by the absent features alone ("absence-consistent" problems) or none of the above ("equiprobable" problems, in which the pattern of features was equally likely under the two alternative hypotheses). We manipulated the ratio of present-to-absent features orthogonally to the previous factor, because it might affect either the occurrence or the strength of feature-positive effects. Indeed, if it is true that feature-positive effects descend from the fact that, in general, occurrences are less likely than non-occurrences (e.g., Newman et al., 1980), then scenarios in which the number of absent features are less than the number of present features could direct attention to the former and improve the chances that they are attended. Therefore, in six of the problems, present and absent letters came in the same number (ratio of present-to-absent = 2:2); in six other problems, there were more present than absent letters (3:2), and in the remaining six problems, there were less present than absent letters (2:3) (this manipulation also varied the overall amount of letters, four in some problems and five in others).

Embedded within the main factorial design described above, we also varied non-orthogonally the informational strength (calculated as either ΔI or probability gain, Table 2 shows the utility values) of the sets of letters in order to allow correlational analyses between the informational strength and participants' choices. In the 12 non-equiprobable problems, the 4 or 5 letters overall conveyed between .12 and .32 bits of information (in terms of probability gain the range was between .2 and .32), corresponding to an increase in the probability of the correct hypothesis from the initial p (H) = .5 to a minimum posterior probability of p (H | E) = .7 and a maximum of p (H | E) = .82 (see Table 1). We chose these values of bits of information because in three previous experiments (reported in Cherubini, Russo, Rusconi, D'Addario, & Boccuti, 2009) the average threshold of information sensitivity that was measured in 130 non-expert participants engaged in abstract tasks similar to the present ones was between .12 and .18 bits. Accordingly, we ensured that our clues were informative enough to be perceived by participants. In the 6 equiprobable problems, of course, the whole set of letters overall conveyed 0 bits of information and the probability gain of all clues was null (see Table 2).

In non-equiprobable problems, the present clue subset conveyed (if its informativeness was measured while ignoring the other set) from .93 to .98 bits (probability gain: from .49 to .5), while the absent clue subset transmitted from .92 to .98 bits of information (probability gain: from .49 to .5). Hence, the two subsets of clues were highly informative. Equiprobable problems were used for presenting weaker subsets of features so that the ΔI of the subsets of features was varied on five levels (from very low, that is .01 bits, to high, that is .45 bits) for both present features and absent features (see Table 2; the probability gain relative to the present clue subset varied on five levels, from a minimum of .07 to a maximum of .37, while the probability gain relative to the absent clue subset varied on six levels, range: .06-.37).

Balancing of present/absent clue informativeness

We compared the amount of information, operationalized in terms of information gain and probability gain, conveyed by the present vs. absent clues in order to ensure that an asymmetry favoring the present over the absent clues could not be attributed to an asymmetry in the information value of these two kinds of clues. The bits of information conveyed by the present clue subset in each problem ($M = .68$, $SD = .41$) were not significantly different from those transmitted by the absent clue subset ($M = .68$, $SD = .41$), $t(17) = -.01$, two-tailed $p = .989$. Also when computing the utility values in terms of probability gain there was not a significant difference between the present clue subset ($M = .38$, $SD = .17$) and the absent clue subset ($M = .38$, $SD = .17$), $t(17) = .05$, two-tailed $p = .964$.

We also considered the informativeness of each clue as a possible source of asymmetry. In particular, participants' judgments might be directed by the high diagnosticity of a single clue. We thus computed the informativeness of each clue considering it as if it were present, according to both information gain and probability gain. In terms of information gain there was a slight asymmetry favoring the present clues ($M = .42$, $SD = .31$) over the absent clues ($M = .34$, $SD = .21$), which nonetheless was not statistically significant, $t(41) = 1.48$, two-tailed $p = .146$. Furthermore, the mean difference of .08 bits is below the threshold of participants' information sensitivity found in previous similar experiments (Cherubini et al., 2009). In terms of probability gain, the asymmetry favoring the present clues ($M = .32$, $SD = .14$) over the absent clues ($M = .3$, $SD = .12$) was less pronounced, $t(41) = .62$, two-tailed $p = .537$. Accordingly, we considered the clues sufficiently balanced to avoid any overrating of present features due to an asymmetry of features' informational strength.

Presentation format of the probabilistic information

Across the three experiments, we planned to control whether the format of the probability information affected the occurrence or the magnitude of feature-positive effects. In all previous hypothesis-testing studies that used explicit probabilities, values were used to describe the probabilities of feature occurrences. The complementary probabilities of non-occurrences, thus, were implicit and had to be derived by the participants. We conjectured that the explicit presentation of non-occurrence probabilities might reduce the cognitive load required to take them into proper account, and, at the same time, draw attention to their diagnostic relevance, thus possibly weakening feature-positive effects. In the first study, we only presented the probabilities of occurrences (the most typical manipulation used in previous studies). In the second study, we presented both the probabilities of occurrences and the complementary probabilities of non-occurrences. In the third study, we only presented the probabilities of non-occurrences.

Main dependent variables and main predictions

In all experiments, responses were primarily classified as presence-consistent, absence-consistent, or equiprobable. Presence-consistent responses were those mentioning the hypothesis that was supported by the present features (regardless of whether they were correct responses or not), and similarly absence-consistent responses reflected choices for the hypothesis supported by the absent features. According to this classification, a feature-positive effect should manifest itself as an increase of present-consistent responses with respect both to the chance level and to absent-consistent and equiprobable responses.

For further analyses, responses were re-classified as correct or incorrect. Correct responses were those in which the hypothesis supported by the whole set of features was chosen for non-equiprobable problems as well as those in which equiprobable responses were made in response to equiprobable problems; all the other responses were deemed incorrect. According to the latter classification, a feature-positive effect should manifest itself as an increase in correct responses for presence-consistent problems as compared to absence-consistent and equiprobable problems. Moderation of feature-positive effects by the present-to-absent features ratio or, between experiments, by the format of the probabilistic information is possible. Specifically, we expected that more attention should be apportioned to absent features in problems in which they are rare (3:2 problems) and in Experiments 2 and 3, in which the probabilities of non-occurrences are explicitly reported. Finally, we asked all participants to rate their confidence in each response on a 1-to-7 rating scale. According to this variable, a feature-positive effect might be observed by an increase in confidence when responses are presence-consistent rather than absence-consistent or equiprobable (or, in terms of correct/incorrect responses, by an increase of confidence in correct responses to presence-consistent problems as opposed to correct responses for all other problems).

Results

Analysis of participants' choices when the likelihoods of occurrences were presented

Table 3 reports the percentages and the standard errors of the means of presence-consistent, absence-consistent and equiprobable responses for each one of the nine experimental cells derived by the type of response × ratio of present-to-absent features experimental design relative to Study 1, in which only the probabilities of features' occurrence were made explicit to the participants. In all conditions but one, presence-consistent responses were significantly more frequent than chance. The exception was the condition in which the two decks were equiprobable and the ratio of present-to-absent features was 3:2. The absence-consistent responses were at chance level in most conditions. They dropped below chance level in the condition in which the correct response was presence-consistent and the ratio was either 2:2 or 2:3 as well as the condition in which the two decks were equiprobable and the ratio was 2:3. The frequencies of equiprobable responses were significantly less than chance in all conditions except for the conditions in which the two decks were equiprobable. These findings hint at a strong feature-positive effect. The response suggested by the present features was the preferred one in most conditions, both when it was the correct response (upper rows in Table 3) and when it was incorrect (middle and bottom rows in Table 3), corroborating the conjecture that present features are the ones most considered when evaluating which hypothesis fits best with a set of data.

Correct responses and presence-consistent responses when the likelihoods of occurrences were presented

Table 3 hints at a possible interaction between the type-of-correct-response factor and the present-to-absent-features-ratio factor. In order to explore this interaction, we analyzed the mean rates of correct responses (the bold diagonal in Table 3). Because the ANOVA is an improper test for count data ranging from zero to two *per* cell (e.g., Jaeger, 2008), we ran a generalized linear model for repeated measures with a Poisson distribution for the response variable by means of the SAS™ statistical package, factoring the type of correct response (presence-consistent vs. absence-consistent vs. equiprobable) and the ratio of present-to-absent features (2:2 vs. 3:2 vs.2:3). The first-level effect of the type of correct response was significant, $\chi^2 = 24.44$, df = 2, $p < .0001$ ($M_{\text{presence-consistent}} = 1.31$, $M_{\text{absence-consistent}} = .65$, $M_{\text{equiprobable}} = .55$), confirming that correct responses were more frequent in the presence-consistent than in the absence-consistent, $\chi^2 = 11.59$, df = 1, $p = .0007$ (Bonferroni correction), or equiprobable, $\chi^2=24.31$, df=1, $p < .0001$, conditions. The first-level effect of the ratio of present-to-absent features was not significant, $\chi^2 = 2.47$, df = 2, $p = .29$. Beyond suggesting that the ratio of present-to-absent features does not have by itself a main influence on the frequency of correct responses, this finding also shows that the different number of clues in the three conditions (five clues in the 3:2 and 2:3 conditions vs. four clues in the 2:2

conditions) did not have appreciable effects on responses. The two-way interaction was significant, $\chi^2 = 11.29$, df = 4, $p < .05$. The interaction probably emerged from the decrease of correct responses for presence-consistent problems in the 3:2 present-to-absent ratio condition and from the increase of correct responses in the absence-consistent and equiprobable problems in the 3:2 present-to-absent ratio condition (see Figure 1, left panel, and Table 3, bold diagonal). This finding is consistent with the prediction that absent clues are apportioned more attention when they are less frequent than present clues. A similar trend, this time indicated by a first-level effect for the ratio of present-to-absent features, was observed for the occurrence of presence-consistent responses, regardless of their correctness (column 1 of Table 3).

