

This article is Copyright © 2012, Canadian Psychological Association. The published article is available as “Online First” at the CPA journal website:

<http://psycnet.apa.org/?&fa=main.doiLanding&doi=10.1037/a0028173>

This article may not exactly replicate the final version published in the CPA journal. It is not the copy of record.

The final publication is available at

<http://psycnet.apa.org/index.cfm?fa=browsePA.volumes&jcode=cep>

Running head: THE FEATURE-POSITIVE EFFECT IN HYPOTHESIS EVALUATION

Moderators of the Feature-Positive Effect in Abstract Hypothesis-Evaluation Tasks

Patrice Rusconi^a, Franca Crippa^a, Selena Russo^b, Paolo Cherubini^a

^aUniversity of Milano-Bicocca

^bUniversity of Trento

Author Note

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

The authors would like to thank an anonymous reviewer for pointing out the Treisman and Souther's (1985) article as well as for insightful criticisms on an earlier version of Cherubini et al. (in press).

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 6448 3736. Fax: ++39 02 6448 3706. E-mail: patrice.rusconi1@unimib.it

Abstract

Three studies using abstract materials tested possible moderators of the feature-positive effect in hypothesis evaluation whereby people use the presence of features more than their absence to judge which of two competing hypotheses is more likely. Drawing on a distinction made in visual perception research, we tested whether the feature-positive effect emerges both when using non-substitutive features, which can be removed without replacement by other features, and substitutive features, the absence of which implies the presence of other features (e.g., the color red, the absence of which entails the presence of another color). Furthermore, we tested whether presenting to participants both the clue occurrence probabilities (which are needed to consider clue presence) and their complements (which are needed to gauge the impact of the absent clues) decreased the feature-positive effect. The results showed that regardless of the type of feature (i.e., non-substitutive vs. substitutive), participants provided more responses consistent with an evaluation of the subset of present clues compared to all other kinds of responses. However, the use of substitutive features combined with an explicit presentation format of probabilistic information had a debiasing effect. Furthermore, the use of substitutive features negated participant sensitivity to the rarity of clues, whereby the feature-positive effect decreased when there was one absent clue and two present clues for problems in which the exclusive consideration of the presence of features did not suggest the correct response.

Keywords: hypothesis testing; the feature-positive effect; substitutive features; the rarity effect

Moderators of the Feature-Positive Effect in Abstract Hypothesis-Evaluation Tasks

The feature-positive effect (FPE) refers to the tendency of human and non-human animals to search for or use occurrences (rather than the non-occurrences) as cues to accomplish a task. Jenkins and Sainsbury first introduced this term to describe the asymmetry in pigeon performance when learning perceptual discrimination. Pigeons learned to discriminate between food-reinforced trials (positive trials) and non-reinforced trials (negative trials) faster when a black dot (i.e., the feature) appeared on the response key for positive trials and when it was absent on negative trials rather than the reverse (Jenkins & Sainsbury, 1969, 1970). This phenomenon has been described in similar forms across a variety of species (including humans), ages, experimental settings, and cognitive processes (e.g., Bourne & Guy, 1968; Estes, 1976; Fazio, Sherman, & Herr, 1982; Jenkins & Ward, 1965; Mandel & Lehman, 1998; Nahinsky & Slaymaker, 1970; Neisser, 1963; Newman, Wolff, & Hearst, 1980; Wells & Lindsay, 1980).

The factors that undermine the FPE have been investigated rarely. For example, in a study of human visual perception, Treisman and Souther (1985) compared search times for the presence versus the absence of a target in a display of distractors. In their first experiment, participants detected a circle with an intersecting line in a display of circles without such lines (i.e., the presence condition) faster than a circle without an intersecting line among circles with these lines (i.e., the absence condition). The lines are “features that can be removed without being replaced by others” (Treisman & Souther, 1985); that is, they are non-substitutive features because they do not need to be replaced by other features when absent. In their second experiment, they compared searches for presence and absence using color to distinguish between targets. Specifically, participants searched for a green target among red and black distractors (presence condition) or looked for a non-green target (which thus was either red or black) among green distractors (absence condition). Color is a “substitutive feature” (Treisman & Souther, 1985); that is, the absence of a color implies the presence of another color. Present and absent features should be equally distinctive and detectable using this method. In fact, the results did not reveal an asymmetry between participant latencies in

the presence condition compared with the absence condition. Treisman and Souther (1985) suggested that the absence of the substitutive feature (green) was spontaneously recoded as the presence of another (red or black).

The present article is concerned with whether the use of substitutive features moderates the FPE in human reasoning, specifically in belief revision. We focused on the evaluation stage of hypothesis development, wherein people interpret available data to determine whether they confirm, disconfirm, or whether they are irrelevant to the hypothesis being tested (e.g., Klayman, 1995; McKenzie, 2004, 2006). This issue is relevant not only for theoretical reasons but also for potential applications. Given that the FPE with regard to hypothesis evaluation could have detrimental results (e.g., when judges consider circumstantial evidence as more convincing than the absence of evidence), reducing this phenomenon might enhance efficient judgment and decision making.

Recently, Cherubini and his colleagues devised probabilistic tasks to examine the FPE with regard to human hypothesis evaluation directly (Cherubini, Rusconi, Russo, & Crippa, in press; Rusconi, Marelli, Russo, D'Addario, & Cherubini, in press). Specifically, participants determined from which of two card decks the experimenter drew a card (Cherubini et al., in press; Rusconi et al., in press). Each card featured between 0 and 5 letters. Participants were informed which letters were printed on the drawn card and which were not. A table showed the number of cards that depicted all possible letters in each deck. Participants were also informed that the experimenter had selected the deck randomly. The absence of letters was conveyed to participants by denying their presence, that is, via negative instructions such as "There is neither a D nor an F on the drawn card" and by showing to participants that they were not printed on the drawn card. The authors also manipulated the letter distribution presentation format in the cards of the two decks by presenting participants with the distributions of letter occurrences, those of letter non-occurrences, or both. The findings revealed a robust FPE regardless of the letter distribution presentation format.

