

# Framing effects in inference tasks—and why they are normatively defensible

CRAIG R. M. MCKENZIE

*University of California, San Diego, California*

Framing effects occur when logically equivalent redescriptions of objects or outcomes lead to different behaviors, and, traditionally, such effects have been seen as irrational. However, recent evidence has shown that a speaker's choice among logically equivalent attribute frames can implicitly convey (or "leak") normatively relevant information about the speaker's reference point, among other things. In a reinterpretation of data published elsewhere, in this article it is shown that some common effects in inference tasks (covariation assessment and hypothesis testing) can also be seen as framing effects, thereby expanding the domain of framing. It is also shown that these framing effects are normatively defensible because normatively relevant information about *event rarity* is leaked through the description of data and through the phrasing of hypotheses, thereby broadening the information leakage approach to explaining framing effects. Information leakage can also explain why framing effects in such inference tasks disappear under certain conditions.

A trend in research on reasoning is to explain, in rational terms, behavior that has traditionally been seen as irrational or biased (Anderson, 1990; Gigerenzer, 1991; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Hilton, 1995; Klayman & Ha, 1987; Oaksford & Chater, 1994, 1996; Schwarz, 1996). Sometimes the revised conclusion is that the purportedly irrational behavior is consistent with a normative viewpoint, just not with the one originally considered to apply to the task. In this article, a similar point is made in the context of framing effects in inference tasks.

Framing effects occur when logically equivalent redescriptions of outcomes or objects lead to different behaviors. Such effects have been shown to occur in a variety of tasks, but traditionally they have not been associated with inference tasks. Using data published elsewhere (McKenzie & Mikkelsen, 2000, in press), in this article it will be shown that some common phenomena in covariation assessment and hypothesis testing can lead to framing effects. In particular, participants judge covariation differently when presented with logically equivalent descriptions of data, and they perceive different observations as more informative when presented with logically equivalent hypotheses. Demonstrating framing effects in inference tasks expands the current conception of framing.

Furthermore, framing effects are virtually always interpreted as irrational, and the framing effects described

in this article would normally be interpreted this way as well. However, through a recent approach suggesting that (at least some) framing phenomena indicate *rational* behavior, it will be argued that framing effects in these inference tasks are normatively justifiable. The reason is that the different descriptions of the data and hypotheses, though logically equivalent, implicitly "leak" information about which events are rare and which are common, and this leaked information is normatively relevant.

In the next section of the article, three traditional categories of framing effects are described briefly. Subsequently, the recently introduced "information leakage" approach to explaining framing effects is described. Next, some common phenomena in covariation assessment and hypothesis testing are shown to lead to framing effects, but it is argued that the effects are normatively justified. In the last section, implications of the article are discussed and areas of future research are suggested.

## THREE CATEGORIES OF FRAMING EFFECTS

In a recent review, Levin, Schneider, and Gaeth (1998) described three categories of framing effects: attribute framing effects, risky choice framing effects, and goal framing effects. *Attribute framing* refers to situations in which an object is described in terms of a single attribute and the description takes one of two logically equivalent forms. For example, Levin and Gaeth (1988) found that describing ground beef as "75% lean" rather than "25% fat" led participants to rate it as leaner, of higher quality, and less greasy (see also Sanford, Fay, Stewart, & Moxey, 2002). Similar effects have been found when outcomes of medical treatment are described in terms of " $X\%$  survive" rather than " $100 - X\%$  die" (Levin, Schnittjer, &

---

This research was supported by NSF Grants SES-0079615 and SES-0242049. The author thanks Nick Chater, Jonathan Nelson, Mike Oaksford, Shlomi Sher, and an anonymous reviewer for their many helpful comments. Correspondence should be addressed to C. R. M. McKenzie, Department of Psychology, University of California San Diego, 9500 Gilman Drive, MC 0109, La Jolla, CA 92093-0109 (e-mail: cmckenzie@ucsd.edu).

Thee, 1988; Marteau, 1989; Wilson, Kaplan, & Schneiderman, 1987). In attribute framing, one frame is usually positive (e.g., % lean, % survival) and one negative (e.g., % fat, % mortality), and the general finding is that positive frames lead to more favorable responses than do negative frames.

*Risky choice framing tasks* usually present participants with two options: one sure thing and one gamble. The different frames result from the options' being phrased either in terms of gains or in terms of losses. Tversky and Kahneman's (1981) "Asian disease problem" illustrates this class of framing problem:

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

A majority of participants selected Program A over Program B. A separate group was presented with the same cover story, but with the two programs reformulated as follows:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Although Program C is simply a rewording of Program A and Program D is a rewording of Program B, only a minority of participants preferred C to D. That is, group preferences reversed when the options were merely rephrased. Programs A and B are phrased in terms of lives saved, whereas Programs C and D are phrased in terms of lives lost. According to the traditional explanation of the framing effect, people prefer Program A to Program B because they are generally risk averse for gains, but they prefer Program D to Program C because they are generally risk seeking for losses (Kahneman & Tversky, 1979; see also Kahneman & Tversky, 1984; Tversky & Kahneman, 1981, 1986). Like attribute framing effects, risky choice framing effects have been demonstrated using a wide variety of tasks (for reviews, see Kühberger, 1998; Levin et al., 1998).

The third type of framing effect described by Levin et al. (1998) is goal framing. In a typical goal-framing task, there are two actions that one can take, and either the advantages of one action or the (logically equivalent) disadvantages of the other action are mentioned. Here, the question is which framing more successfully leads participants to take a particular course of action. For example, Meyerowitz and Chaiken (1987) demonstrated that women were more apt to engage in breast self-examination (BSE) when it was stressed that *not* engaging in BSE *decreased* the chances of finding a tumor

early than when it was stressed that engaging in BSE *increased* the chances of finding a tumor early.

## INFORMATION LEAKAGE AND FRAMING EFFECTS

Traditionally, all three types of framing effects have been interpreted as indicating irrational or biased behavior. However, Sher and McKenzie (2004) have argued that if framing effects are to be taken as evidence of irrationality, it is not sufficient that the frames in question be logically equivalent. Instead, they must be *information equivalent*. Frames are information equivalent only if no normatively relevant information can be inferred from the speaker's choice of frame. The fact that a speaker chooses one utterance over another that is logically equivalent can convey implicit information that is relevant to the listener.