We statistically explored this interaction by means of another generalized Poisson model, featuring the number of presence-consistent responses as the dependent variable and factoring the type of problem and the ratio of present-to-absent features. The analysis yielded a significant first-level effect for the ratio of present-to-absent features, $\chi^2 = 8.65$, df = 2, $p = .0132$ ($M_{2:2\text{ problems}} = 1.18$; $M_{3:2\text{ problems}} = .97$; $M_{2:3\text{ problems}} = 1.21$). This finding corroborates the idea that present features drive attention less when the absent features are rare than when they are equally frequent or more frequent than present features. The first-level effect of the type-of-problems factor was also significant, $\chi^2 = 14.18$, df = 2, $p < .001$ ($M_{\text{presence-consistent problems}} = 1.31$, $M_{\text{absence-consistent problems}} = 1.12$, $M_{\text{equiprobable problems}} = .93$), suggesting that, although participants in aggregate form had an overall preference for presence-consistent responses, they were also sensitive to the formal correctness of the response (presence-consistent vs. equiprobable: $\chi^2 = 12.43$, df = 1, $p = .0004$; presence-consistent vs. absence-consistent: $\chi^2 = 5.58$, df = 1, $p = .0181$; absent-consistent versus equiprobable was not significant, $\chi^2 = 3.15$, df = 1, $p = .0757$). The two-way interaction was not significant (Figure 1, right panel).

Analysis of participants' choices when likelihoods of both occurrences and non-occurrences were presented

Table 4 reports the percentages and the standard errors of the means of presence-consistent, absence-consistent and equiprobable responses in each experimental condition when participants were provided with both the probability distributions of occurrences and the likelihoods of non-occurrences (i.e., Study 2). In all conditions but one, presence-consistent responses were significantly more frequent than chance. The exception was the same as in Study 1–that is, the condition in which the two decks were equiprobable and the ratio of present-to-absent features was 3:2. Equiprobable responses were at chance level or significantly below it in all conditions. The absence-consistent responses were below chance level in most conditions, except for equiprobable 3:2 problems (in which they were above the chance level of 33%; this was the only condition in which they were preferred to presence-consistent responses) and either absence-consistent or equiprobable 2:3 problems (in which they were at chance level). These preliminary tests apparently replicated the strong feature-positive effect observed in the previous study: The response suggested by the present features was the preferred one in most conditions, both when it was the correct response (upper rows in Table 4) and when it was incorrect (middle and bottom rows in Table 4), with only one exception.

Correct responses and presence-consistent responses when the likelihoods of both occurrences and non-occurrences were presented

The frequency of correct responses was analyzed by means of a generalized repeated-measures model for a Poisson distribution, factoring the type of correct response (presence-consistent vs. absence-consistent vs. equiprobable) and the ratio of present-to-absent features (2:2 vs. 3:2 vs. 2:3). The first-level effect for the type of correct response was significant, $\chi^2 = 19.98$, df $= 2$, $p < .0001$ ($M_{\text{presence-consistent}} = 1.36$, $M_{\text{absence-consistent}} = .48$, $M_{\text{equiprobable}} = .53$), confirming that correct responses were more frequent in presence-consistent than in the absence-consistent, $\chi^2 = 19.17$, df $= 1$, $p < .0001$ (Bonferroni correction) or equiprobable, $\chi^2 = 24.99$, df $= 1$, $p < .0001$, conditions. The first-level effect of the ratio of present-to-absent features was not significant, $\chi^2 = 4.42$, df $= 2$. The two-way interaction was significant, $\chi^2 = 10.9$, df $= 4$, $p < .05$. The statistical results closely match those of Study 1. However, in Study 1, the interaction was driven by a relative increase in correct responses for absence-consistent and equiprobable problems with a 3:2 ratio of present-to-absent features along with a decrease in accuracy in the 3:2 presence-consistent problems. By contrast, in this study, the interaction was probably due to the higher frequency of correct responses in the equiprobable problems compared to the absence-consistent problems for the 2:2 and 3:2 conditions, which reversed into a greater accuracy in the absence-consistent problems than in the equiprobable problems for the 2:3 condition (see Figure 2, left panel, and Table 4, bold diagonal).

Similar to Study 1, a generalized Poisson model featuring the number of presence-consistent responses as the dependent variable and factoring the type of problem and the ratio of present-to-absent features yielded significant first-level effects for the type of problem, $\chi^2 = 19.98$, df = 2, $p < .0001$ ($M_{\text{presence-consistent problems}} = 1.36$, $M_{\text{absence-consistent problems}} = 1.18$, $M_{\text{equiprobable problems}} = 0.80$) and the ratio of present-to-absent features, $\chi^2 = 8.17$, df = 2, $p < .05$ ($M_{\text{2:2 problems}} = 1.09$, $M_{\text{3:2 problems}} = 1.04$, $M_{\text{2:3 problems}} = 1.21$). The former effect replicates the one found in Study 1, showing that, beyond generally preferring the presence-consistent responses, participants were also partly sensitive to the correct responses. The latter effect shows a tendency for the preference towards present-consistent responses to decrease for the problem versions in which there were two absent clues (i.e., the 2:2 and 3:2 problems) as compared to those in which they were three (the 2:3 problem versions), and it was qualified by the significant two-way interaction, $\chi^2 = 14.28$, df = 4, $p < .01$ (see the first column of Table 2 and Figure 2, right panel). The interaction probably emerged from the decrease in presence-consistent responses in the 3:2 condition of the equiprobable problems in comparison to the 2:2 and 2:3 conditions as well as from the increase in presence-consistent responses in the 2:3 condition of the equiprobable problems in comparison to the 2:2 and 3:2 conditions It might derive from a rarity effect, in turn favoring either absent features (in terms of a drop of presence-consistent responses in the equiprobable 3:2 problems) or present features (in terms of an increase in the presence-consistent responses in the equiprobable 2:3 problems).

Analysis of participants' choices when the likelihoods of non-occurrences were presented

Table 5 reports percentages and the standard errors of the means of presence-consistent, absence-consistent and equiprobable responses in each experimental condition in Study 3, in which we exclusively presented to participants the probabilities of non-occurrences. In all of the presence-consistent and absence-consistent problems, presence-consistent responses were significantly more frequent than chance. An exception was with equiprobable problems, with presence-consistent responses at chance level in all conditions, whereas, in Studies 1 and 2, they were at chance level only in the 3:2 versions of the equiprobable problems. Absence-consistent responses, which were mostly below chance levels in the previous studies, were mostly at chance level in the present study, possibly indicating a marginal increase in the attendance to absent features. Equiprobable responses were mostly below chance level, except for the equiprobable problems, in which they were at chance level. Divergences from Studies 1 and 2 are small: The overall pattern still suggests a rather strong, quite generalized preference for attending to present features over absent ones.

Correct responses and presence-consistent responses when the likelihoods of non-occurrences were presented

We analyzed the rate of correct answers by means of a generalized repeated-measures model for a Poisson distribution, factoring the type of correct response (presence-consistent, absence-consistent, equiprobable) and the ratio of present-to-absent features (2:2, 3:2, 2:3). The first-level effect for the type of correct response was once again significant, $\chi^2 = 11.14$, df = 2, $p < .005$ ($M_{\text{presence-consistent}} = 1.2$, $M_{\text{absence-consistent}} = .51$, $M_{\text{equiprobable}} = .67$), confirming that correct responses were more frequent in the presence-consistent than in the absence-consistent, $\chi^2 = 10.59$, df = 1, $p = .0011$ (Bonferroni correction), or equiprobable, $\chi^2 = 9.15$, df = 1, $p = .0025$, problems. As occurred in the previous studies, the first-level effect for the ratio of present-to-absent features was not significant, $\chi^2 = .80$, df = 2. However, in contrast to the previous studies, the two-way interaction was also not significant (see Figure 3, left panel): Hence, the ratio of present-to-absent features in this version of the task had no appreciable effects whatsoever on the frequency of correct responses.

A second generalized Poisson model featured the number of presence-consistent responses as the dependent variable and factored the type of problem and the ratio of present-to-absent features. Similar to the two previous studies, it yielded a significant first-level effect for the type of problem, $\chi^2 = 17.58$, df = 2, $p < .0005$ ($M_{\text{presence-consistent problems}} = 1.2$, $M_{\text{absence-consistent problems}} = 1.13$, $M_{\text{equiprobable problems}} = .67$). Surprisingly the effect shows that the presence-consistent responses were more frequent in the presence-consistent problems in comparison to the equiprobable problems, $\chi^2 = 16.64$, df = 1, $p < .0001$ (Bonferroni correction) and in the absence-consistent problems in comparison to the equiprobable problems, $\chi^2 = 146.96$, df = 1, $p < .0001$. However, presence-consistent responses were not significantly more frequent in presence-consistent vs. absence-consistent problems, $\chi^2 = .60$, df = 1, $p = .4368$ (see Figure 3, right panel). If anything, this pattern hints at a strengthening, instead of a weakening, of the feature-positive effect in this version of the task, as far as the rate of presence-consistent responses are concerned. The first-level effect of the ratio of present-to-absent features was not significant, $\chi^2 = .27$, df = 2, nor it was the two-way interaction, $\chi^2 = 6.41$, df = 4, confirming that the rarity of features in this study did not appreciably affect the overall preference for the responses suggested by present features.

Confidence ratings

We then analyzed the mean confidence ratings toward the correct responses (derived from a 1-to-7 rating scale, in which 7 indicates extreme confidence and 1 reflects no confidence) in the three experiments. Participants expressed their confidence only relative to the responses they chose, thus the dataset included a relatively large number of missing values requiring a powerful statistical analysis. Accordingly, we analyzed the data by means of a linear mixed-effects model which provides greater statistical robustness than ANOVA (e.g., Blom & Baayen, in press). In the analysis, the experiment and the correct-response type were introduced as potentially significant fixed effects. Participant, questionnaire version, and the different judgmental problem were modeled as random-effect factors. Fixed-effect factors were modeled by contrasting each level of a factor to a specified reference level. The levels of experiment (the reference level is shown in boldface) were: **Study 1**, Study 2, Study 3. The levels of correct-response type were: presence-consistent, **absence-consistent**, equiprobable. We started with a full factorial model, which was progressively simplified by removing the predictors that did not significantly contribute to the goodness of fit of the model. We tested both first-level effects and the interaction between the fixed-effect factors. The statistical significance of the fixed effects was determined using a Markov chain Monte Carlo (MCMC) sampling algorithm with 10,000 samples.