The method that Cherubini et al. (in press) and Rusconi et al. (in press) used to convey the absence of features might have caused an imbalance such that present and absent clues were

unequally distinctive and detectable. Letters, unlike colors, are not substitutive features: other features do not necessarily replace them when absent (e.g., a card might reveal a blank in place of a letter). Treisman and Souther (1985, Experiment 1) showed that when non-substitutive features are used, there is a presence-absence asymmetry in search latencies. Accordingly, the use of negative sentences (e.g., “There is neither a D nor an F on the drawn card”) and non-substitutive features to convey an absence in the studies by Cherubini and colleagues might have directed participant attention to the present clues, thereby artificially decreasing the value of the absent clues. This hypothesis rests on two findings. First, people have greater difficulty handling negative relative to positive sentences (e.g., Hearst, 1991; Van Wallendael, 1995; Wason, 1959, 1961). Therefore, the negative instructions used to describe the absence of letters might have required participants a further processing step relative to the affirmative instructions used to describe their presence. Second, present and absent clues with non-substitutive features do not require the same amount of perceptual processing (Treisman & Souther, 1985).

Bayesian background

Bayes’ theorem has become a widely used method in the hypothesis-testing literature for updating initial beliefs. This theorem can be expressed in odds (e.g., Beyth-Marom & Fischhoff, 1983; Cherubini, Rusconi, Russo, Di Bari, & Sacchi, 2010; Fischhoff & Beyth-Marom, 1983) as follows:

$$\frac{p(H|D)}{p(\neg H|D)} = \frac{p(H)}{p(\neg H)} \times \frac{p(D|H)}{p(D|\neg H)}$$

where “ $p()$ ”, “ $|$ ”, “ \neg ”, “ H ”, “ $\neg H$ ”, and “ D ” denote “the probability of”, “given that”, the logical symbol for negation, the hypothesis to be tested, the alternate hypothesis, and the new evidence, respectively. The terms are defined as follows from left to right: (1) the posterior odds, that is, the ratio of the probability that the hypothesis is true over the probability that the alternate hypothesis is true given the same information; (2) the prior odds, that is, the ratio of the probability that the hypothesis is true over the probability that the alternate hypothesis is true prior to collecting

new data; (3) the likelihood ratio (LR), that is, the ratio of the probability of observing new data given that the hypothesis is true over the probability of observing the same data given that the alternate hypothesis is true. The LR provides a measure of the evidential strength of the data and is of interest in this study because the tendency to be swayed by the LR of a feature's presence, $p(D | H)/p(D | \neg H)$, over the LR of a feature's absence, $p(\neg D | H)/p(\neg D | \neg H)$, is evidence of an FPE in hypothesis evaluation.

There are several alternate models of the informativeness of a datum (e.g., Crupi, Tentori, & Gonzalez, 2007; Nelson, 2005, 2008; Nelson, McKenzie, Cottrell, & Sejnowski, 2010). One of these models is information gain (IG), which differs from LR in that it considers the probative impact of data by accounting for both the prior probabilities and the posterior probabilities of the hypotheses. Specifically, IG is based on Shannon's (1948) definition of entropy (i.e., uncertainty), whose unit of measure is *bit* and is defined as follows:

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

where X is a discrete random variable, and x_i are the possible values whose probability are $p(x_i)$. IG is the difference between the entropy prior to receiving new information (“ D ”) and the entropy after receiving new information. This equation is as follows.

$$IG = \sum_{i=1}^n p(x_i) \log_2 p(x_i) - \sum_{i=1}^n p(x_i | D) \log_2 p(x_i | D)$$

In other words, IG models the information value of new evidence, which may be either present or absent, by measuring how much the new data either reduce or increase the initial uncertainty about which of the hypotheses is more likely¹.

Surprisal and clue rarity

Measuring the degree to which the occurrence of a piece of evidence is “surprising” is possible with a term within the entropy formula described above. Specifically, “surprisal” (Tribus, 1961) can be expressed as follows.

$$-\log_2 p(x_i)$$

That is, whenever the probability of an outcome is 1, the surprise of seeing it is 0 bits.

Conversely, a highly unlikely outcome is very surprising. For example, when the probability of an outcome is .01, the surprisal is 6.64 bits. According to this notion (also called “self-information” in information theory), the occurrence of a rare event is more diagnostic than that of a common event (for similar arguments, see the “rarity assumption” in Oaksford & Chater, 1994; see also Anderson, 1990, pp. 149–160; McKenzie, 2006, p. 580; McKenzie & Mikkelsen, 2000, 2007). Sensitivity to rare events is relevant to the present paper because Newman et al. (1980) hypothesized that the FPE might have evolved because non-occurring events are relatively more common, and thus less diagnostic, than occurring events in nature. The present studies tested whether participant sensitivity to the rarity of data moderates the FPE.

Overview of the present studies

Three studies tested three possible moderators of the FPE in hypothesis evaluation. Those moderators are (1) the use of substitutive versus non-substitutive features; (2) the relative number of present versus absent features; and (3) the presentation format of feature probabilities (from which the LRs can be derived).

Based on Treisman and Souther’s (1985) visual search results, FPEs that originate from perceptive mechanisms occur when non-substitutive features are used but not when substitutive features are used. Therefore, if an FPE is observed with substitutive features in our hypothesis-testing task, then it probably originated from the later stages of information processing (i.e., reasoning). Study 1 tested this prediction using non-substitutive features, whereas Study 2 tested this prediction using substitutive features.

Furthermore, based on Newman et al.’s (1980) hypothesis that the FPE evolved as a consequence of the information provided by relatively rare stimuli occurrences compared with relatively common stimuli non-occurrences, we expect that the FPE decreases when there are fewer

absent clues than present clues. All of the studies presented in this article tested this prediction by manipulating the relative number of present versus absent clues.

Finally, we conjectured that the FPE should decrease as a consequence of a two-part manipulation. First, the perceptual processes of both present and absent clues should be equated when substitutive features are used (see Treisman & Souther, 1985, Experiment 2). Second, we expected to equate the computational processing of the values associated with both types of clues by presenting participants with both the feature probabilities (which are useful to determine the impact of the present clues) and their complements (which are relevant to assess the impact of the absent clues). Cherubini et al. (in press) found that this explicit presentation format of probabilistic information did not appreciably affect participants' feature-positive choices; however, these authors used non-substitutive features. Accordingly, we hypothesized that equating both the perceptual processing and computational steps required to assess the value of the present and absent clues should debias the FPE. Study 3 tested this prediction.

Study 1

Study 1 provided data to test whether feature type moderates the FPE. Previous studies (Cherubini et al., in press; Rusconi et al., in press) investigating the FPE in hypothesis evaluation have already identified that such an effect does occur when non-substitutive features are used. However, we required new data collected using non-substitutive features that could be converted into substitutive features; otherwise, any differences between studies using substitutive features and those using non-substitutive features may be attributed to the stimuli, that is, to whether or not they can be converted into substitutive features. This study was designed to address this methodological issue and test the reliability of the FPE in previous hypothesis-evaluation studies.