The idea that logically equivalent utterances are sometimes interpreted differently can be found in pragmatic analyses of language use. Such analyses are concerned not just with what a speaker *says*, but also with what a speaker *means* (see, e.g., Hilton, 1995; Schwarz, 1996). For example, Johnson-Laird (1968a, 1968b) found that passive and active forms of sentences, although logically equivalent, convey different information about the relative prominence of the subject and object (e.g., in "The man was kissed by the woman," the man is intended and interpreted to be more prominent than in "The woman kissed the man"; see also Moxey & Sanford, 2000).

In the context of attribute framing, McKenzie and Nelson (2003) and Sher and McKenzie (2004) have shown that speakers do not randomly choose among logically equivalent frames but are instead systematically influenced by reference points. In particular, speakers are more likely to describe an object in terms of a given property when that property has increased rather than decreased relative to their reference point. To take a simple example, when given a choice between describing a 4-oz cup with water at the 2-oz line as "half empty" or "half full," participants are more likely to describe it as "half empty" if the cup used to be full than if it used to be empty (McKenzie & Nelson, 2003; see also Sher & McKenzie, 2004). This is because the cup's emptiness has increased if it was previously full, whereas its fullness has increased if it was previously empty. Thus, logically equivalent frames can "leak" different information (Sher & McKenzie, 2004), in this case information about whether the speaker considers the cup relatively full or relatively empty. Similar effects were found when participants described a new medical treatment outcome. When given a choice between describing the new treatment outcome in terms of either " $X\%$  survive" or " $100 - X\%$  die," they were more likely to choose the survival frame when, in comparison with the old treatment, the new treatment led to relatively many survivors than when it led to relatively few (McKenzie & Nelson, 2003). This example shows that *normatively relevant* information can be leaked via logi-

cally equivalent frames. In this case, the speaker's choice of frame reveals whether the speaker considers the treatment relatively successful or relatively unsuccessful. It is hard to deny the normative relevance of this leaked information. (Imagine that the information implicit in the speaker's choice of frame were made explicit.) Furthermore, "listeners" presented with a frame made accurate inferences about the speaker's reference point. For example, they were more likely to infer that the old treatment led to more deaths when the new treatment was described in the terms of " $X\%$  survive" rather than in terms of " $100 - X\%$  die."

In both the glass and medical treatment experiments described above, speakers' reference points systematically influenced their choice among logically equivalent frames, and listeners tended to infer the same reference points when presented with a given frame. Logical equivalence does not guarantee information equivalence, and listeners may often be justified in responding differently to logically equivalent frames.

The information leakage account of framing effects developed thus far (McKenzie & Nelson, 2003; Sher & McKenzie, 2004) has focused primarily on attribute framing problems and on the leakage of reference point information by the speaker's choice of frame. One way to view the research conducted up to this point is that it serves as an existence proof: At least some of the logically equivalent frames used by framing researchers are *not* information equivalent. However, the logic of the approach is limited neither to attribute framing nor to reference point leakage. For example, Sher and McKenzie (2004) have suggested that, more generally, a speaker's choice of attribute frame reveals what the speaker considers notable or essential about the object being described. Indeed, normatively relevant information might be leaked through speakers' frame selection in a variety of ways in a variety of tasks. In the next section, it is argued that a new type of normatively relevant information is leaked in a new category of framing effect.

#### A FOURTH CATEGORY: FRAMING EFFECTS IN INFERENCE TASKS

In the remainder of this article, I will show that some common findings in covariation assessment and hypothesis testing can lead to a new type of framing effect.<sup>1</sup> Logically equivalent descriptions of data and of hypotheses can lead to different behaviors, and such behaviors have traditionally been interpreted as nonnormative and biased. In both cases, however, a new (i.e., not previously discussed) kind of information is leaked by the different descriptions, and this provides a rational explanation of why the framing effects occur: Information about *event rarity* is leaked by the descriptions, and this information has normative relevance. In other words, listeners infer what is rare on the basis of how hypotheses and data are described, and behaving differently when one is presented with the different (albeit logically equivalent) descrip-

tions is normatively justifiable. Bolstering this point of view is the fact that the framing effects disappear when it is clear which events are rare, in which case listeners no longer need to infer rarity on the basis of the speaker's choice of description.

#### Covariation Assessment

Behavioral phenomena such as learning (Hilgard & Bower, 1975), categorization (Smith & Medin, 1981), and judging causation (Cheng, 1997; Cheng & Novick, 1990, 1992; Einhorn & Hogarth, 1986) rely on an ability to assess how variables covary. Crocker (1981) pointed out that people's ability to accurately assess covariation allows them to explain the past, control the present, and predict the future. Because of its importance, much research has been devoted to covariation assessment since the groundbreaking studies of Inhelder and Piaget (1958) and Smedslund (1963; for reviews, see Allan, 1993; McKenzie, 1994).

In a typical covariation task, there are just two variables, each with levels of presence and absence, creating the familiar  $2 \times 2$  contingency matrix. Such a matrix is depicted in Figure 1 for Variables  $X$  and  $Y$ . Cell A corresponds to the joint presence of  $X$  and  $Y$ , Cell B to the presence of  $X$  and the absence of  $Y$ , Cell C to the absence of  $X$  and the presence of  $Y$ , and Cell D to the joint absence of  $X$  and  $Y$ . Imagine, for example, being presented with the following information regarding a treatment ( $X$ ) and recovery from an illness ( $Y$ ): 15 people received the treatment and recovered (Cell A), 5 people received the treatment and did not recover (Cell B), 9 people did not receive the treatment and recovered (Cell C), and 3 people did not receive the treatment and did not recover (Cell D). Participants might be asked to assess the strength of the relation between the two variables given the four cell values, or they might be asked which of two matrices shows a stronger relation. Sometimes the observations are presented sequentially to participants, whereas other times the information is summarized (as above).

The model considered normative by virtually all covariation researchers is  $\Delta p = (A/[A + B]) - (C/[C + D])$ , where A, B, C, and D correspond to the respective cell values (Allan, 1993; Jenkins & Ward, 1965; Kao & Wasserman, 1993; Mandel & Lehman, 1998; Shaklee &

		Variable Y	
		Present	Absent
Variable X	Present	Cell A	Cell B
	Absent	Cell C	Cell D

Figure 1. The four cells of a  $2 \times 2$  contingency matrix.