The final model included the correct-response type as significant fixed-effect factor and participant, questionnaire version, and judgmental task as significant random-effect factors. In particular, the presence-consistent level of correct-response type was not significant (MCMC mean = .26, *p*MCMC = .104), indicating that confidence toward the correct presence-consistent response was not significantly different from confidence toward the correct absence-consistent response. By contrast, the equiprobable level of correct-response type was significant (MCMCmean = -.56, *p*MCMC = .002), indicating that confidence in the correct equiprobable responses was significantly lower than confidence in the correct absence-consistent responses. We also took the presence-consistent correct response as the reference level in order to compare presence-consistent vs. equiprobable correct-response types. It emerged that the equiprobable level of correct-response type was significant (MCMC mean = -.82, *p*MCMC = .0001), indicating that confidence in the correct equiprobable responses was significantly lower than confidence in the correct presence-consistent responses. Neither the first-level effect of the experiment nor the experiment × correct-response type interaction emerged as significant predictors of confidence toward the correct responses (see Figure 4, left panel).

We ran a second linear mixed-effects model on the confidence ratings expressed in the presence-consistent, absence-consistent, and equiprobable responses provided by participants. Experiment and response type were modeled as fixed-effect factors, and participant, questionnaire version, and the judgmental task were modeled as random-effect factors. Fixed-effects factors were modeled by contrasting each level of a factor to a specified reference level. The levels of experiment (the reference level is shown in boldface) were: **Study 1**, Study 2, and Study 3. The levels of response type were: presence-consistent, **absence-consistent**, and equiprobable. We tested both first-level effects and the interaction between the fixed-effect factors. We used the same procedure as in the previous analysis, in which an initial full factorial model was progressively simplified by removing predictors that did not significantly improve the goodness of fit of the model. The statistical significance of the fixed effects was determined using a Markov chain Monte Carlo (MCMC) sampling algorithm with 10,000 samples.

The final model included response type as significant fixed-effect factor, and participant, questionnaire version and the judgmental task as significant random-effect factors. In particular, the equiprobable response-type level was significant (MCMC mean = -.35, *p*MCMC = .0001), indicating that participants' confidence in the equiprobable responses they gave significantly decreased with respect to confidence in the absence-consistent responses. Also the presence-consistent response-type level was significant (MCMC mean = .15, *p*MCMC = .005), indicating that participants were significantly more confident in the presence-consistent responses than in the absence-consistent responses. In order to contrast the presence-consistent level to the equiprobable level we also took the presence-consistent response type as the reference level. The equiprobable response-type level was significant (MCMC mean = -.5, *p*MCMC = .0001), indicating that confidence in equiprobable responses was significantly lower than confidence in presence-consistent responses. Neither the first-level effect of experiment nor the experiment × response type interaction were significant predictors of confidence ratings (see Figure 4, right panel).

Correlations between the normative utility values and presence-/absence-consistent responses

By aggregating data across participants for the 18 problems, we performed tests of correlation between the number of presence-consistent and absence-consistent choices in each problem and either the ΔI in bits or the utility values defined by the probability gain conveyed in that problem by the two subsets of present or absent clues (see Table 2 for the exact values). In Study 1, both the ΔI and the probability gain of present clues were strongly correlated with the number of presence-consistent choices, $r_{\Delta I} = .61$, $N = 18$, two-tailed $p = .008$, $r_{\text{P.G.}} = .62$, $N = 18$, two-tailed $p = .006$. In Study 2, we found an even stronger association of the normative utility values of the present clue subset with the number of presence-consistent responses, $r_{\Delta I} = .82$, $N = 18$, two-tailed $p < .001$, $r_{\text{P.G.}} = .84$, $N = 18$, two-tailed $p < .001$. The results of Study 3 confirmed this tendency, $r_{\Delta I} = .91$, $N = 18$, two-tailed $p < .001$, $r_{\text{P.G.}} = .88$, $N = 18$, two-tailed $p < .001$. These findings indicate that, in aggregated form, participants were sensitive to the amount of information conveyed by the present clues. In other words, the more the present clues were informative, the more likely the choice of the presence-consistent response was, irrespective of the information format of the probabilistic information they received.

Conversely, a different pattern emerged when considering the association between absence-consistent responses and the utility of the absent clue subset. In Study 1, there was not a significant correlation between absence-consistent choices and either the ΔI of absent clues, $r = .06$, $N = 18$, two-tailed $p = .814$, or the probability gain, $r = .01$, $N = 18$, two-tailed $p = .978$. In Study 2, there was a significant negative correlation between absence-consistent responses and either the ΔI of absent clues, $r = -.6$, $N = 18$, two-tailed $p = .009$, or the probability gain, $r = -.6$, $N = 18$, two-tailed $p = .008$. Study 3 confirmed this significant negative association, $r_{\Delta I} = -.56$, $N = 18$, two-tailed $p = .016$, $r_{\text{P.G.}} = -.62$, $N = 18$, two-tailed $p = .006$. Hence, these results indicate that either the amount of information conveyed by absent clues did not appreciably affect the decision to choose or not choose the absence-consistent response (Study 1), or the more the absent clues were informative, the less likely the choice of the absence-consistent response was (in Studies 2 and 3, in which the probability distributions of non-occurrences were available).

Cross-experimental analyses and discussion

The three studies were run sequentially, and participants were not assigned randomly to the three samples. Apart from that, the studies were homogeneous: The participants came from the same pool; the procedure and stimuli were the same, except for the presentation format of the probabilistic information; and the sample sizes were the same. Hence, a statistical cross-examination of the three studies could theoretically be reliable. The mean number and percentage of choices (ranging from 0 to 2) for each response in each problem across the 126 participants (252 responses) to the three studies are reported in Table 6.

Presence-consistent responses were more frequent than chance in all conditions for both correct and incorrect responses, with the exception of the equiprobable problems with a 3:2 ratio of present-to-absent features, in which they were at chance level. All other responses (again, both correct and incorrect ones) were at chance level or below it. The pattern hints at a strong feature-positive effect, which is only slightly moderated by the correctness of responses and by the ratio of present-to-absent features.

Analysis of correct responses across the three studies

We analyzed the relative frequency of correct responses (Table 6, bold diagonal) by means of a generalized mixed model for a Poisson distribution, factoring the type of problems and the ratio of present-to-absent features within-participants and the format of probabilistic information as a between-groups variable. The first-level effect for the type of problems was significant, $\chi^2 = 55.87$, df = 2, $p < .0001$ ($M_{\text{presence-consistent problems}} = 1.29$, $M_{\text{absence-consistent problems}} = .55$, $M_{\text{equiprobable problems}} = .59$). This finding indicates that, independent of all other factors, namely the ratio of present-to-absent features and the presentation format of probabilistic information, correct responses were more frequent when they were backed by present features than when they were congruent with absent features, $\chi^2 = 40.70$, df = 1, $p < .0001$ (Bonferroni correction), or were inconsistent with both present and absent features (i.e., equiprobable problems), $\chi^2 = 54.79$, df = 1, $p < .0001$ (see Figure 5, left panel). Notice that, because of the structure of the problems, this means that correct responses were more frequent when they were *opposite* to the responses congruent with the absent features. The first-level effects for the ratio of present-to-absent features, $\chi^2 = 5.05$, df = 2, $p = .08$, and for the presentation format of the probabilistic information were not significant, $\chi^2 = 2.13$, df = 2, $p = .34$. All of the two-ways interactions did not reach significance. The three-way interaction was significant, $\chi^2 = 21.22$, df = 8, $p < .01$, probably originating from the different trends of the type of problems × ratio of present-to-absent features two-way interactions across the three experiments.

Analysis of presence-consistent responses across the three studies

A second generalized mixed model for a Poisson distribution with the same factors as the previous one was run to analyze the frequency of presence-consistent responses (Table 6, first column). It yielded a significant first-level effect for the type of problem, $\chi^2 = 49.46$, df = 2, $p < .0001$ ($M_{\text{presence-consistent problems}} = 1.29$, $M_{\text{absence-consistent problems}} = 1.14$, $M_{\text{equiprobable problems}} = .80$). Presence-consistent responses were significantly more frequent in presence-consistent than absence-consistent, $\chi^2 = 8.46$, df = 1, $p = .0036$ (Bonferroni correction) and equiprobable problems, $\chi^2 = 40.53$, df = 1, $p < .0001$. However, presence-consistent responses were also significantly more frequent in absence-consistent than in equiprobable problems, $\chi^2 = 27.61$, df = 1, $p < .0001$. Although the increased frequency of presence-consistent responses in presence-consistent problems, in which they were correct, in comparison to absence-consistent problems, in which they were incorrect, shows a residual sensitivity to the formal correctness of responses, their increased amount in comparison to equiprobable problems (that is apparent also for presence-consistent responses in absence-consistent problems) probably reflects the fact that, in the latter problems, the amount of information conveyed by present or absent clues was very small. Thus, the finding supports the idea that participants are mostly sensitive to the amount of information conveyed by present clues, as shown by correlations with the ΔI and the probability gain of the present clue subsets in each problem (see below). The first-level effect for the ratio of present-to-absent features was also significant, $\chi^2 = 12.62$, df = 2, $p < .005$ ($M_{2:2\text{ problems}} = 1.10$, $M_{3:2\text{ problems}} = 1.00$, $M_{2:3\text{ problems}} = 1.13$). However, this effect is best accounted for by the significant type of problem × ratio of present-to-absent features two-way interaction, $\chi^2 = 22.96$, df = 4, $p < .0001$, which shows that presence-consistent responses were indeed less frequent in the 3:2 problems, but only in the equiprobable problems (see Figure 5, right panel, and Table 6, first column). Thus, the rarity of absent clues can draw attention to absent features, albeit exclusively in circumstances in which the amount of information conveyed by the two subsets of present or absent clues is tiny. No other first-level effects or interactions reached significance.