Method

Participants

A total of 42 volunteers (20 female, 22 male, mean age = 21.8 years, $SD = 3.4$, range: 19-37 years; mean education = 15.5, $SD = 2$) took part in the study.

Materials

Participants received a booklet that included a cover page (upon which they provided personal information, e.g., gender, age, and years of education), written instructions, and 12 judgment tasks.

Each task included a card draw. Participants were told that a card was drawn from one of two 100-card decks. The deck was chosen at random (i.e., the prior probabilities of the two decks were equal: $p(H_1) = p(H_2) = .5$). For each problem, participants surmised from which deck (Deck 1 or Deck 2) a card was drawn. Cards either featured red shapes (i.e., triangles, circles, squares, or pentagons) or they did not. Specifically, each card had between 0 and 3 or 0 and 4 shapes printed on its face, depending on the type of problem (see the Design section below). Participants were told which red shapes were present and which were absent from the drawn card in each task. Participants were also shown a picture of the drawn card. A table reported the number of cards (out of the 100 within each deck) that were printed with each of the 3 or 4 shapes. The table values were different for each of the 12 tasks.

Table 2 provides an example of a table and picture that were provided to participants. For example, the “12” in Table 2 indicates that 12 of the 100 cards from Deck 2 had a red triangle printed on them (Thus, a red triangle was missing on 88 cards.) The exemplar problem featured a drawn card with a red triangle and a red circle, but a red square was absent. The instructions stated, “There is a red triangle and a red circle on the card but no red square”. Thus, the presence of features was conveyed via an affirmative instruction, whereas the absence of features was conveyed via a negative instruction.

For each problem, participants selected from which of the two decks the card was most likely drawn by checking one of three boxes labeled “Deck 1”, “Deck 2”, or “equiprobable”. If participants were unable to decide the source of the drawn card, then they checked the “equiprobable” box. The order of the three alternative conclusions was balanced; thus, there were six versions of the questionnaire².

The normatively correct response to the problem illustrated in Table 2 (i.e., the response derived from the application of Bayes' theorem) is "Deck 2". This response is identical to the one that Bayes' theorem suggests given only the absent clue, whereas only considering the present clues would suggest the answer is "Deck 1" (see Problem 4 in Tables 1-3). The subsets of present and absent clues suggested opposing decks in all problems (see Table 3).

The instructions clarified that the number of cards printed with a red shape was independent of the number of cards printed with other red shapes. This clarification was meant to convey to participants that the features (i.e., the red shapes) were class-conditionally independent of each other. Thus, the marginal totals of the two decks should not necessarily sum to 100. Formally, the table values were the frequencies that corresponded to the constituent probabilities of the clue occurrence LRs. That is,

$$p(D | H_1) \text{ and } p(D | H_2)$$

where " D " indicates the shapes (i.e., red triangle, red circle, red square, and red pentagon), " H_1 " denotes Deck 1, and " H_2 " denotes Deck 2. Table 1 reports the constituent probabilities of the LRs for all problems.

The instructions reported a sample probability table and taught participants how to read it. Participants were also informed of the correct response for the sample problem, and they were prompted to not use a calculator and to not take notes when completing the task. The booklet also contained the 12 problems.

Design

The 12 problems originated from a 3×2 within-participants design with 2 problems *per* cell. One factor was correct-response type. The normatively favored hypothesis was the same hypothesis favored by the subsets of present clues (i.e., "presence-consistent" problems), absent clues (i.e., "absence-consistent" problems), or neither (i.e., "equiprobable" problems) (see Table 3). We also manipulated the present-to-absent feature ratio to test the rarity effect. There was an equal

number of present and absent clues in 6 problems (i.e., 2:2 problems) but more present than absent clues in the remaining 6 problems (i.e., 2:1 problems).

Balancing factors

We ensured that different clues were not asymmetrically diagnostic to avoid participant preferences for present versus absent clues based on informativeness considerations.

When the posterior probabilities of the two hypotheses were unequal, the diagnosticity of all the clues was .2781 IG bits (see Table 3). This value corresponds to a shift of $p = .3$ from the prior probabilities. The level of diagnosticity for all the clues was set in accordance with the results of previous similar hypothesis-testing studies involving 130 non-experts. These studies found that the participant sensitivity threshold for clue diagnosticity was between .12 and .18 bits (Cherubini, Russo, Rusconi, D'Addario, & Boccuti, 2009). In the problems in which the hypotheses had equal posterior probabilities, the overall informativeness of the new information was null given that the prior probabilities were also equal.

The diagnosticity of the present clue subset in the 4 equiprobable problems was similar to that of the absent clue subset, regardless of the number of present and absent clues. The mean of the absolute difference in IG between the present and absent clues across problems was .0014 (range = .0004-.0022). The mean deviation between the strength of present clues versus those of the absent clues in the 8 non-equiprobable problems was .3087 (range = .1670-.4263). However, the latter discrepancy in informativeness between the present versus the absent clues was balanced across the non-equiprobable problems. In half of these problems (i.e., Problems 3, 6, 10, and 11 in Table 3), the present clue subset had a mean of .3101 bits more than the absent clues subset. For the remaining problems (i.e., Problems 2, 4, 5, and 9 in Table 3), the absent clue subset had a mean of .3073 bits more than the present clue subset. There was not a significant difference between the present clue subset ($M = .42$, $SD = .36$) and the absent clue subset ($M = .42$, $SD = .34$) with regard to the IG across all 12 problems, $t(11) = .01$, two-tailed $p = .99$.

We also computed the IG of each clue in the 12 problems independent of the other clues and their present or absent status (i.e., the IG of each clue was computed as if it were present). No significant difference was observed between the IG of the present clues ($M = .27$, $SD = .18$) and the IG of the absent clues ($M = .34$, $SD = .2$), $t(17) = -1.06$, two-tailed $p = .306$. Furthermore, the maximum present clue IG was .5463 bits, which was lower than the maximum absent clue IG (.6627 bits). No present clue had a higher IG than the maximum absent clue IG. In contrast, only one of the 18 absent clues had a higher IG (.6627 bits) than the maximum present clue IG. Notably, the clue with the highest IG was a present clue in 6 problems, whereas the clue with the highest IG was an absent clue in the remaining 6 problems. We concluded that our stimuli were sufficiently balanced to ensure that absent clues were not overlooked due to the greater informativeness of present clues.