Tucker, 1980; Ward & Jenkins, 1965; Wasserman, Dornier, & Kao, 1990). A closely related measure (see Allan, 1980),  $\varphi$ , which is a special case of Pearson's product-moment correlation coefficient, is the (traditional) normative measure used in this article:  $\varphi = (AD - BC) / [(A + B)(C + D)(A + C)(B + D)]^{1/2}$ . Using one measure or the other does not alter any points made herein. Both  $\varphi$  and  $\Delta p$  range between  $-1$  and  $1$ ; to the extent that they are close to  $1$  ( $-1$ ), there is a strong positive (negative) relation between the variables:  $Y$  is more (less) likely to be present when  $X$  is present than when  $X$  is absent. When  $\varphi$  and  $\Delta p = 0$ , as in the above example,  $X$  and  $Y$  are independent. Note that  $\varphi$  and  $\Delta p$  are *statistical summaries* of the four cell frequencies and treat each frequency as equally important.

Probably the most robust finding in the covariation literature is that participants do not treat the four cells as equally important. In particular, observation of the joint presence of the variables (Cell A) has the largest impact on judgments, and observation of the joint absence of the variables (Cell D) has the smallest impact. This has been shown by regressing strength judgments onto cell frequencies (Mandel & Lehman, 1998; Schustack & Sternberg, 1981) by asking participants directly which cells are most important (Crocker, 1982; Wasserman et al., 1990), by inferring cell importance and/or which rule participants use on the basis of patterns of responses (Arkes & Harkness, 1983; Kao & Wasserman, 1993; Levin, Wasserman, & Kao, 1993; Shaklee & Tucker, 1980; Ward & Jenkins, 1965; Wasserman et al., 1990), and by a meta-analysis of covariation research (Lipe, 1990). The impact of Cells B and C falls between that of Cells A and D (Crocker, 1982; Kao & Wasserman, 1993; Levin et al., 1993; Lipe, 1990; Wasserman et al., 1990). Because the four cells are equally important in calculating  $\varphi$  and  $\Delta p$ , differential cell impact has been routinely interpreted as nonnormative. For example, Kao and Wasserman (1993) stated, "It is important to recognize that unequal utilization of cell information implies that nonnormative processes are at work" (p. 1365), and Mandel and Lehman (1998) attempted to explain differential cell informativeness in terms of a combination of two reasoning biases.

Because the largest difference in impact is between Cells A and D, both of which provide evidence of a positive relation between the variables, these two cells are the focus of this section, both theoretically and empirically.

**A framing effect in covariation assessment.** Although not seen in this light before, the preference for Cell A data over Cell D data can lead to a framing effect merely by switching which level of each variable is treated as present or absent. Usually, when a variable has levels of *present* and *absent* (e.g., when a symptom is either present or absent), the labeling is not arbitrary. In the study described below, however, the variables were chosen so that the level to be treated as present or absent could be manipulated.

In a recent article, McKenzie and Mikkelsen (in press) investigated participants' use of Cell A and Cell D information. In one experiment, their participants imagined

that they were trying to uncover the factors that determine whether people have personality type  $X$  or personality type  $Y$ . The participants were told that everyone has one personality type or the other. The factor they were examining was genotype, and everyone was said to have either genotype  $A$  or genotype  $B$ . In order to find out if there was a relationship between genotype and personality type, they saw each individual's records stating whether he or she had genotype  $A$  (yes or no) and whether he or she had personality type  $X$  (yes or no). Note that these records were described in terms of the presence or absence of genotype  $A$  and of personality type  $X$ .

Among other, related tasks, the participants were shown two different (purportedly random) samples of 9 people (see Table 1, Condition 1). The numbers in the columns indicate the number of people falling into each category for each sample. For instance, 6 of the 9 people in Sample 1 had genotype  $A$  and personality type  $X$ . Table 1 shows a slightly condensed version of what the participants saw. Specifically, rather than appearing to the immediate right of Sample 1, Sample 2 appeared below Sample 1 (along

**Table 1**  
**Results for the Four Conditions From a Covariation Study**  
**Reported by McKenzie and Mikkelsen (in press)**

Condition	Sample 1	Sample 2	Cell
<b>Condition 1 (Abstract)</b>			
Genotype $A$ : Yes/Personality type $X$ : Yes	6	1	A
Genotype $A$ : Yes/Personality type $X$ : No	1	1	B
Genotype $A$ : No/Personality type $X$ : Yes	1	1	C
Genotype $A$ : No/Personality type $X$ : No	1	6	D
"Which sample provides stronger evidence of a relation?"	76%	24%	
<b>Condition 2 (Abstract)</b>			
Genotype $B$ : No/Personality type $Y$ : No	6	1	D
Genotype $B$ : No/Personality type $Y$ : Yes	1	1	C
Genotype $B$ : Yes/Personality type $Y$ : No	1	1	B
Genotype $B$ : Yes/Personality type $Y$ : Yes	1	6	A
"Which sample provides stronger evidence of a relation?"	26%	74%	
<b>Condition 3 (Concrete)</b>			
Emotionally disturbed: Yes/Drop out: Yes	6	1	A
Emotionally disturbed: Yes/Drop out: No	1	1	B
Emotionally disturbed: No/Drop out: Yes	1	1	C
Emotionally disturbed: No/Drop out: No	1	6	D
"Which sample provides stronger evidence of a relation?"	73%	27%	
<b>Condition 4 (Concrete)</b>			
Emotionally healthy: No/Graduate: No	6	1	D
Emotionally healthy: No/Graduate: Yes	1	1	C
Emotionally healthy: Yes/Graduate: No	1	1	B
Emotionally healthy: Yes/Graduate: Yes	1	6	A
"Which sample provides stronger evidence of a relation?"	67%	33%	

Note—The participants did not see the information pertaining to cells presented in the rightmost column. Most participants considered the sample in which the large frequency corresponded to Cell A (rather than to Cell D) to provide the strongest evidence of a relationship—except in Condition 4, in which the participants knew that Cell A observations were common. In that condition, most participants considered the large Cell D sample to provide the strongest support.

with a repetition of the descriptions of the cells). Furthermore, the participants did not see the cell label. Note that the frequencies for Cells B and C are held constant; the only difference between the two samples is whether the large frequency corresponds to Cell A or to Cell D. The participants were then asked, "In your opinion, which of the two samples provides stronger support for a relationship between genotype and personality type?" (No definition of "relationship" was provided to the participants.) Table 1 shows that most of the participants (76%) selected Sample 1, in which the high frequency corresponded to the joint presence of the two variables (the yes/yes category), or Cell A. This finding replicates many previous results showing that participants are influenced more by Cell A than by Cell D.