Confidence ratings and sensitivity to the clue informativeness across the three studies

The mean confidence across the 18 problems in the three studies was positively correlated with the number of presence-consistent responses, $r = .67$, $N = 18$, two-tailed $p = .002$, whereas it was negatively correlated with the number of equiprobable responses, $r = -.76$, $N = 18$, two-tailed $p < .001$. Conversely, the mean confidence did not correlate significantly with either the number of correct responses, $r = .37$, $N = 18$, two-tailed $p = .135$, or with the number of absence-consistent responses, $r = -.26$, $N = 18$, two-tailed $p = .296$. That is, the more participants chose presence-consistent responses, the more they trusted their choices, whereas the more they chose equiprobable responses, the less they trusted their choices. By contrast, confidence did not appreciate as a function of either the actual number of formally correct choices or the number of absence-consistent choices.

By aggregating data across participants for the 18 problems, we performed tests of correlation between the number of presence-consistent and absence-consistent choices in each problem and either the ΔI in bits or the probability gain conveyed in that problem by the two subsets of present or absent clues. Both the ΔI and the probability gain of present clues were positively correlated with the number of presence-consistent choices, $r_{\Delta I} = .85$, $N = 18$, two-tailed $p < .001$, $r_{\text{P.G.}} = .86$, $N = 18$, two-tailed $p < .001$. Conversely, both the ΔI and the probability gain of absent clues were negatively correlated with the number of absence-consistent choices, $r_{\Delta I} = -.51$, $N = 18$, two-tailed $p = .031$, $r_{\text{P.G.}} = -.56$, $N = 18$, two-tailed $p = .015$.

General Discussion

This study lends conclusive support to one main finding and less strong support to some ancillary findings, which merit further investigation.

Main finding: The feature-positive effect on the evaluation of alternative hypotheses

Results of the current study show that people overrate the information conveyed by the presence of clues in comparison to that conveyed by the absence of other clues when they evaluate available data for establishing which of two competing hypotheses is the most likely. Previous studies have reported this tendency (e.g., Fischhoff & Beyth-Marom, 1983; Slowiaczek et al., 1992), but no conclusive empirical evidence could directly support it. To our knowledge, the only study that directly investigated this issue with quasi-experimental methods failed to find support for it (Christensen-Szalansky & Bushyhead, 1981), although the authors attributed their negative finding to possible artifacts. The Christensen-Szalansky and Bushyhead (1981)'s study used Bayes' theorem as criterion for evaluating physicians' calibration, but its nature was correlational (Slovic & Lichtenstein, 1971). Accordingly, in the present experiments we used orthogonal designs instead of intercorrelated cues that are representative of the real world.

In the present studies, there are at least four sources of converging evidence for the occurrence of a relatively strong feature-positive effect in the evaluation of alternative hypotheses:

1) In all experiments, the hypothesis consistent with the information conveyed by present clues and therefore inconsistent with the information conveyed by absent clues was preferred significantly above chance level in most conditions, regardless of whether it was the formally correct response or not. There were only a few exceptions, with presence-consistent responses at chance level, that emerged in some instances in which the two hypotheses were formally equiprobable. However, in those problems, the subset of present clues was formally very weak (that is, it conveyed a very low ΔI/probability gain). Because participants were sensitive mostly to the information conveyed by present clues (see point 4, below), it is not surprising that, in those problems, their preference for the positive-consistent responses was weakened.
2) In all of the studies, the formally correct responses were chosen significantly more often when they were consistent with the responses indicated by the present clues than when

they were consistent with absent clues or were inconsistent with both present and absent clues (i.e., equiprobable problems).

3) Overall, across the three experiments, the mean confidence in the presence-consistent responses was higher than the mean confidence in either absence-consistent or equiprobable responses (see Figure 4, right panel). Furthermore, as shown by the cross-experimental analysis, the mean confidence toward responses to the 18 problems, across participants, was positively correlated with the number of presence-consistent choices that were selected for those problems. It did not correlate significantly with the number of correct or absence-consistent choices and it was negatively correlated with the number of equiprobable responses.
4) In all studies, the amount of information (as measured by ΔI and probability gain) conveyed by the subset of present clues in each problem correlated positively with the number of presence-consistent choices on that problem, across participants. The amount of information conveyed by the subset of absent clues either did not correlate significantly (Study 1) or was negatively correlated with the number of absence-consistent choices (Studies 2-3). These intriguing findings suggest that, although humans are probably sensitive to some extent to the informativeness of data (e.g., Cherubini et al., 2009; Oaksford & Chater, 1994), this is mostly the case when they evaluate the meaning of occurrences. Apparently, people can sometimes perceive the informativeness of the absence of some features, in particular when the probability distributions of non-occurrences are explicit. However, in those instances, on average, they do not revise their beliefs consistently with that information.

These converging pieces of evidence are mostly independent of the presentation format of the probabilities of the clues under the two alternative hypotheses, which was manipulated across the three studies. They are also mostly independent of the ratio of present-to-absent features presented in each problem, which was manipulated within each study.

Ancillary findings: Possible moderators of the feature-positive effect

The influence of the rarity of the absent clues

Rarity effects concern the apportionment of increased attention to rare events in contrast to common ones (e.g., Feeney, Evans & Clibbens, 2000; Feeney, Evans & Venn, 2008; Green & Over, 2000; McKenzie & Mikkelsen, 2000, 2007; Oaksford & Chater, 1994; 2003; in legal contexts, for example, see Loftus, 1976; Wells & Lindsay, 1980). We included in our initial predictions a hypothesis that was based on rarity effects, conjecturing that participants would possibly pay more heed to absent clues when they were rare in comparison to present ones. The prediction followed from Newman et al. (1980)'s evolutionary-based argument that feature-positive effects originate from the fact that, in nature, the occurrence of events is less common than the non-occurrence of events and thus is, in a very general sense, more informative. Following from that argument, in specific contexts in which absent clues occur less than present clues, an opposite trend to pay heed to absent clues could arise. Accordingly, we devised different versions of each problem, varying the ratio of present-to-absent clues along three levels (2:2; 3:2; 2:3). Results were inconclusive with respect to the original prediction. A slight weakening of the feature positive effects occurred in Study 1 in the 3:2 problems, as shown by the type of problems × ratio of present-to-absent two-way interaction for the frequency of correct responses observed in that study. However, the interaction, although it was still significant, followed a distinctively different pattern in Study 2 and was not significant in Study 3 (thus giving rise to the three-way interaction observed in the cross-experimental analyses of correct responses). The cross-experimental analyses of the presence-consistent responses showed a decrease of presence-consistent choices occurring in the 3:2 equiprobable problems only, that is, in those problems in which the present clues were least informative. This set of different findings suggests that the rarity of absent clues might, in some circumstances, draw attention to them, but this effect is not systematic, and it apparently interacts with the presentation format of probabilistic information as well as with the amount of information conveyed by the stimuli in ways that are in need of further specification.

The effect of the presentation format of the probabilistic information

In most past experiments on hypothesis testing and evaluation that used explicit probabilistic information, only the probabilities of the occurrences of different features were communicated to participants (e.g., Cherubini et al., 2010; McKenzie, 2006; Skov & Sherman, 1986; Slowiaczek et al., 1992). We conjectured that this format might inflate feature-positive effects, because the probabilities of non-occurrences have to be inferred by complementation. Accordingly, we systematically changed the way probabilities were communicated to the participants across the three studies: In Study 1, we communicated the probabilities of occurrences; in Study 2, we communicated the probabilities of occurrences and non-occurrences; in Study 3, we communicated exclusively the probabilities of non-occurrences. Contrary to the initial conjecture, the probabilistic format did not have appreciable effects on choices, as shown by the cross-experimental analyses. It had some effect on confidence ratings, although the two linear mixed models we ran showed that neither the first-level effect of the type of experiment nor the experiment × response type interaction were significant, thus indicating that the pattern was homogeneous across the studies (see Figure 4). Apparently, communicating explicitly the probabilities of non-occurrences gave a hint to the participants that those probabilities should be considered and thus they perceived their diagnosticity but they seemed to disregard this information when choosing the most likely hypothesis as shown by the significant negative correlations between the informativeness (calculated as ΔI and probability gain) conveyed by the absent clue subset and the number of absence-consistent choices in Studies 2 and 3 (but not in Study 1, in which non-occurrences were not explicitly presented to participants).

The only other appreciable effect of the different probabilistic formats is its interaction with the rarity of absent features, as discussed in the previous paragraph.

Conclusion

The present scrutiny shows that, in the evaluation stage of hypothesis development, the occurrence of clues is systematically overrated with respect to the non-occurrence of other clues.

The tendency to neglect the significance of the dog that failed to bark, as noted by Arthur Conan Doyle and mentioned by Ross (1978) and Fischhoff and Beyth-Marom (1983), far from being supported only by anecdotes, can be robustly observed in abstract laboratory tasks with fully explicit probabilistic information. The tendency is not appreciably or systematically weakened in those contexts in which non-occurrences are rare in comparison to occurrences, neither is it by the overt display of the probabilities of non-occurrences. Furthermore, this feature-positive effect influences confidence towards judgments: On average, participants trusted judgments based on occurrences more than those based on non-occurrences (although this effect was undermined by the overt presentation of the probabilities of non-occurrence, see Figure 4). Finally, participants showed a remarkable sensitivity to the amount of information conveyed by the occurrence of stimuli (as shown by a positive correlation between the ΔI/probability gain of and the number of presence-consistent choices), but they were either insensitive to the amount of information conveyed by non-occurrences (Study 1) or their sensitivity did not actually help them to consider absent clues in their eventual decisions (Studies 2-3).