Dependent variable

We classified participant responses regarding the most likely hypothesis into three groups: those that considered only the present clues (i.e., “presence-consistent” responses), those that considered only the absent clues (i.e., “absence-consistent” responses), or those that considered neither (i.e., “equiprobable” responses). The correctness of participants’ responses provided an additional classification. Specifically, their responses were normatively correct when they chose either the normatively favored deck in the non-equiprobable problems (i.e., “presence-consistent” responses in “presence-consistent” problems and “absence-consistent” responses in “absence-consistent” problems) or “equiprobable” responses in “equiprobable” problems. All other responses were classified as incorrect. According to this classification, a preference for presence-consistent responses over either absence-consistent or equiprobable responses in absence-consistent and equiprobable problems revealed an FPE. Furthermore, although presence-consistent responses to presence-consistent problems were normatively correct, greater accuracy in presence-consistent problems compared with absence-consistent and equiprobable problems indicates an FPE.

Procedure

Volunteers for the study completed the questionnaire in quiet places or study rooms. They received the booklet upon providing consent. There were no time constraints; thus, participants were able to complete the task at their own pace. Upon completion, participants were thanked and debriefed.

Results

The number of presence-consistent, absence-consistent, and equiprobable responses

Table 4 shows the results from a series of hypergeometric tests on the number of presence-consistent, absence-consistent and equiprobable responses for each type of problem. Presence-consistent responses were significantly preferred over the other response types in all conditions but two: the 2:1 absence-consistent and equiprobable problems. Furthermore, participant accuracy was higher in presence-consistent problems compared with absence-consistent and equiprobable problems (see the bold diagonal line in Table 4). These findings suggest that an FPE affected participants, but the rarity of clues moderated this effect. In particular, the FPE decreased when there was one absent clue and two present clues, although this result held only when the presence-consistent responses were incorrect.

Analysis of presence-consistent responses

The previous analysis suggests that clue rarity (present-to-absent feature ratio) and problem type (correct-response type) interacted to influence the number of presence-consistent responses. Thus, we explored this interaction. Given that ANOVAs on count data can yield spurious results (e.g., Jaeger, 2008), we analyzed the number of presence-consistent responses (range = 0 to 2 *per* cell) using a repeated-measures generalized linear model with a Poisson distribution for the response variable using the SASTM statistical package. This within-participants study used a 3 (correct-response type: presence-consistent vs. absence-consistent vs. equiprobable) \times 2 (present-to-absent feature ratio: 2:2 vs. 2:1) design. There was a significant first-level effect of the present-to-absent feature ratio, $\chi^2 = 15.82$, $df = 1$, $p < .0001$, due to the larger number of presence-consistent responses in 2:2 ($M = 1.25$, $SD = .75$) versus 2:1 problems ($M = .93$, $SD = .82$). There was also a

significant first-level effect of correct-response type, $\chi^2 = 10.68$, $df = 2$, $p = .0048$, indicating that the number of presence-consistent responses was higher in the presence-consistent problems ($M = 1.32$, $SD = .73$) compared with absence-consistent ($M = .96$, $SD = .81$), $\chi^2 = 8.28$, $df = 1$, Bonferroni adjusted $p = .008$, and equiprobable problems ($M = .99$, $SD = .81$), $\chi^2 = 7.16$, $df = 1$, Bonferroni adjusted $p = .0148$. Finally, the interaction between correct-response type and the present-to-absent feature ratio was significant, $\chi^2 = 7.15$, $df = 2$, $p = .0280$. Specifically, there were no differences in the number of presence-consistent responses between the 2:2 ($M = 1.33$, $SD = .72$) and 2:1 ($M = 1.31$, $SD = .75$) problems in the presence-consistent problems (i.e., when presence-consistent responses were correct), $\chi^2 = .03$, $df = 1$, Bonferroni adjusted $p = 1$. Conversely, the number of presence-consistent responses was greater for 2:2 ($M = 1.19$, $SD = .83$) versus 2:1 ($M = .74$, $SD = .73$) problems in the absence-consistent problems, $\chi^2 = 11.65$, $df = 1$, Bonferroni adjusted $p = .0018$. Similarly, there were more presence-consistent responses for the 2:2 ($M = 1.24$, $SD = .69$) versus 2:1 problems ($M = .74$, $SD = .86$) in the equiprobable problems, $\chi^2 = 11.31$, $df = 1$, Bonferroni adjusted $p = .0024$.

Discussion

The results of Study 1 provide evidence for the reliability of the FPE in hypothesis-evaluation tasks that use absent, non-substitutive features (see Cherubini et al., in press; Rusconi et al., in press). In fact, participants chose significantly more presence-consistent responses than absence-consistent and equiprobable responses in most conditions. Furthermore, the greater accuracy of participants in presence-consistent problems compared with the other problem types provides evidence of the FPE. The rarity of absent clues undermined the FPE, however, which is in line with Newman et al.'s (1980) hypothesis as well as the tenet of information theory that states that rare observations are valued as more informative than common observations. However, participant sensitivity to the rarity of absent clues was selective in that it emerged only for the problems in which the presence-consistent responses were incorrect (i.e., in absence-consistent and equiprobable problems). The interaction between correct-response type and the present-to-absent

feature ratio reveals this finding. Thus, participants were partially sensitive to the formal correctness of the responses. The significant first-level effect of correct-response type also provides support for this sensitivity. Accordingly, only when normative soundness corroborated the presence-consistent responses did the rarity of the absent clues not undermine the FPE.

Study 2

Study 2 investigated whether the FPE persisted when features were substitutive, that is, when a feature's absence entailed the presence of another feature. For example, when a figure is present, the absence of a color implies the presence of a different color.

Method

Participants

A total of 42 volunteers (23 female, 19 male, mean age = 21.9 years, $SD = 3.6$, range: 19-39 years; mean education = 15.6, $SD = 2.3$) took part in the study.

Materials, procedure and design

The materials, procedure and design were identical to those applied in Study 1 except for the use of substitutive features. This method entailed the presence of a picture of the drawn card with the absent clues as well as the present clues (see Table 5). Furthermore, we used affirmative instructions when presenting the absent clues to participants (e.g., in the problem illustrated in Table 5, the instructions read, "There is a red triangle, a red circle and a yellow square on the drawn card").

Results

The number of presence-consistent, absence-consistent, and equiprobable responses

Table 6 reports the results of a series of hypergeometric tests on the number of presence-consistent, absence-consistent, and equiprobable responses for each problem type. Participants chose more presence-consistent responses than absence-consistent and equiprobable responses in all conditions. The FPE also emerged in the form of greater participant accuracy in the presence-consistent problems compared with the other problem types (see bold diagonal line in Table 6).