In another condition (see Table 1, Condition 2), the labeling of the observations in terms of "yes" and "no" was simply reversed, but the relabeling maintained the logical identity of each observation. Rather than indicate whether or not each person had genotype *A* and personality type *X*, the records showed whether each person had genotype *B* (yes or no) and personality type *Y* (yes or no). For example, a person labeled in Condition 1 as genotype *A*/personality type *X* (Cell A) was relabeled in Condition 2 as not-genotype *B*/not-personality type *Y* (Cell D). The participants were presented with the two samples of 9 people shown in the second section of Table 1. Note that these two sample descriptions are logically equivalent to their counterparts presented earlier (Condition 1). For example, the results of Sample 1 are the same in both conditions; the categories are simply labeled differently. Nonetheless, because of the larger impact of Cell A relative to Cell D, Table 1 shows that the participants' responses were reversed: Now most participants reported that Sample 2 provided stronger evidence of a relation between genotype and personality type. Although McKenzie and Mikkelsen (in press) did not interpret their results in this manner, the reversal of responses to the logically equivalent evidence demonstrates a framing effect.

**Making normative sense of the framing effect.** Despite the apparent errors by participants, information leakage, combined with a Bayesian approach to the covariation task, provides a rational explanation of the larger impact of joint presence observations over joint absence observations and the resulting framing effect. If it is assumed that variables that are either present or absent tend to be absent more often than present, then (1) from a Bayesian perspective joint presence is in fact more informative than joint absence (Anderson, 1990; Anderson & Sheu, 1995; McKenzie & Mikkelsen, in press) and (2) identifying (implicitly or explicitly) a level of a variable as present or absent leaks information about event rarity.

The Bayesian account assumes that participants approach the task as one of *induction* rather than as one of statistical summary. Participants are assumed to treat the cell values as samples from a larger population of instances and to be trying to distinguish between compet-

ing hypotheses about the larger population—for example, whether there is a positive relationship versus no relationship between the variables. Observation of the presumably rare case of Cell A discriminates better between the competing hypotheses. If absence of the two variables is common, then it would not be surprising to see both variables absent—a Cell D observation—even if the variables were independent. In contrast, it would be surprising to observe their joint presence—Cell A—especially if the variables were independent. Joint presence would provide stronger support than would joint absence for the hypothesis that the variables are related.

For illustrative purposes, assume that one hypothesis, H1, corresponds to a moderate positive contingency between *X* and *Y*,  $\rho = .5$  (where  $\rho$  is the population's hypothesized  $\phi$ ), and that the alternative hypothesis, H2, corresponds to noncontingency,  $\rho = 0$ . In other words, assume that participants are trying to determine how likely it is that there is a moderate contingency between the variables rather than none. (These assumptions are relaxed below.) Now we can ask how informative Cell A and Cell D observations are under these conditions. From a Bayesian perspective, data are informative, or diagnostic, to the extent that they help distinguish between the hypotheses under consideration. This is quantified through the use of likelihood ratios. The numerator of a likelihood ratio corresponds to the probability of observing the data assuming that H1 is true, and its denominator corresponds to the probability of observing the same data assuming that H2 is true. A datum is diagnostic to the extent that its likelihood ratio differs from 1. In the case of a Cell A observation, the likelihood ratio is  $p(\text{Cell A} | \text{H1})/p(\text{Cell A} | \text{H2})$ . Because in H2  $\rho = 0$ —that is, *X* and *Y* are independent— $p(\text{Cell A} | \text{H2}) = p(X)p(Y)$ , where  $p(X)$  and  $p(Y)$  correspond to the relative frequencies, or subjective probabilities, of the presence of the variables, *X* and *Y*, respectively, in the population of interest. Similarly,  $p(\text{Cell D} | \text{H2}) = [1 - p(X)][1 - p(Y)]$ . Assume further that  $p(X) = p(Y) = .1$ —that is, that the presence of each variable is rare. The denominators of the likelihood ratios for A and D observations equal .01 and .81, respectively. The respective numerators of the likelihood ratios for  $\rho = .5$  (H1) are .055 and .855.<sup>2</sup> Thus, the likelihood ratios corresponding to A and D observations are 5.5 and 1.06, respectively, showing in this example that a Cell A observation is much more diagnostic than a Cell D observation. Indeed, the Cell D likelihood ratio is very close to 1, indicating that it is virtually uninformative.

The two panels in Figure 2 show the two cells' likelihood ratios as a function of  $p(X)$  and  $p(Y)$ , which were orthogonally manipulated between .1 and .9 in steps of .1 (resulting in 81 data points in each panel). H1 was  $\rho = .1$ , and H2 was  $\rho = 0$ . (The low  $\rho$  value for H1 was used because there are low upper bounds on  $\rho$  when one of  $p(X)$  or  $p(Y)$  is low and the other is high.) Each cell's informativeness is heavily influenced by  $p(X)$  and  $p(Y)$ . The left panel shows that Cell A is most informative when  $p(X)$  and  $p(Y)$  are both low, and the right panel shows

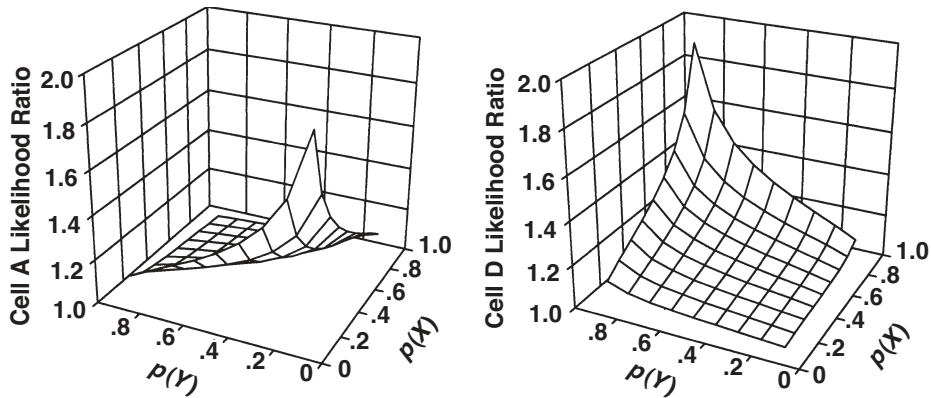


Figure 2. Cell A and Cell D likelihood ratios as a function of  $p(X)$  and  $p(Y)$ . The left panel shows that Cell A is most informative when  $p(X)$  and  $p(Y)$  are low—i.e., when the presence of  $X$  and the presence of  $Y$  are both rare. The right panel shows that Cell D is most informative when  $X$  and  $Y$  are both common.

that Cell D is most informative when  $p(X)$  and  $p(Y)$  are both high.