The feature-positive effect in hypothesis evaluation might have important consequences for confirmation biases (e.g., Klayman, 1995; McKenzie, 2004, 2006). The most common information-gathering testing strategy is positive testing, consisting of the search for clues whose occurrence is consistent with the hypothesis under examination. That is, when a hypothesis is tested positively, occurrences confirm it, whereas non-occurrences confute it. This information-gathering strategy, if coupled with the tendency to overrate occurrences in comparison to non-occurrences, would give rise to a systematic tendency to improperly confirm the tested hypothesis (e.g., Klayman, 1995; McKenzie, 2004, 2006). This type of confirmation bias might have relevant implications in circumstances in which people cannot resort to previous knowledge about the relationships among cues in the real world (situations that were exemplified in our experiments), for example when evaluating technical reports (e.g., statistical write-ups), possibly contributing to inappropriate evaluations of the acquired data and ultimately to inefficient decisions.

This type of confirmation bias might also have important, detrimental side effects in contexts in which the rigorous testing of hypotheses is of critical importance, such as in scientific research, forensic practice, medical diagnosis (e.g., Christensen-Szalansky and Bushyhead's 1981; Scandellari, 2005) and health behaviors (e.g., Rassin, Muris, Franken, & van Straten, 2008). It might also have relevant consequences in the social domain, where feature-positive effects have been proven to occur (e.g., Fazio, Sherman, & Herr, 1982) and other types of confirmatory tendencies toward stereotypes are already known (e.g., Fiedler & Walther, 2004). However, because the present experiments used abstract problems only, estimating the impact of the feature-positive effect on the evaluation of hypotheses in practical domains will require further investigation. In this sense, it has yet to be clarified whether and under what circumstances the feature-positive effect might diminish or even reverse to a feature-negative effect (FNE, Fiedler, Eckert, & Poysiak, 1989) when evaluating competing hypotheses.

More generally, it has to be noted that although research on judgment and decision making has pointed out people's difficulty to adhere to Bayesian principles in explicit tasks, recent studies have offered Bayesian accounts of more implicit processes suggesting sophisticated abilities of adult, children, and infants (e.g., Griffiths & Tenenbaum, 2006; Téglás, Vul, Girotto, Gonzalez, Tenenbaum, & Bonatti, 2011; Tenenbaum, Griffiths, & Kemp, 2006). Accordingly, future studies should elucidate whether and in which kind of hypothesis-testing tasks people might exhibit a Bayesian-like weighing of both present and absent features.

References

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes,42*, 88–110. doi:10.1016/0749-5978(88)90021-0

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory & Cognition, 10*, 511-519.

Beyth-Marom, R. & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology, 45*, 1185–1195. doi:10.1037//0022-3514.45.6.1185

Blom, E., & Baayen, H. R. (in press). The impact of verb form, sentence position, home language and L2 proficiency on subject-verb agreement in child L2 Dutch. *Applied Psycholinguistics*.

Bourne, L. E. Jr. & Guy, D. E. (1968). Learning conceptual rules. II: The role of positive and negative instances. *Journal of Experimental Psychology, 77*, 488–494. doi:10.1037/h0025952

Cherubini, P., Rusconi, P., Russo, S., Di Bari, S., & Sacchi, S. (2010). Preferences for different questions when testing hypotheses in an abstract task: Positivity does play a role, asymmetry does not. *Acta Psychologica*, *134,* 162–174. doi:10.1016/j.actpsy.2010.01.007

Cherubini, P., Russo, S., Rusconi, P., D'Addario, M., & Boccuti, I. (2009). *Il ragionamento probabilistico nella diagnosi medica: sensibilità e insensibilità alle informazioni*. In P. Giaretta, A. Moretto, G. F. Gensini, & M. Trabucchi (Eds.), *Filosofia della medicina: Metodo, modelli, cura ed errori* (pp. 541-564). Bologna: Il Mulino.

Christensen-Szalanski, J. J. J., & Bushyhead, J. B. (1981). Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 928–935. doi:10.1037//0096-1523.7.4.928

Cohen, L. J. (1977). *The probable and the provable*. Oxford: Clarendon Press.

Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian theories of evidential support: Normative and descriptive considerations. *Philosophy of Science, 74,* 229–252.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83*, 37-64. doi:10.1037//0033-295X.83.1.37

Fazio, R. H., Sherman, S. J., & Herr P. M. (1982). The feature-positive effect in the self-perception process: Does not doing matter as much as doing? *Journal of Personality and Social Psychology, 42*, 404–411. doi:10.1037/0022-3514.42.3.404

Feeney, A., Evans, J. & Venn, S. (2008). Rarity, pseudodiagnosticity and Bayesian reasoning. *Thinking & Reasoning, 14,* 209–230. doi:10.1080/13546780801934549

Feeney, A., Evans, J. St. B. T., & Clibbens, J. (2000). Background beliefs and evidence interpretation. *Thinking and Reasoning*, 6, 97-124.Fiedler, K, Eckert, C., & Poysiak, C. (1989). Asymmetry in human discrimination learning: Feature positive effect or focus of hypothesis effect? *Acta Psychologica, 70*, 109–127. doi:10.1016/0001-6918(89)90015-2

Fiedler, K., Eckert, C., & Poysiak, C. (1989). Asymmetry in human discrimination learning: Feature positive effect or focus of hypothesis effect? *Acta Psychologica, 70*, 109–127. doi:10.1016/0001-6918(89)90015-2

Fiedler, K., & Walther, E. (2004). *Stereotyping as inductive hypothesis testing*. Hove, U.K.: Psychology Press.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90*, 239–260. doi:10.1037//0033-295X.90.3.239

Green, D. W., & Over, D. E. (2000). Decision theoretic effects in testing a causal conditional. *Current Psychology of Cognition*, 19, 51–68.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science,17*, 767–773. doi:10.1111/j.1467-9280.2006.01780.x

Hearst, E., & Wolff, W. T. (1989). Addition versus deletion as a signal. *Animal Learning & Behavior, 17*, 120-133.

Hovland, C. I., & Weiss, W. (1953). Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology, 45*, 175-182. doi:10.1037/h0062351

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446. doi:10.1016/j.jml.2007.11.007

Jenkins, H. M., & Sainsbury, R. S. (1969). The development of stimulus control through differential reinforcement. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 123-161). Halifax: Dalhousie University Press.

Jenkins, H. M., & Sainsbury, R. S. (1970). Discrimination learning with the distinctive feature on positive or negative trials. In D. Mostofsky (Ed.), *Attention: Contemporary theory and analysis.* New York: Appleton-Century-Crofts.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79*, 1-17.

Kao, S-F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1363-1386. doi:10.1037//0278-7393.19.6.1363

Klayman, J. (1995). Varieties of confirmation bias. *The Psychology of Learning and Motivation, 32*, 385-418. doi: 10.1016/S0079-7421(08)60315-1

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228. doi: 10.1037/0033-295X.94.2.211

Loftus, E. F. (1976). Unconscious transference in eyewitness identification. *Law and Psychology Review, 2*, 93-98.

Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General, 127*, 269-285. doi:10.1037//0096-3445.127.3.269

McKenzie, C. R. M. (2004). Hypothesis testing and evaluation. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 200-219). Malden, MA, US: Blackwell Publishing.

McKenzie, C. R. M. (2006). Increased sensitivity to differentially diagnostic answers using familiar materials: Implications for confirmation bias. *Memory & Cognition, 34*, 577–588. doi:10.3758/BF03193581

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin & Review, 7*, 360-366.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology, 54,* 33-61. doi:10.1016/j.cogpsych.2006.04.004

Nahinsky, I. D., & Slaymaker, F. L. (1970). Use of negative instances in conjunctive concept identification. *Journal of Experimental Psychology, 84*, 64-68. doi:10.1037/h0028951

Neisser, U. (1963). Decision-time without reaction-time: Experiments in visual scanning. *The American Journal of Psychology, 76*, 376−385. doi:10.2307/1419778

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review, 112*, 979–999. doi: 10.1037/0033-295X.112.4.979

Nelson, J. D. (2008). Towards a rational theory of human information acquisition. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 143-163). Oxford, UK: Oxford University Press.

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science, 21*, 960–969. doi:10.1177/0956797610372637

Newman, J., Wolff, W. T., & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 630-650. doi:10.1037//0278-7393.6.5.630

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220. doi: 10.1037/1089-2680.2.2.175

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608-631. doi:10.1037//0033-295X.101.4.608

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review and re-evaluation. *Psychonomic Bulletin & Review*, *10,* 289–318.

Rassin, E., Muris, P, Franken, I., & van Straten, M. (2008). The feature-positive effect and hypochondriacal concerns. *Behaviour Research and Therapy, 46*, 263–269. doi:10.1016/j.brat.2007.11.003

Ross, L. (1978). The intuitive psychologist and his shortcomings: distortions in the attribution process. In L. Berkowitz (Ed.), *Cognitive theories in social psychology: papers from advances in experimental social psychology*. New York: Academic Press.

Scandellari, C. (2005). *La diagnosi clinica: Principi metodologici del procedimento decisionale*. Milano: Masson.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379−423, 623−656.

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22*, 93−121. doi:10.1016/0022-1031(86)90031-4

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6*, 649−744. doi:10.1016/0030-5073(71)90033-X

Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory & Cognition, 20*, 392−405.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4*, 165−173. doi:10.1111/j.1467-9450.1963.tb01324.x

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36*, 1202−1212. doi:10.1037/0022-3514.36.11.1202

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332*, 1054−1059. doi:10.1126/science.1196404

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*, 309−318. doi:10.1016/j.tics.2006.05.009

Treisman, A. & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*, 285−310. doi:10.1037//0096-3445.114.3.285

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, *12*, 129-140. doi:10.1080/17470216008416717

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88*, 776-784. doi:10.1037//0033-2909.88.3.776

Table 1

Formal properties of the 18 problems used in the three experiments (the conditional probabilities of the present clues are in boldface).