Compared with Study 1, absent clue rarity did not undermine the FPE. In contrast, presence-consistent responses occurred most frequently for problems in which there was one absent clue, two present clues, and the presence-consistent responses were correct (i.e., presence-consistent problems). The FPE decreased in 2:1 compared with 2:2 problems for both absence-consistent and equiprobable problems, but this effect was smaller than that of Study 1.

Analysis of presence-consistent responses

As in Study 1, we explored the effects of the present-to-absent feature ratio and correct-response type. The number of presence-consistent responses was analyzed using a generalized repeated-measures model with a Poisson distribution. The predictors were correct-response type (the normatively correct response was presence-consistent, absence-consistent or equiprobable) and the present-to-absent feature ratio (2:2 vs. 2:1). There was a significant first-level effect of correct-response type, $\chi^2 = 11.19$, $df = 2$, $p = .0037$, which showed that the number of presence-consistent responses was higher for presence-consistent problems ($M = 1.39$, $SD = .76$) compared with equiprobable ($M = .95$, $SD = .73$), $\chi^2 = 10.25$, $df = 1$, Bonferroni adjusted $p = .0028$, and absence-consistent problems, although only marginally ($M = 1.15$, $SD = .84$), $\chi^2 = 4.46$, $df = 1$, Bonferroni adjusted $p = .0694$. Conversely, there was not a significant first-level effect of the present-to-absent feature ratio, $\chi^2 = .14$, $df = 1$, $p = .7074$, due to the lack of differences in the number of presence-consistent responses between the 2:2 ($M = 1.13$, $SD = .79$) and 2:1 problems ($M = 1.20$, $SD = .80$). There was also a significant interaction between correct-response type and the present-to-absent feature ratio, $\chi^2 = 12.91$, $df = 3$, $p = .0048$. Specifically, participants provided more presence-consistent (i.e., normatively correct) responses for 2:1 ($M = 1.64$, $SD = .66$) versus 2:2 problems ($M = 1.14$, $SD = .78$) for presence-consistent problems, $\chi^2 = 10.26$, $df = 1$, Bonferroni adjusted $p = .0042$. Conversely, there was no difference in the choice of presence-consistent responses between 2:1 ($M = 1.05$, $SD = .82$) and 2:2 ($M = 1.26$, $SD = .86$) absence-consistent problems, $\chi^2 = 2.79$, $df = 1$, Bonferroni adjusted $p = .2841$. Finally, there was no significant difference in the number of

presence-consistent responses between 2:1 ($M = .90$, $SD = .73$) and 2:2 equiprobable problems ($M = 1$, $SD = .73$), $\chi^2 = .53$, $df = 1$, Bonferroni adjusted $p = 1$.

Discussion

The results of Study 2 are at odds with those of Treisman and Souther (1985, Experiment 2). Although we used substitutive features, we found evidence for an FPE, whereas Treisman and Souther (1985, Experiment 2) did not find a presence-absence asymmetry in the search latencies for substitutive features. The finding of an FPE regardless of the type of feature (i.e., substitutive in Study 2 versus non-substitutive in Study 1) suggests that it is not merely due to difficulties in perceptual processing for this type of hypothesis-evaluation task.

Equating the perceptual processing of the present and absent clues did not attenuate the FPE; moreover, the use of substitutive features strengthened the FPE compared with Study 1. In fact, presence-consistent responses were chosen significantly more often than all other response types regardless of either the type of correct response or the ratio of present-to-absent features. Accordingly, we found evidence against our expectation of a decreased FPE when there are fewer absent clues than present clues. The lack of a significant present-to-absent feature ratio first-level effect reveals this finding. Surprisingly, the rarity effect reversed in terms of an increased FPE for the presence-consistent problems in which there was only one absent clue (i.e., 2:1 problems). Apparently, the use of substitutive features perceptually equates absent clues with present clues such that participants failed to consider the rarity of absent clues as a cue for responding consistent with the absent clue subset alone.

Finally, as in Study 1, participants exhibited a residual sensitivity to formal correctness: they chose presence-consistent responses more often when they were correct than when they were not. The significant correct-response type first-level effect revealed this finding. However, the marginally significant difference in the number of presence-consistent responses between presence-consistent and absence-consistent problems suggests that this sensitivity was weaker than that of Study 1. This finding is consistent with the view that Study 2 had a larger FPE than Study 1.

Study 3

Study 3 tested the prediction that presenting participants with substitutive features and the LR constituent probabilities of occurrences and non-occurrences should undermine the FPE. We expect to equate the perceptual processing of both present and absent clues as well as the computational steps required by the assessment of the value associated with both types of clues. In fact, the use of substitutive features should cause the present and absent clues to be equally perceptible (Treisman & Souther, 1985, Experiment 2). At the computational level, participants do not need to infer the complements from the table to gauge the evidential strength of the absent clues. Moreover, this method uses affirmative instructions to convey present and absent clues. Thus, both types of clues should be indistinguishable from each other.

Method

Participants

A total of 66 volunteers (29 female, 37 male, mean age = 23.3 years, $SD = 3.4$, range: 19-42 years) took part in the study.

Materials, procedure and design

The materials, procedure and design were identical to those used in Study 2, except that participants viewed both the probabilities of feature occurrences and those of feature non-occurrences (see Table 7), and the presentation order of the 12 judgmental tasks was randomized. The balance of the presentation order of the three alternative conclusions was identical to the previous studies.

Results

The number of presence-consistent, absence-consistent, and equiprobable responses

Table 8 reports the results of a series of hypergeometric tests on the number of presence-consistent, absence-consistent, and equiprobable responses for each problem type. Contrary to previous studies and in keeping with our hypothesis, the FPE was virtually undetectable. In fact, participants chose presence-consistent responses more frequently than the other response types for

the presence-consistent problems. Furthermore, they chose absence-consistent responses more often than presence-consistent and equiprobable responses for the absence-consistent problems. However, they did not choose equiprobable responses more often than the other responses for the equiprobable problems. The latter finding might be caused by the lower informativeness of both present and absent clue subsets with regard to the equiprobable problems compared with presence-consistent and absence-consistent problems (see Table 3).