Of course, assumptions were made about the specific competing hypotheses in the analysis above, but the conclusion is considerably more general. If it is assumed only that the marginal probabilities,  $p(X)$  and  $p(Y)$ , do not change under the competing hypotheses, then *the competing hypotheses are irrelevant* to the conclusion that Cell A is more informative than Cell D. The conclusion is true whenever  $p(X) < 1 - p(Y)$  (Horwich, 1982; Mackie, 1963; McKenzie & Mikkelsen, 2000). Thus, whenever  $p(X)$  and  $p(Y)$  are both low (i.e., less than .5), Cell A is normatively more informative than Cell D, at least from a Bayesian perspective (see also Anderson, 1990, pp. 149–160; McKenzie & Mikkelsen, in press).

When one is presented with two entire matrices and asked which provides stronger evidence for H1 (rather than H2), a reasonable rule would be to select the matrix that results in the largest likelihood ratio,  $p(E|H1)/p(E|H2)$ , where E corresponds to all the data contained in the matrix. In terms of the specific matrices presented to participants above, the logic is simply that the matrix with the large cell corresponding to the most diagnostic evidence in favor of a relationship provides stronger evidence in favor of a relationship than the matrix with the large cell corresponding to the least diagnostic evidence (holding evidence against a positive relationship, Cells B and C, constant, as is done here). As was illustrated above, the likelihood ratio for the same matrix values can differ depending on beliefs or assumptions about the marginal probabilities—that is, about which events are rare.

Do variables that are either present or absent tend to be absent more often than present? The answer will depend on the specific circumstances, but in the vast majority of cases the answer appears to be yes: Most things are not red, most things are not mammals, most people do not have a fever, and so on.<sup>3</sup> It seems plausible that people have learned through a lifetime of experience that the presence of a variable is rarer than its absence and

that, therefore, observing the joint presence of two variables is usually more informative than observing their joint absence when trying to determine if the variables are related. A Cell A “bias” might reflect deeply rooted tendencies that are highly adaptive outside the laboratory. Combining the normative Bayesian account with information leakage—the notion that labeling a level of a variable as present or absent leaks normatively relevant information about event rarity—provides a rational explanation of the above framing effect in covariation judgment.

**Making the framing effect disappear.** So far, the account above could be seen as a mere post hoc rationalization of irrational behavior. However, the account leads to the following prediction: If event rarity is already known by participants and there is therefore no need to *infer* rarity from how the variables or observations are described, then the framing effect will disappear. Interestingly, this also leads to the (equivalent) prediction that the robust Cell A “bias” will reverse under certain conditions, becoming a Cell D “bias!” These predictions are predicated on the assumption that participants are sensitive to event rarity when making inferences, as has been shown elsewhere (McKenzie & Amin, 2002; McKenzie & Mikkelsen, 2000; Oaksford & Chater, 1994, 1996, 2003; Oaksford, Chater, & Grainger, 1999; Oaksford, Chater, Grainger, & Larkin, 1997).

The most promising way to demonstrate such sensitivity is to use variables with which participants are familiar (McKenzie & Mikkelsen, 2000). Ideally, participants would already know how common the levels of each variable are. To test this, McKenzie and Mikkelsen (in press) asked participants in the *concrete condition* of their experiment, described above, to imagine that they worked at a large high school and were trying to uncover factors that would help determine students’ “high school outcome” (i.e., whether they drop out or graduate). The factor being examined was students’ “emotional status.” All students were said to undergo a thorough psycholog-

ical examination in their freshman year and to be categorized as either emotionally disturbed or emotionally healthy. Although it was assumed that the participants knew that dropping out and being emotionally disturbed are rare, these factors were reinforced in the task instructions. Except for the change in cover story, the materials were identical to those used in the “abstract” scenario presented above.

These “concrete” participants were told that they had access to the records of former students in order to find out if there was a relationship between students’ emotional status and their high school outcome. Half of the participants were told that each record stated whether the student was emotionally disturbed (yes or no) and whether the student dropped out (yes or no). Thus, the presence (i.e., the “yes” level) of each variable was rare, making a Cell A observation rarer than a Cell D observation. The participants were then presented with the two samples of nine observations (see Condition 3 in Table 1), one with many Cell A observations and one with many Cell D observations, and were asked, “In your opinion, which of the two samples provides stronger support for a relationship between emotional status and high school outcome?” The Bayesian account predicts the same result that has been found in earlier covariation studies, including the one reported above: Because presence is rare in this condition, the participants should judge the large Cell A sample as providing stronger evidence of a relation between emotional health and high school outcome. Indeed, Table 1 shows that most participants selected the large Cell A sample.

The key condition was the remaining one. Some participants were presented with the same concrete scenario with the labeling simply reversed, just as it had been in the abstract condition (see Condition 4 in Table 1). Rather than state whether each student was emotionally disturbed and whether each dropped out, the records stated whether each was emotionally healthy (yes or no) and whether each graduated (yes or no). Now the *absence* of each of these variables was rare, making Cell A more common than Cell D. The Bayesian perspective leads to a prediction for this condition that is the opposite of all previous covariation findings: Participants will find Cell D observations most informative. Indeed, McKenzie and Mikkelsen (in press) found this to be the case. As is shown in Table 1, only 33% of these participants selected the sample with the large Cell A frequency as providing stronger support; that is, most found the large Cell D sample more supportive. Because it was clear to the participants in the concrete conditions which events were rare and which were common, they had no need to infer rarity on the basis of how the variables or observations were described. The result was that the framing effect demonstrated in the abstract conditions virtually disappeared in the concrete conditions.

McKenzie and Mikkelsen’s (in press) results appear to be the first demonstration of a reversal of the Cell A “bias.” Their results provide strong evidence for the hypothesis that the robust Cell A bias demonstrated over the past four decades stems from (1) participants’ Bayes-

like approach to the task and (2) their (perhaps implicit) default assumption that presence is rare. When there is good reason to believe that absence is rare, Cell D is deemed more informative, just as the Bayesian approach predicts.