Problem	Deck	Prior probabilities	Likelihoods					Posterior probabilities
			p(B)	p(C)	p(D)	p(F)	p(G)	
1	1	.5	**.43**	**.8**	.89	.93		.8
	2	.5	**.01**	**.08**	.1	.1		.2
2	1	.5	**.03**	**.29**	.35	.65	.25	.5
	2	.5	**.3**	**.2**	.9	.35	.62	.5
3	1	.5	**.85**	**.8**	**.95**	.95	.96	.23
	2	.5	**.04**	**.44**	**.3**	.1	.1	.77
4	1	.5	**.35**	**.2**	.14	.4		.5
	2	.5	**.1**	**.98**	.39	.4		.5
5	1	.5	**.01**	**.11**	**.8**	.3	.2	.21
	2	.5	**.8**	**.75**	**.76**	.96	.9	.79
6	1	.5	**.9**	**.7**	.95	.96		.26
	2	.5	**.02**	**.22**	.1	.1		.74
7	1	.5	**.9**	**.7**	.9	.9	.9	.27
	2	.5	**.02**	**.22**	.2	.4	.2	.73
8	1	.5	**.5**	**.7**	**.3**	.5	.35	.5
	2	.5	**.09**	**.88**	**.97**	.4	.26	.5
9	1	.5	**.02**	**.16**	.1	.1		.76
	2	.5	**.5**	**.7**	.94	.96		.24
10	1	.5	**.8**	**.2**	.15	.45	.85	.5
	2	.5	**.07**	**.68**	.4	.35	.4	.5
11	1	.5	**.02**	**.16**	**.5**	.1	.05	.74
	2	.5	**.7**	**.6**	**.45**	.95	.95	.26
12	1	.5	**.09**	**.88**	.85	.16		.5
	2	.5	**.1**	**.5**	.2	.75		.5
13	1	.5	**.02**	**.16**	.2	.12	.1	.79
	2	.5	**.5**	**.7**	.9	.9	.85	.21
14	1	.5	**.09**	**.88**	**.97**	.3	.76	.5
	2	.5	**.83**	**.22**	**.55**	.8	.35	.5
15	1	.5	**.01**	**.11**	.15	.1	.1	.27
	2	.5	**.85**	**.6**	.8	.8	.9	.73
16	1	.5	**.85**	**.65**	**.65**	.89	.94	.74
	2	.5	**.02**	**.16**	**.3**	.1	.05	.26
17	1	.5	**.01**	**.11**	.16	.15		.3
	2	.5	**.75**	**.5**	.9	.95		.7
18	1	.5	**.8**	**.5**	.6	.7	.96	.82
	2	.5	**.01**	**.08**	.3	.15	.1	.18

Table 2

The properties of the 18 problems used in the three experiments: Deck favored by the consideration of the present-clue subset, absent-clue subset or all clues (i.e., correct), and utility values of the present clues, absent clues, and all clues according to four optimal-experimental-design models (I.G. = Information Gain; K.-L. = Kullback-Leibler distance; P.G. = Probability Gain).

Problem	Correct response	Response suggested by present clues	Response suggested by absent clues	I.G./K.-L. of all clues	I.G./K.-L of present clues	I.G./K.-L of absent clues	P.G./Impact of all clues	P.G./Impact of present clues	P.G./Impact of absent clues
1	Deck 1	Deck 1	Deck 2	.28	.98	.92	.3	.5	.49
2	equiprobable	Deck 2	Deck 1	0	.45	.45	0	.37	.37
3	Deck 2	Deck 1	Deck 2	.22	.93	.98	.27	.49	.5
4	equiprobable	Deck 2	Deck 1	0	.02	.02	0	.08	.09
5	Deck 2	Deck 2	Deck 1	.26	.98	.94	.29	.5	.49
6	Deck 2	Deck 1	Deck 2	.17	.94	.98	.24	.49	.5
7	Deck 2	Deck 1	Deck 2	.16	.94	.97	.23	.49	.5
8	equiprobable	Deck 1	Deck 2	0	.02	.02	0	.08	.08
9	Deck 1	Deck 2	Deck 1	.2	.93	.97	.26	.49	.5
10	equiprobable	Deck 1	Deck 2	0	.22	.22	0	.27	.27
11	Deck 1	Deck 2	Deck 1	.17	.93	.97	.24	.49	.5
12	equiprobable	Deck 1	Deck 2	0	.04	.04	0	.11	.11
13	Deck 1	Deck 2	Deck 1	.26	.93	.98	.29	.49	.5
14	equiprobable	Deck 2	Deck 1	0	.01	.01	0	.07	.06
15	Deck 2	Deck 2	Deck 1	.16	.98	.95	.23	.5	.49
16	Deck 1	Deck 1	Deck 2	.17	.97	.94	.24	.5	.49
17	Deck 2	Deck 2	Deck 1	.12	.97	.94	.2	.5	.49
18	Deck 1	Deck 1	Deck 2	.32	.98	.93	.32	.5	.49

Table 3

*Percentages (standard errors of the means in parentheses) of each type of choice in each type of problem in Study 1. There were two problems in each experimental cell, and thus the frequency of responses ranged from 0 to 2. Percentages were computed out of 84 total responses (because of rounding, some row totals do not exactly equal 100 for percentages). Correct responses are in boldface. The p*s *of the binomial tests comparing actual answers to a chance level of 33% are reported as "*", meaning Bonferroni adjusted* $p < .05$, *"**", meaning Bonferroni adjusted* $p \leq .01$, *or "***", meaning Bonferroni adjusted* $p \leq .001$.

		Responses		
	Present-to-absent ratio	Presence-consistent	Absence-consistent	Equiprobable
Presence consistent Problems	2:2	**75% (.1) *****	17% (.09) ***	8% (.06) ***
	3:2	**56% (.11) *****	32% (.11)	12% (.07) ***
	2:3	**65% (.12) *****	20% (.1) **	14% (.09) ***
Absence consistent Problems	2:2	57% (.13) ***	**30% (.11)**	13% (.09) ***
	3:2	51% (.12) ***	**38% (.12)**	11% (.06) ***
	2:3	60% (.12) ***	**30% (.11)**	11% (.08) ***
Equiprobable problems	2:2	45% (.12) *	31% (.12)	**24% (.1)**
	3:2	38% (.12)	27% (.11)	**35% (.12)**
	2:3	56% (.11) ***	20% (.08) **	**24% (.1)**

Table 4

*Percentages (standard errors of the means in parentheses) of each type of choice in each type of problem in Study 2. There were 18 problems (2 per cell), N = 42. The stars report the level of significance against chance level (set at .33): * = Bonferroni adjusted p < .05; ** = Bonferroni adjusted p ≤ .01; *** = Bonferroni adjusted p ≤ .001. Correct responses are in boldface.*

		Responses		
	Present-to-absent ratio	Presence-consistent	Absence-consistent	Equiprobable
Presence consistent Problems	2:2	**68% (.11) *** **	15% (.08) ***	17% (.09) ***
	3:2	**69% (.11) *** **	13% (.08) ***	18% (.1) **
	2:3	**67% (.12) *** **	20% (.1) **	13% (.08) ***
Absence consistent Problems	2:2	52% (.14) ***	**23% (.1) * **	25% (.11)
	3:2	63% (.12) ***	**21% (.09) * **	15% (.08) ***
	2:3	62% (.13) ***	**27% (.12)**	11% (.07) ***
Equiprobable problems	2:2	43% (.13) *	21% (.1) *	**36% (.12)**
	3:2	24% (.11)	46% (.12) **	**30% (.11)**
	2:3	54% (.1) ***	32% (.1)	**14% (.09) *** **

Table 5

*Percentages (standard errors of the means in parentheses) of each type of choice in each type of problem in Study 3. There were 18 problems (2 per cell), N = 42. The stars report the level of significance against chance level (set at .33): * = Bonferroni adjusted p < .05; ** = Bonferroni adjusted p ≤ .01; *** = Bonferroni adjusted p < .001. Correct responses are in boldface.*

		Responses		
	Present-to-absent ratio	Presence-consistent	Absence-consistent	Equiprobable
Presence consistent Problems	2:2	**65% (.12) *****	26% (.11)	8% (.06) ***
	3:2	**62% (.11) *****	18% (.1) **	20% (.1) **
	2:3	**52% (.14) *****	26% (.12)	21% (.11) *
Absence consistent Problems	2:2	52% (.14) ***	**25% (.11)**	21% (.11) *
	3:2	61% (.13) ***	**26% (.11)**	13% (.08) ***
	2:3	57% (.13) ***	**25% (.12)**	18% (.1) **
Equiprobable problems	2:2	35% (.12)	38% (.11)	**27% (.1)**
	3:2	27% (.11)	37% (.12)	**36% (.12)**
	2:3	38% (.12)	24% (.1)	**37% (.12)**

Table 6

*Percentages (standard errors of the means in parentheses) of each type of choice in each type of problem in the three studies. There were 18 problems (2 per cell), N = 126. The stars report the level of significance against chance level (set at .33): * = Bonferroni adjusted p ≤ .05; ** = Bonferroni adjusted p < .01; *** = Bonferroni adjusted p < .001. Correct responses are in boldface.*

		Responses		
	Present-to-absent ratio	Presence-consistent	Absence-consistent	Equiprobable
Presence consistent Problems	2:2	**69% (.06) *** **	19% (.06) ***	11% (.04) ***
	3:2	**62% (.06) *** **	21% (.06) ***	17% (.05) ***
	2:3	**62% (.07) *** **	22% (.06)***	16% (.05) ***
Absence consistent Problems	2:2	54% (.08) **** **	**26% (.06) ** **	20% (.06) ***
	3:2	58% (.07) **** **	**29% (.06) * **	13% (.04) ***
	2:3	60% (.07) **** **	**27% (.07)* **	13% (.05) ***
Equiprobable problems	2:2	41% (.07) **	30% (.06)	**29% (.06)**
	3:2	30% (.07)	37% (.07)	**33% (.07)**
	2:3	49% (.06) ***	25% (.05) **	**25% (.06) ** **