Analysis of presence-consistent responses

As in the previous studies, we analyzed the number of presence-consistent responses using a generalized repeated-measures model with a Poisson distribution. The predictors were correct-response type (the normatively correct response was presence-consistent, absence-consistent or equiprobable) and the present-to-absent feature ratio (2:2 vs. 2:1). There was a significant first-level effect of correct-response type, $\chi^2 = 43.96$, $df = 2$, $p < .0001$, due to the higher number of presence-consistent responses for presence-consistent problems ($M = 1.29$, $SD = .68$) compared with absence-consistent ($M = .37$, $SD = .54$), $\chi^2 = 43.45$, $df = 1$, Bonferroni adjusted $p < .0002$, and equiprobable problems ($M = .83$, $SD = .77$), $\chi^2 = 21.39$, $df = 1$, Bonferroni adjusted $p < .0002$. As in Study 2 (but contrary to Study 1), the present-to-absent feature ratio first-level effect was not significant, $\chi^2 = 1.43$, $df = 1$, $p = .2323$. This result was due to the lack of a difference between the presence-consistent responses in 2:1 ($M = .89$, $SD = .82$) versus 2:2 problems ($M = .77$, $SD = .71$). The interaction between correct-response type and the present-to-absent feature ratio was significant, $\chi^2 = 14.23$, $df = 3$, $p = .0026$. Similar to Study 2 (but in contrast to Study 1), participants provided more presence-consistent responses in the 2:1 ($M = 1.52$, $SD = .68$) than the 2:2 condition ($M = 1.06$, $SD = .6$) for presence-consistent problems, $\chi^2 = 13.64$, $df = 1$, Bonferroni adjusted $p = .0006$. In contrast, there was no difference in the number of presence-consistent responses between the 2:1 ($M = .41$, $SD = .58$) and 2:2 conditions ($M = .33$, $SD = .51$) for the absence-consistent problems, $\chi^2 = .61$, $df = 1$, Bonferroni adjusted $p = 1$. Similarly, the difference between the 2:1 ($M =$

.74, $SD = .75$) and 2:2 conditions ($M = .91$, $SD = .78$) was also not significant for equiprobable problems, $\chi^2 = 1.86$, $df = 1$, Bonferroni adjusted $p = .5172$.

Discussion

The results of Study 3 confirmed much of our hypothesis: the combined use of substitutive features and an explicit presentation format of probabilistic information had a debiasing effect. In fact, participants' choices were usually in keeping with normative considerations (i.e., they preferred presence-consistent responses in presence-consistent problems and absence-consistent responses in absence-consistent problems). Their performances diverged from the Bayesian prescriptions only for the equiprobable problems. This result may reflect the lower diagnosticity of the present and absent clue subsets compared with the other problem types (see Table 3). The preference for presence-consistent responses in 2:2 equiprobable problems and the higher accuracy in presence-consistent problems compared with absence-consistent problems might be due to a residual facilitation in interpreting the relevant LRs of the present clues. In fact, there was a direct correspondence between the color red of the present shapes and the header "red" in the probability table. In contrast, participants had to recode the specific color of the absent shapes (i.e., green, blue, or yellow) to match the table header "other color" (see Table 7). Finally, we did not find evidence of a rarity effect, which replicates the results of Study 2. Participants did not provide significantly more presence-consistent responses in the 2:2 than the 2:1 conditions for the absence-consistent and equiprobable problems. In contrast, we found that the rarity effect was reversed for presence-consistent problems, whereby participants chose more presence-consistent responses in 2:1 compared with 2:2 problems.

General Discussion

This study is the first to show that the use of substitutive features, the absence of which implies the presence of another, does not by itself undermine the FPE in abstract hypothesis-evaluation tasks. Such an effect would result in participants overemphasizing the evidence they received to determine which of two competing hypotheses was the more likely. This finding is

inconsistent with a previous visual perception study that did not find a significant search-latency presence-absence asymmetry when using substitutive features (Treisman & Souther, 1985, Experiment 2). As shown by Treisman and Souther (1985), substitutive features have the advantage of equating the perceptual processes that underlie the elaboration of present clues with those that do so for absent clues. Furthermore, substitutive features allowed us to convey the absence of features through affirmative instructions (Studies 2 and 3). This method avoids possible confounds due to people's difficulty with handling negative information (e.g., Hearst, 1991; Van Wallendael, 1995; Wason, 1959, 1961). Accordingly, the failure of substitutive features to debias the FPE (Study 2) suggests that perceptual processes and the difficulties in interpreting negative information do not by themselves play a central role in causing this phenomenon in abstract hypothesis-evaluation tasks.

However, we found that using substitutive features and explicitly presenting probabilistic information had a debiasing effect (Study 3). This finding is not surprising because we presented participants with (1) a picture of a drawn card featuring both the present clues (i.e., red shapes) and the absent clues (i.e., green, blue, or yellow shapes) and (2) the explicit clue occurrence LRs (which are needed to consider clue presence) as well as their complements, that is, the clue non-occurrence LRs (which are needed to gauge the evidential strength of the absent clues). In addition, we used affirmative instructions to present both present and absent clues. This method caused the present and absent clues to be essentially indistinguishable from each other. Thus, neither the use of substitutive features (Study 2) nor the presentation format of the probabilistic information (Cherubini et al., in press) appreciably undermined participants' feature-positive choices by themselves; however, the combination of these manipulations did (Study 3). In other words, people might consider and use the absence of features when both the perceptual processing and the computational processing that underlie the assessment of present clues are equated with those that underlie the evaluation of absent clues and when both the presence and absence of features are conveyed via affirmative instructions.

The present studies also tested the possible moderator effect of clue rarity; we hypothesized that the FPE would decrease when there are fewer absent clues than present clues. This prediction originated from the information theory tenet that rare observations are more diagnostic than common observations (i.e., self-information or surprisal; Tribus, 1961; see also the “rarity assumption”, e.g., Anderson, 1990; McKenzie, 2006; McKenzie & Mikkelsen, 2000, 2007; Oaksford & Chater, 1994). Furthermore, Newman et al. (1980) hypothesized that the FPE evolved because the occurrence of events is relatively less common and thus more diagnostic than the non-occurrence of events in nature. We found a diminished FPE when there were fewer absent clues than present clues (i.e., 2:1 problems) in Study 1 but only for problems in which the present clues did not suggest the correct response. Conversely, we found a reversal of the rarity effect (i.e., more presence-consistent responses in 2:1 versus 2:2 problems) in Studies 2 and 3 but only for the problems in which the present clues suggested the correct response. Thus, the use of substitutive features in Studies 2 and 3 eliminated participant sensitivity to data rarity, whereas non-substitutive features fostered it in Study 1. Apparently, designing the present and absent clues to be equally distinctive and detectable caused the rare clues to be less salient. The absence of the rarity effect in Study 1 and the reversal of the rarity effect in Studies 2 and 3 in the presence-consistent problems indicate that also sensitivity to normative considerations influenced participant use of the rare clues in a way that deserves additional empirical investigation. Overall, these findings do not support Newman et al.’s (1980) hypothesis. In fact, we found that participants did not place more weight on *all* rare stimuli but only on *some* of them; for example, those that are more perceptually salient than common stimuli.