Finally, note that the behavior of both the abstract and the concrete groups is explained in terms of their sensitivity to rarity: The abstract group exploited knowledge about how labeling leaks information about which events are rare, and the concrete group exploited knowledge about which observations were actually rare. This is why there was a framing effect for the abstract group but not for the concrete group.

### Hypothesis Testing

An important area of psychological investigation is concerned with how people search for and use data when testing hypotheses (for reviews, see Evans, 1989; Klayman, 1995; Klayman & Ha, 1987; McKenzie, 2004; Nickerson, 1998). Since Wason’s (1960) research on the 2–4–6 task, it has been clear that, when testing hypotheses, people tend to examine cases that they expect to work rather than those they expect not to work (Klayman & Ha, 1987). For example, in Wason’s task, participants try to determine what rule the experimenter has in mind that generates triples of numbers, an example of which is 2–4–6. They form an initial hypothesis, such as “numbers must increase by two,” and then test triples that conform to the hypothesis (e.g., 8–10–12) rather than triples that do not conform (e.g., 8–9–10). As another example, when asked to test whether someone is an extravert, participants tend to ask questions about information that one would expect to be true of an extravert, and they tend not to ask questions about information one would expect to be true of an introvert (Snyder, 1981; Snyder & Swann, 1978; see also Fischhoff & Beyth-Marom, 1983; McKenzie, 1994, 1998, 1999; McKenzie, Wixted, Noelle, & Gyurjyan, 2001; Mynatt, Doherty, & Tweney, 1977). This hypothesis-testing strategy was originally thought to lead to a “confirmation bias” because it often led participants in laboratory experiments to incorrectly believe their initial, tentative hypotheses. However, Klayman and Ha (1987) showed that the strategy works well under what appear to be naturally occurring conditions (one of which is that the events targeted by the hypothesis are rare). The authors suggested the neutral-phrase *positive test strategy* to refer to the strategy of testing cases expected to work.

In tasks in which participants are asked to test hypotheses of the form, “If  $X_1$ , then  $Y_1$ ,” where variables  $X$  and  $Y$  each have two levels ( $X_1$  and  $X_2$ ,  $Y_1$  and  $Y_2$ ), positive testing manifests itself in the following way: Participants consider an  $X_1$ & $Y_1$  observation to be more supportive than an  $X_2$ & $Y_2$  observation, although both observations support the hypothesis. In other words, confirming observations that are mentioned in the hypothesis are deemed more informative than confirming observations that are not mentioned in the hypothesis.

**A framing effect in hypothesis testing.** Although not interpreted this way before, the tendency to consider the observations mentioned in hypotheses to be most informative can lead to a framing effect. Using a hypothesis-testing task with a cover story similar to that of the abstract covariation task used by McKenzie and Mikkelsen (in press), McKenzie and Mikkelsen (2000) had participants imagine that they were researchers investigating a possible relation between genetics and personality type. They were told that everyone has either genotype *A* or genotype *B* and that everyone has either personality type *X* or personality type *Y*. Some participants tested the following hypothesis: "If a person has personality type *Y*, then he/she has genotype *B*" (or " $Y \rightarrow B$ ").<sup>4</sup> Of the first 2 people observed, 1 had genotype *A* and personality type *X* (*A*&*X*) and 1 had genotype *B* and personality type *Y* (*B*&*Y*). Both of these observations support the hypothesis, but when asked which provided stronger support, more than 70% of these participants chose the mentioned *B*&*Y* observation. Other participants were asked to test the hypothesis "If a person has genotype *A*, then he/she has personality type *X*" (or " $A \rightarrow X$ "), and almost 80% of these participants selected the mentioned *A*&*X* observation as more supportive. The results for these two "abstract" groups are illustrated in the leftmost panel of Figure 3. The tall first column shows that most participants selected the *B*&*Y* observation when testing  $Y \rightarrow B$ , and the short second column shows that few selected the *B*&*Y* observation when testing  $A \rightarrow X$ . (Although the abstract groups had no explicit information regarding rarity, the *B*&*Y* observation is referred to as the "rare observation" in Figure 3, for reasons that will become clear shortly.)

On the face of it, these participants' behavior is peculiar for the following reason: The two hypotheses are logically equivalent (one is the contrapositive of the other), and, therefore, whichever observation supports one hypothesis more strongly must also support the other hypothesis more strongly.<sup>5</sup> Nonetheless, because of the tendency to consider mentioned observations more supportive, participants selected different observations depending on which logically equivalent hypothesis was presented to them. In other words, a framing effect occurred (although McKenzie & Mikkelsen, 2000, did not interpret their results as such).

**Making normative sense of the framing effect.** One can make normative sense of the framing effect if, when testing  $X1 \rightarrow Y1$ , participants assume that *X1* and *Y1* (the

mentioned events) are rare relative to *X2* and *Y2* (the unmentioned events). If this were so, then, from a Bayesian perspective, the mentioned observation would be *normatively* more informative than the unmentioned observation. (Essentially the same normative analysis presented in the previous section on covariation assessment is applicable here. See McKenzie, Ferreira, Mikkelsen, McDermott, & Skrable, 2001; McKenzie & Mikkelsen, 2000; see also Nickerson, 1996; Oaksford & Chater, 1994, 1996.) In other words, it would be normatively justifiable to treat mentioned observations as more informative if hypotheses tend to be phrased in terms of rare events.

Do people tend to phrase conditional hypotheses in terms of rare events? McKenzie, Ferreira, et al. (2001) presented some participants with the following scenario: A prestigious college receives many applications but admits few applicants. Listed in Table 2 is information regarding 5 high school seniors who applied last year. Next to each applicant is a rating from the college in five categories. As can be seen in the table, in each category 1 candidate was rated "high" and the other 4 were rated "low." On the far right, it is shown that only 1 of the 5 candidates was accepted. On the basis of this information, the participants supplied the best explanation of why the applicants experienced different outcomes by completing the statement, "If applicants \_\_\_\_\_, then \_\_\_\_\_."