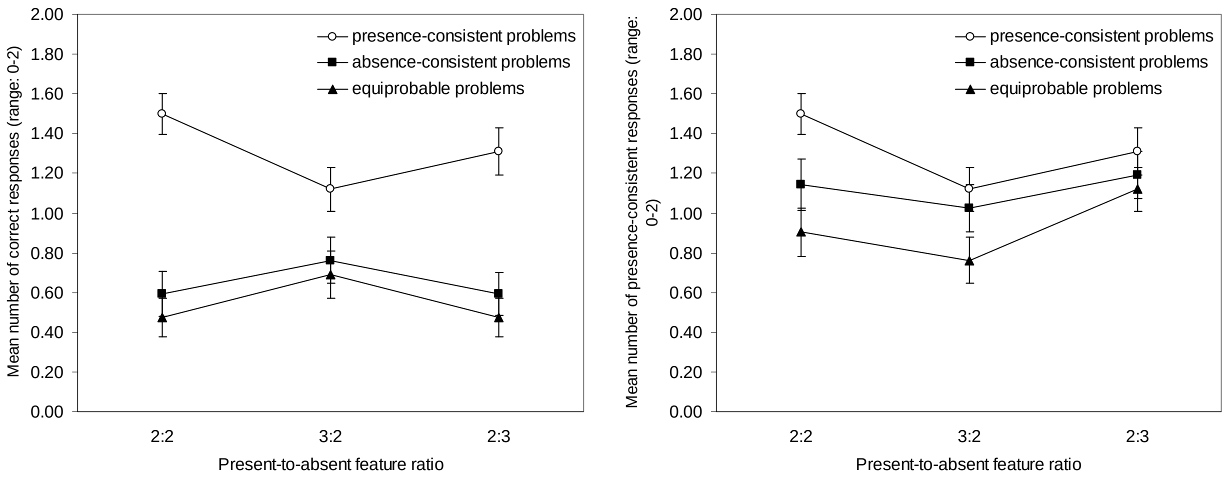


Figure 1. Left panel. The mean number of correct responses as a function of the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3) in Study 1. Right panel. The mean number of presence-consistent responses in Study 1 as a function of the type of problem, that is, the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3). The error bars represent the standard errors of the means (SEMs).

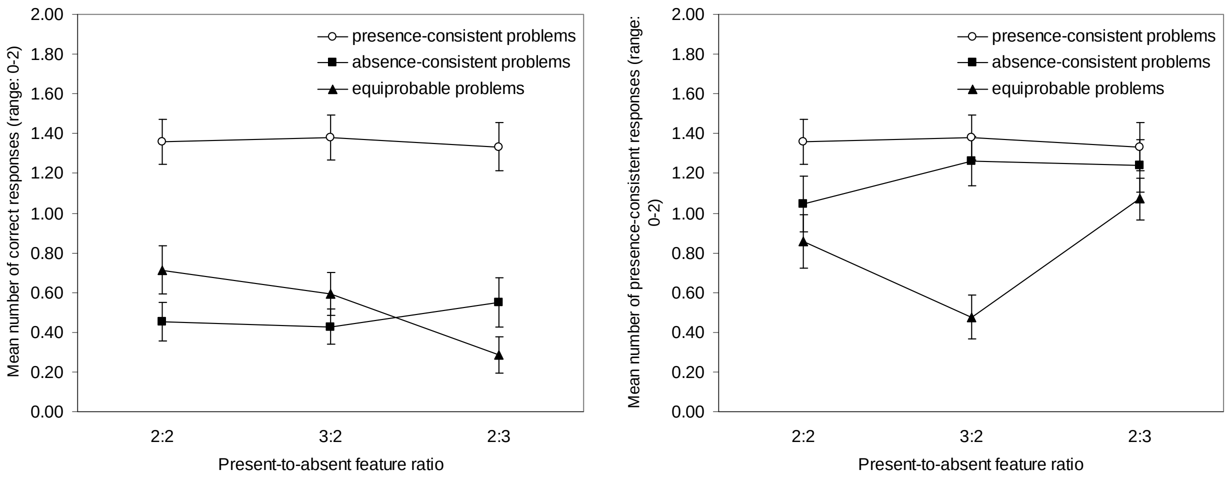


Figure 2. Left panel. The mean number of correct responses as a function of the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3) in Study 2. Right panel. The mean number of presence-consistent responses in Study 2 as a function of the type of problem, that is, the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3). The error bars represent the standard errors of the means (SEMs).

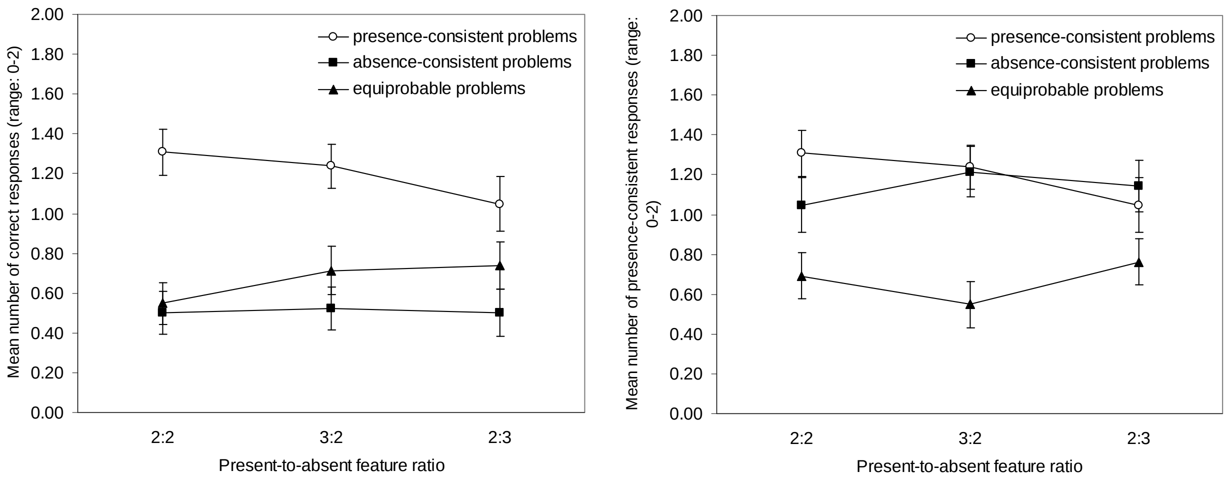


Figure 3. Left panel. The mean number of correct responses as a function of the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3) in Study 3. Right panel. The mean number of presence-consistent responses in Study 3 as a function of the type of problem, that is, the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3). The error bars represent the standard errors of the means (SEMs).

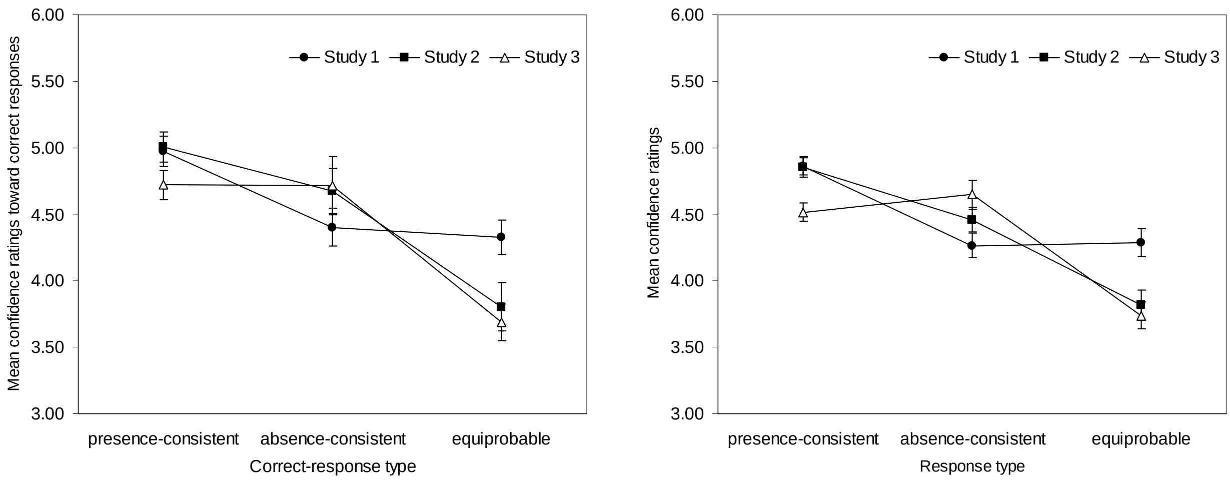


Figure 4. Left panel. The mean confidence ratings toward the correct responses in the three studies when the normative response was presence-consistent, absence-consistent, or equiprobable. Right panel. Participants' mean confidence in each response type they gave (i.e., presence-consistent, absence-consistent, and equiprobable responses) in the three studies. The error bars represent the standard errors of the means (SEMs).