Our research complements numerous studies that have described the FPE in similar forms including discriminative learning (e.g., Newman et al., 1980; Sainsbury, 1971); learning predictive relationships (e.g., Mutter, Haggblom, Plumlee, & Schirmer, 2006); rule, attribute and concept identification (e.g., Bourne & Guy, 1968; Hovland & Weiss, 1953; Nahinsky & Slaymaker, 1970); contingency judgment (Jenkins & Ward, 1965; Kao & Wasserman, 1993; Mandel & Lehman, 1998;

Smedslund, 1963, the “accounting for occurrences hypothesis”, White, 2008, 2009, but for alternate results and interpretations, see McKenzie & Mikkelsen, 2007 and Vallée-Tourangeau, Hollingsworth, & Murphy, 1998, Experiment 2); probability learning (Estes, 1976); and juridical judgment (Arkes, Shoots-Reinhard, & Mayes, in press, Experiment 2; Wells & Lindsay, 1980). The results of the present research are relevant not only because they provide additional evidence that this phenomenon generalizes to abstract hypothesis-evaluation tasks (see also Cherubini et al., in press; Rusconi et al., in press), but also because they illuminate the conditions under which the FPE does not occur.

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Arkes, H. R., Shoots-Reinhard, B., & Mayes, R. S. (in press). Disjunction between probability and verdict in juror decision making. *Journal of Behavioral Decision Making*. doi: 10.1002/bdm.734
- 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
- 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., & Crippa, F. (in press). Missing the dog that failed to bark in the nighttime: on the overestimation of occurrences over non-occurrences in hypothesis testing. *Psychological Research*. doi: 10.1007/s00426-012-0430-3
- 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.
- Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74, 229–252. doi:10.1086/520779
- 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
- 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
- Hearst, E. (1991). Psychology and nothing. *American Scientist*, *79*, 432-443.
- Hearst, E., & Wolff, W. T. (1989). Addition versus deletion as a signal. *Animal Learning & Behavior*, *17*, 120-133. doi:10.3758/BF03207627
- 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* (pp. 239-275). 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. doi:10.1037/h0093874
- 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
- 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. doi:10.3758/BF03212994
- 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
- Mutter, S. A., Haggbloom, S. J., Plumlee, L. F., & Schirmer, A. R. (2006). Aging, working memory, and discrimination learning. *The Quarterly Journal of Experimental Psychology, 59*, 1556–1566. doi:10.1080/17470210500343546
- 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
- 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
- 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.
- Rusconi, P., Marelli, M., Russo, S., D'Addario, M., & Cherubini, P. (in press). Integration of base rates and new information in an abstract hypothesis-testing task. *British Journal of Psychology*. doi:10.1111/j.2044-8295.2012.02112.x
- Sainsbury, R. (1971). The “feature positive effect” and simultaneous discrimination learning. *Journal of Experimental Child Psychology, 11*, 347–356. doi:10.1016/0022-0965(71)90039-7
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423, 623–656.
- 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

- 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
- Tribus, M. (1961). *Thermostatistics and thermodynamics: An introduction to energy, information and states of matter, with engineering applications*. Princeton, NJ: D. van Nostrand Company, Inc.
- Vallée-Tourangeau, F., Hollingsworth, L., & Murphy, R. A. (1998). ‘Attentional bias’ in correlation judgments? Smedslund (1963) revisited. *Scandinavian Journal of Psychology*, *39*, 221–233.
- Van Wallendael, L. R. (1995). Implicit diagnosticity in an information-buying task. How do we use the information that we bring with us to a problem? *Journal of Behavioral Decision Making*, *8*, 245-264. doi:10.1002/bdm.3960080403
- Ward, R., & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 664–672.
doi:10.1037/0096-1523.15.4.664
- Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *11*, 92–107. doi:10.1080/17470215908416296
- Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, *52*, 133-142. doi:10.1111/j.2044-8295.1961.tb00775.x
- 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
- White, P. A. (2008). Accounting for occurrences: A new view of the use of contingency information in causal judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 204–218. doi: 10.1037/0278-7393.34.1.204
- White, P. A. (2009). Accounting for occurrences: An explanation for some novel tendencies in causal judgment from contingency information. *Memory & Cognition*, *37*, 500–513.
doi:10.3758/MC.37.4.500

Footnotes

¹ Although there are several metrics that are able to define the utility values of evidence, none are conclusively known to account for participant performance in hypothesis *evaluation* tasks (Nelson, 2005 and Nelson et al., 2010 addressed this issue relative to the *testing* stage of hypothesis development.) Accordingly, the present study only considers IG.

² Participants also judged the confidence they placed in their responses using a 7-point scale that ranged from 1 = *not confident* to 7 = *very confident*. Confidence ratings may provide us with potential converging evidence for the FPE by showing the extent to which participants are convinced that judgments based on the featural presence (rather than on the featural absence or on all available evidence) are accurate. Although we measured participant confidence in the responses they provided for all studies presented in this article, we do not discuss these analyses for the sake of brevity and because they do not provide direct evidence for the moderator role of the variables tested.




Table 1

Formal properties of the 12 judgmental tasks of the three studies (the likelihoods of the present clues are in boldface).