It is easy to see that only SAT scores correlate perfectly with whether the applicants were rejected or accepted. However, a choice still remains as to how to complete the statement. The participants could write, "If applicants have high SAT scores, then they will be accepted" or "If applicants have low SAT scores, then they will be rejected." Both statements are accurate, but the former phrasing targets the rare events and the latter targets the common ones. The authors found that 88% filled in the conditional with "If applicants have high SAT scores, then they will be accepted"—that is, the participants mentioned the rare rather than the common event. Another group was presented with the same task, but the college was said to be a local one that did not receive many applications and admitted most applicants. "Accepted" and "rejected" were merely reversed in the above scenario, as were "high" and "low." Everything else was the same. Now only 35% filled in the conditional with "If applicants have high SAT scores, then they will be accepted." Most of the participants targeted the rare events, "If applicants have low SAT scores, then they will be rejected." Thus, whether particular events were

**Table 2**  
Example of a Scenario Used by McKenzie, Ferreira, Mikkelsen, McDermott, and Skrable (2001)

Applicant	GPA	SAT Scores	Letters of			Extracurricular	Application Outcome
			Recommendation	Interview			
Alice	Low	Low	High	Low	Low	Rejected	
Bill	Low	High	Low	Low	Low	Accepted	
Cindy	Low	Low	Low	Low	High	Rejected	
Dennis	Low	Low	Low	High	Low	Rejected	
Emily	High	Low	Low	Low	Low	Rejected	



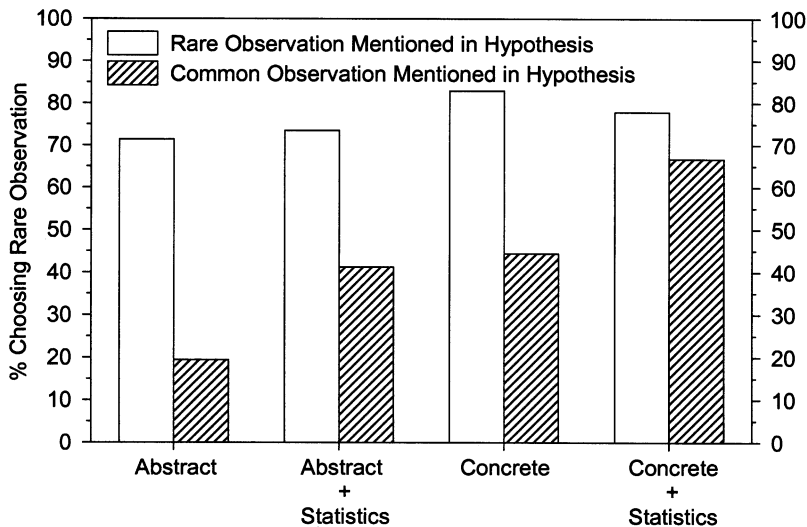
mentioned depended on whether they were rare. Similar results were found using other scenarios with different content.

In short, people appear to have a tendency to phrase conditional hypotheses in terms of rare rather than common events. This provides a rational explanation of the fact that “listeners” consider different data more supportive when hypotheses are rephrased in logically equivalent ways: How “speakers” choose to phrase hypotheses leaks normatively relevant information about event rarity. Mentioned confirming observations are seen as more informative than unmentioned confirming observations, probably because they usually *are* more informative.

**Making the framing effect disappear.** As with covariation assessment, the framing effect in hypothesis testing can be made to disappear by making clear which events are rare. When this is done, participants no longer need to rely on how hypotheses are phrased to infer event rarity and can instead focus on event rarity itself. McKenzie and Mikkelsen (2000) told additional participants testing each of the above hypotheses that few people have genotype *B* (and most have genotype *A*) and few have personality type *Y* (and most have personality type *X*)—information that, from a Bayesian perspective, makes the *B&Y* observation the most supportive because it is the rarest. As is shown in Figure 3, these “abstract + statistics” participants were about as likely as the “abstract” participants to select the *B&Y* observation when testing  $Y \rightarrow B$  (compare the first

two white columns). However, this is not too surprising because the *B&Y* observation is mentioned in the hypothesis in both cases. More interesting are the results when the *B&Y* observation was not mentioned in the hypothesis (hatched columns). As can be seen, the “abstract + statistics” group was about twice as likely as the “abstract” group to select the *B&Y* observation when testing “ $A \rightarrow X$ ” (compare the first two hatched columns). That is, participants were more likely to select the *unmentioned* observation if they were told that it was rare rather than if they were told nothing.

Of course, the framing effect did not go away; it was merely reduced. However, the participants were asked to test hypotheses that were abstract and unfamiliar. Even the rarity information provided was arbitrary and probably had little meaning for the participants. As with the covariation studies discussed earlier, one might expect that sensitivity to rarity would increase when participants were presented with familiar variables that tap into their real world knowledge regarding rarity. To test this, McKenzie and Mikkelsen (2000) told additional participants to imagine they were researchers examining a possible relation between mental health and AIDS. These participants tested one of two *concrete* hypotheses: “If a person is HIV+, then he/she is psychotic” and “If a person is mentally healthy, then he/she is HIV-.” They then selected whether a person who is HIV+ and psychotic or a person who is HIV- and mentally healthy provided



**Figure 3.** Results of the hypothesis-testing study (McKenzie & Mikkelsen, 2000). Shown is the percentage of participants selecting the rare observation as a function of (1) whether the task was abstract or concrete, (2) whether statistical information about rarity/commonality was provided, and (3) whether the rare observation was mentioned in the hypothesis. (The “abstract” group had no information about rarity.) Generally, the participants were more likely to correctly select the rare observation as more informative when (1) the task was concrete, (2) statistical information was provided, and (3) the rare observation was mentioned in the hypothesis. Most interesting is that the participants in the “concrete + statistics” group (far right) often selected the rare observation regardless of whether it was mentioned in the hypothesis, showing that the framing effect virtually disappeared for this group.

stronger support for the hypothesis that they were testing. Again, the two hypotheses are logically equivalent and both observations support both hypotheses. However, the HIV+/psychotic observation is relatively rare, and participants presumably knew this. Figure 3 shows that when these “concrete” participants tested “mentally healthy → HIV–,” almost half of them selected the rare HIV+/psychotic person (hatched column). That is, the *unmentioned* observation was often seen as more supportive if it was rare.