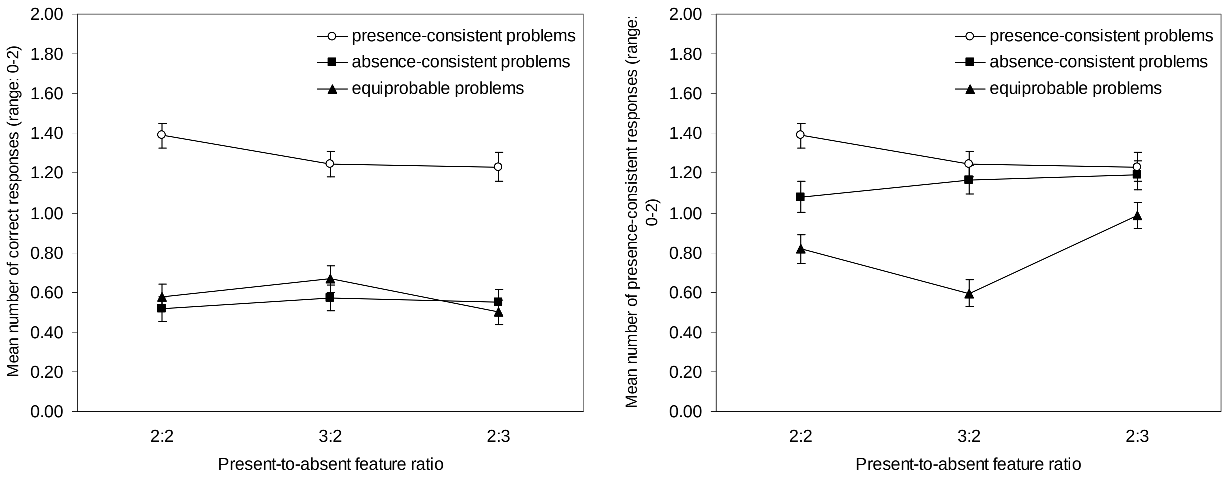


Figure 5. Left panel. The mean number of correct responses across the three studies as a function of the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3). Right panel. The mean number of presence-consistent responses across the three studies as a function of the type of problem, that is, the correct-response type (i.e., presence-consistent, absence-consistent, and equiprobable problems) and of the present-to-absent feature ratio (i.e., 2:2, 3:2; 2:3). The error bars represent the standard errors of the means (SEMs).

Appendix

Instructions given to participants in the three studies (we report both the original versions in Italian and the English translations).

ISTRUZIONI

In queste pagine troverai 18 problemi riguardanti due mazzi composti da 100 carte ciascuno. In alcuni problemi, su ogni carta di ciascun mazzo sono riportate da **0** a **4** lettere consonanti (scelte tra B, C, D, F). In altri problemi ogni carta contiene da **0** a **5** lettere consonanti (scelte tra B, C, D, F, G).

Study 1 version:

In ogni problema una tabella descrive quante carte in ciascun mazzo riportano una data lettera. Il numero di carte che riportano una lettera è del tutto **indipendente** dal numero di carte che riportano ogni altra lettera.
Ad esempio, una tabella come:

	B	C	D	**F**
mazzo 1	46	21	9	38
mazzo 2	12	88	72	56

sta a indicare che nel mazzo 1, 46 carte contengono la lettera B, 21 la C, 9 la D, 38 la F. Nel mazzo 2, 12 carte contengono la B, 88 la C, 72 la D, e 56 la F.

Study 2 version:

In ogni problema una tabella descrive quante carte in ciascun mazzo riportano una data lettera e quante invece non la riportano. Il numero di carte che riportano o meno una lettera è del tutto **indipendente** dal numero di carte che riportano, o meno, ogni altra lettera.
Ad esempio, una tabella come:

	B		C		D		**F**	
	sì	no	sì	no	sì	no	sì	no
mazzo 1	46	54	21	79	9	91	38	62
mazzo 2	12	88	88	12	72	28	56	44

sta a indicare che nel mazzo 1: 46 carte contengono la lettera B e 54 non la contengono, 21 carte riportano la C e 79 non la riportano, 9 carte contengono la D e 91 non la contengono, 38 carte contengono la F e 62 non la contengono. Inoltre, nel mazzo 2: 12 carte contengono la B e 88 non la contengono, 88 carte contengono la C e 12 non la contengono, 72 carte riportano la D e 28 non la riportano, 56 carte contengono la F e 44 non la contengono.

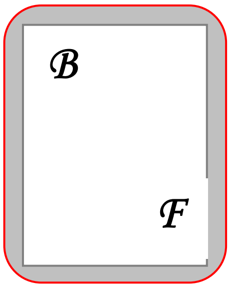
Study 3 version:

In ogni problema una tabella descrive quante carte in ciascun mazzo **non** riportano una data lettera. Il numero di carte che non riportano una lettera è del tutto **indipendente** dal numero di carte che non riportano ogni altra lettera.
Ad esempio, una tabella come:

	B	C	D	**F**
mazzo 1	54	79	91	62
mazzo 2	88	12	28	44

sta a indicare che nel mazzo 1, 54 carte non contengono la lettera B, 79 non riportano la C, 91 non contengono la D, 62 non riportano la F. Nel mazzo 2, 88 carte non contengono la B, 12 non riportano la C, 28 non contengono la D, e 44 non riportano la F.

Immagina che lo sperimentatore scelga a caso il mazzo da cui estrarre, a caso, **una sola** carta, e non ti dica di quale mazzo si tratti. Ti comunica, però, se e quali lettere vi siano sulla carta estratta. Ad esempio, riferendosi ai mazzi 1 e 2 della tabella qui sopra, potrebbe dirti: "la carta che ho estratto ha una **B** e una **F**, ma non ha né la C né la D". Una raffigurazione della carta, in questo esempio, potrebbe essere di questo tipo:



Naturalmente, in base alle poche informazioni a disposizione non puoi stabilire con certezza da quale dei due mazzi sia stata estratta la carta. Tuttavia, ti viene richiesto di stabilire se, alla luce del suo contenuto, la carta è più probabilmente del mazzo 1 o del mazzo 2. Se in alcuni problemi non riesci a decidere di quale mazzo sia probabilmente la carta, rispondi che i due mazzi sono egualmente probabili.
Nell'esempio appena riportato, la risposta corretta è la seguente: "è più probabile che la carta provenga dal mazzo 1".

In dettaglio, per ogni problema:

- se ritieni che sia più probabile che la carta provenga dal **mazzo 1**, segna con una croce la casella posta sotto la tabella con scritto "**mazzo 1**";
- se ritieni che sia più probabile che la carta provenga dal **mazzo 2**, segna con una croce la casella posta sotto la tabella con scritto "**mazzo 2**";
- se ritieni che la probabilità di estrarre la carta dai due mazzi sia uguale, segna con una croce la casella posta sotto la tabella con scritto "**equiprobabili**".

Dopo aver risposto, indica quanto ti fidi che la tua risposta sia corretta, segnando con una croce un numero sulla scala numerata da 1 a 7 (1= poca fiducia; 7 = molta fiducia) che trovi alla fine di ciascun problema.

INSTUCTIONS

In the following pages, you will be presented with 18 problems concerning two decks of cards. Each deck consists of 100 cards. In some problems, each card of each deck shows between **0** and **4** consonants (chosen among B, C, D, F). In other problems, each card shows between **0** and **5** consonants (chosen among B, C, D, F, G)

Study 1 version:

For each problem, a table describes how many cards in each deck have a given letter printed on their face. The number of cards showing a letter is totally **independent** of the number of cards showing any other letter.
For example, a table as the following:

	B	C	D	**F**
mazzo 1	46	21	9	38
mazzo 2	12	88	72	56

indicates that in deck 1, 46 cards have the letter B, 21 the letter C, 9 the letter D, 38 the letter F. In deck 2, 12 cards have the letter B, 88 the letter C, 72 the letter D, and 56 the letter F.

Study 2 version:

For each problem, a table describes how many cards in each deck have a given letter printed on their face and how many cards do not. The number of cards showing or not showing a letter is totally **independent** of the number of cards showing, or not showing, any other letter.
For example, a table as the following:

	B		C		D		**F**	
	sì	no	sì	no	sì	no	sì	no
mazzo 1	46	54	21	79	9	91	38	62
mazzo 2	12	88	88	12	72	28	56	44

indicates that in deck 1: 46 cards have the letter B and 54 do not, 21 cards have the letter C and 79 do not, 9 cards have the letter D and 91 do not, 38 cards have the letter F and 62 do not. Furthermore, in deck 2: 12 cards have the letter B and 88 do not, 88 cards have the letter C and 12 do not, 72 cards have the letter D and 28 do not, 56 cards have the letter F and 44 do not.

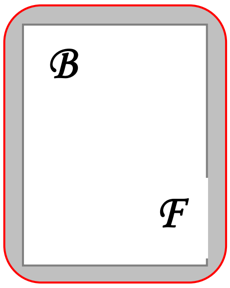
Study 3 version:

For each problem, a table describes how many cards in each deck **do not** have a given letter printed on their face. The number of cards that do not show a letter is totally **independent** of the number of cards that do not show any other letter.
For example, a table as the following:

	B	C	D	**F**
mazzo 1	54	79	91	62
mazzo 2	88	12	28	44

indicates that in deck 1, 54 cards do not have the letter B, 79 do not have the C, 91 do not have the D, 62 do not have the F. In deck 2, 88 cards do not have the B, 12 do not have the C, 28 do not have the D, and 44 do not have the F.

Imagine that the experimenter randomly chooses the deck from which to draw, at random, **only one** card, without telling you which deck she selected. However, she tells you which letters, if any, are printed on the drawn card. For example, with reference to the decks 1 and 2 of the table above, she could tell you: "the card I drew has a **B** and an **F**, but it has neither a C nor a D". A picture of the card, in this example, could be as the following:



Obviously, based on the limited information available you cannot determine with certainty from which of the two decks the card was drawn. However, based on the content of the card, you are requested to determine whether the card belongs more likely to deck 1 or to deck 2. If in some problems you cannot decide from which deck the card was more likely drawn, you will answer that the two decks are equiprobable.
In the example above, the correct answer is the following: "it is more probable that the card has been drawn from deck 1".

In detail, in each problem:
- if you deem that it is more likely that the card has been drawn from **deck 1**, mark a cross in the box "**deck 1**" located under the table.
- if you deem that it is more likely that the card has been drawn from **deck 2**, mark a cross in the box "**deck 2**" located under the table.
- if you deem that it is equally likely that the card has been drawn from deck 1 and deck 2, mark a cross in the box "**equiprobable**" located under the table.

After making your decision, indicate how much you are confident in the accuracy of your answer. Mark a cross on a 1-to-7 rating scale (1 = not confident; 7 = very confident), which you find at the end of each problem.