Problem	Deck	Prior probabilities	Likelihoods				Posterior probabilities
			<i>p</i> (square)	<i>p</i> (circle)	<i>p</i> (triangle)	<i>p</i> (pentagon)	
1	1	.5	.35	.2	.13	.42	.5
	2	.5	.1	.98	.39	.4	.5
2	1	.5	.91	.7	.9	.9	.2
	2	.5	.1	.23	.12	.06	.8
3	1	.5	.9	.8	.82		.8
	2	.5	.2	.4	.3		.2
4	1	.5	.93	.3	.93		.2
	2	.5	.1	.73	.12		.8
5	1	.5	.12	.11	.1	.24	.8
	2	.5	.87	.69	.9	.57	.2
6	1	.5	.1	.14	.14		.2
	2	.5	.92	.89	.94		.8
7	1	.5	.89	.85	.11	.16	.5
	2	.5	.51	.2	.12	.75	.5
8	1	.5	.2	.13	.35		.5
	2	.5	.98	.39	.1		.5
9	1	.5	.56	.22	.1		.8
	2	.5	.57	.68	.93		.2
10	1	.5	.8	.7	.77	.64	.8
	2	.5	.1	.1	.2	.12	.2
11	1	.5	.14	.2	.28	.1	.2
	2	.5	.9	.1	.93	.95	.8
12	1	.5	.4	.87	.1		.5
	2	.5	.2	.5	.13		.5

Table 2

A sample probability table and a sample picture of a drawn card used in Study 1. The values in the table indicate the number of cards (out of 100) with each shape printed on them.

			
deck 1	93	30	93
deck 2	12	73	10

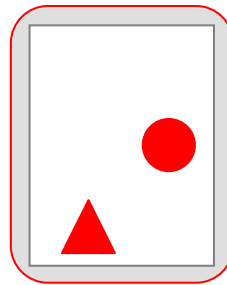


Table 3

The properties of the 12 problems used in the three studies: correct responses, responses suggested by either the present clue subset or the absent clue subset, and the information gain (IG) values of each clue subset as well as all clues.

Problem	Correct response	Response suggested by present clues	Response suggested by absent clues	IG of present clues	IG of absent clues	IG of all clues
1	equiprobable	Deck 2	Deck 1	.0201	.0184	0
2	Deck 2	Deck 1	Deck 2	.7487	.9157	.2781
3	Deck 1	Deck 1	Deck 2	.6150	.1887	.2781
4	Deck 2	Deck 1	Deck 2	.2067	.6260	.2781
5	Deck 1	Deck 2	Deck 1	.6611	.8808	.2781
6	Deck 2	Deck 2	Deck 1	.8769	.6522	.2781
7	equiprobable	Deck 1	Deck 2	.0387	.0375	0
8	equiprobable	Deck 2	Deck 1	.0201	.0224	0
9	Deck 1	Deck 2	Deck 1	.2030	.6260	.2781
10	Deck 1	Deck 1	Deck 2	.8426	.5721	.2781
11	Deck 2	Deck 2	Deck 1	.8020	.4832	.2781
12	equiprobable	Deck 1	Deck 2	.0152	.0148	0

Table 4




The percentages (standard errors of the means in parentheses) of each participant-provided response for all problem types in Study 1. Having devised two versions of each problem type, participants provided 2 responses for each cell of the 3 × 2 design. The total number of participant responses per cell was 84 (N = 42). Asterisks indicate the statistical significance that resulted from the hypergeometric tests performed on the number of responses. Normatively correct responses are in boldface.

	Responses			
	Present-to-absent ratio	Presence-consistent	Absence-consistent	Equiprobable
Presence-consistent	2:2	67% (.11) ***	18% (.1)	15% (.09)
problems	2:1	65% (.12) ***	14% (.08)	20% (.1)
Absence-consistent	2:2	60% (.13) ***	25% (.11)	15% (.1)
problems	2:1	37% (.11)	44% (.13)	19% (.08)
Equiprobable	2:2	62% (.11) ***	25% (.1)	12% (.07)
problems	2:1	37% (.13)	27% (.11)	36% (.12)

*** = exact $p < .001$.

Table 5

A sample probability table and a sample picture of a drawn card used in Study 2 (Problem 4 in Tables 1 and 3). The values in the table indicate the number of cards (out of 100) printed with shapes in red. Study 2 used substitutive features; thus, the absence of the color red denoted the presence of another color. In the example below, the drawn card features a red triangle, a red circle, and a yellow square; that is, the red square is absent.

			
deck 1	93	30	93
deck 2	12	73	10

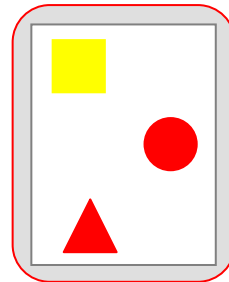


Table 6

The percentages (standard errors of the means in parentheses) of each participant-provided response for all problem types in Study 2. Asterisks indicate the statistical significance that resulted from the hypergeometric tests performed on the number of responses. Normatively correct responses are in boldface

	Responses			
	Present-to-absent ratio	Presence-consistent	Absence-consistent	Equiprobable
Presence-consistent problems	2:2 2:1	57% (.12) *** 82% (.1) ***	18% (.09) 11% (.09)	24% (.09) 7% (.06)
Absence-consistent problems	2:2 2:1	63% (.13) *** 52% (.13) ***	21% (.1) 26% (.1)	15% (.09) 21% (.08)
Equiprobable problems	2:2 2:1	50% (.11) ** 45% (.11) *	19% (.09) 31% (.11)	31% (.11) 24% (.09)

* = exact $p = .027$; ** = exact $p = .002$; *** = exact $p < .001$.

Table 7

A sample probability table and a sample picture of a drawn card used in Study 3 (Problem 4 in Tables 1 and 3). The values in the table indicate the number of cards (out of 100) printed with shapes in red or a different color. As in Study 2, Study 3 used substitutive features; thus, the absence of the color red denoted the presence of another color. In the example below, the drawn card features a red triangle, a red circle, and a yellow square; that is, the red square is absent.

	△		○		red	other color
	red	other color	red	other color		
deck 1	93	7	30	70	93	7
deck 2	12	88	73	27	10	90

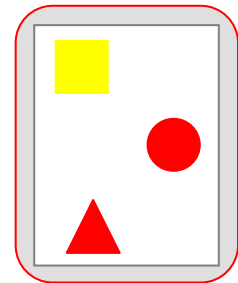


Table 8

The percentages (standard errors of the means in parentheses) of each participant-provided response for all problem types in Study 3. Asterisks indicate the statistical significance that resulted from the hypergeometric tests performed on the number of responses. Normatively correct responses are in boldface

	Responses			
	Present-to-absent ratio	Presence-consistent	Absence-consistent	Equiprobable
Presence-consistent problems	2:2 2:1	53% (.07) *** 76% (.08) ***	20% (.07) 14% (.07)	25% (.08) 11% (.06)
Absence-consistent problems	2:2 2:1	17% (.06) 20% (.07)	45% (.09) ** 64% (.09) ***	38% (.1) 15% (.06)
Equiprobable problems	2:2 2:1	45% (.1) ** 37% (.09)	27% (.08) 41% (.1)	28% (.08) 22% (.08)

** = exact $p = .004$; *** = exact $p < .001$.