A final group of participants was given one of the two concrete hypotheses to test, but was reminded that few people are HIV+ (and most are HIV–) and that few are psychotic (and most are mentally healthy). Figure 3 shows that almost 70% of these “concrete + statistics” participants testing “mentally healthy → HIV–” selected the HIV+/psychotic person—the *unmentioned* observation—as more supportive (hatched column). Note that, regardless of which hypothesis they were testing, the “concrete + statistics” participants were about equally likely to select the HIV+/psychotic person. When real-world knowledge was combined with a validation of their beliefs about rarity, the participants preferred the normatively more supportive rare observation, regardless of whether it was mentioned in the hypothesis. In other words, the framing effect virtually disappeared.

In short, the information leakage account provides rational explanations of why framing effects occur in hypothesis testing tasks (i.e., because how hypotheses are phrased leaks normatively relevant information about event rarity) and why they disappear (i.e., because when event rarity is known, listeners no longer need to infer it on the basis of how the hypothesis is phrased).

## SUMMARY AND IMPLICATIONS

Two common phenomena in inference tasks can lead to framing effects. In covariation assessment, the fact that joint presence affects judgments more than joint absence can lead participants to judge the relation between two variables differently when presented with logically equivalent descriptions of data. In hypothesis testing, the fact that confirming observations mentioned in hypotheses are perceived as more informative than confirming observations not mentioned can lead participants to deem different observations more informative when presented with logically equivalent hypotheses. These phenomena, which have not previously been interpreted in terms of framing effects, have nonetheless traditionally been seen as normative errors or biases (but see Anderson, 1990; Klayman & Ha, 1987). However, through the use of the information leakage account of framing effects developed elsewhere (Sher & McKenzie, 2004; see also McKenzie & Nelson, 2003), a rational explanation of these purportedly irrational behaviors emerges: Although logically equivalent, the rephrased data and hypotheses are not information equivalent. In particular, the different descriptions leak information about which events are rare,

and event rarity is normatively relevant in these tasks (at least from a Bayesian perspective). In covariation assessment, observing the joint presence of variables has a larger impact on judgment, apparently because participants assume that the presence of variables is rare (Anderson, 1990; Anderson & Sheu, 1995; McKenzie & Mikkelsen, in press). In hypothesis testing, mentioned confirming observations are considered more supportive apparently because people assume that hypotheses are phrased in terms of rare events (McKenzie, Ferreira, et al., 2001; McKenzie & Mikkelsen, 2000; see also Einhorn & Hogarth, 1986; Klayman & Ha, 1987; Mackie, 1974; Oaksford & Chater, 1994, 1996). In both cases, participants’ inferences about event rarity appear reasonable and, because in these tasks rare events are normatively more informative than common ones, the information leakage account provides a rational explanation of the framing effects.

Bolstering the information leakage account is the fact that the framing effects all but disappear when it is made clear to participants which events are rare. Under these circumstances, participants no longer need to infer what is rare on the basis of how events are described or which events are mentioned in the hypothesis. When event rarity is known, well-known “biases” can be reversed: If it is clear that the absence of variables is rare, participants’ judgments of covariation are influenced more by the variables’ joint absence than by their joint presence (McKenzie & Mikkelsen, in press), and if it is clear that the hypothesis to be tested mentions common events, then participants consider unmentioned observations more informative than mentioned observations (McKenzie & Mikkelsen, 2000). The fact that the framing effects disappear under predictable conditions shows that the information leakage account of these inferential phenomena is not merely a post hoc rationalization of irrational behavior.

These observations broaden the information leakage account of framing effects. Previously, the account has focused primarily on attribute frames’ leaking reference point information (although Sher & McKenzie, 2004, suggested a more general hypothesis). Now it is clear that the account extends beyond that: Logically equivalent descriptions, or frames, in inference tasks leak normatively relevant information about event rarity.

Seeing the covariation and hypothesis testing findings as framing effects is useful because it facilitates the application of the information leakage approach to the inference tasks. That is, lessons learned from the framing literature were applied to the inference literature. It is also possible that lessons learned from the inference tasks can be applied to more traditional framing tasks. For example, it was argued that framing effects occurred in the inference tasks because frames leak information about event rarity. However, when participants already knew which events were rare and therefore had no need to infer such information, the framing effects virtually disappeared. Perhaps attribute framing effects could be attenuated in a similar way. Attribute framing effects appear to

occur because listeners infer reference points from the speaker's choice of frame, at least in part. It is conceivable that when listeners already have well established reference points and, therefore, have no (or less) need to infer the speaker's reference point, attribute framing effects will be reduced. Regardless of whether or not additional insights can be gleaned from viewing inferential phenomena as framing effects, the fact that the information leakage approach ties together traditional research on framing and research on inference is noteworthy. It is unusual for such seemingly disparate tasks to have a common thread, as is proposed here.

The extent to which information leakage can explain framing effects even more generally (e.g., risky choice framing effects) remains an open question. Whatever the scope of the information leakage account may be, it is important to note that it illuminates framing phenomena not just from a psychological perspective but also from a normative perspective. When the account is applicable—and it may be applicable to a wide range of framing tasks—participants are normatively justified in behaving differently when presented with different frames. Rather than indicate biased or irrational behavior, many framing phenomena might instead indicate behavior that is both rational and psychologically richer than was previously thought.

## REFERENCES

- ALLAN, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, **15**, 147-149.
- ALLAN, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin*, **114**, 325-448.
- ANDERSON, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- ANDERSON, J. R., & SHEU, C.-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition*, **23**, 510-524.
- ARKES, H. R., & HARKNESS, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, **112**, 117-135.
- CHENG, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, **104**, 367-405.
- CHENG, P. W., & NOVICK, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality & Social Psychology*, **58**, 545-567.
- CHENG, P. W., & NOVICK, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, **99**, 365-382.
- CROCKER, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, **90**, 272-292.
- CROCKER, J. (1982). Biased questions in judgment of covariation studies. *Personality & Social Psychology Bulletin*, **8**, 214-220.
- EINHORN, H. J., & HOGARTH, R. M. (1986). Judging probable cause. *Psychological Bulletin*, **99**, 3-19.
- EVANS, J. ST. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum.
- FISCHHOFF, B., & BEYTH-MAROM, R. (1983). Hypothesis testing from a Bayesian perspective. *Psychological Review*, **90**, 239-260.
- GIGERENZER, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology*, **2**, 83-115.
- GIGERENZER, G., HOFFRAGE, U., & KLEINBÖLTING, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, **98**, 506-528.
- HILGARD, E. R., & BOWER, G. H. (1975). *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- HILTON, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, **118**, 248-271.
- HOLLAND, J. H., HOLYOAK, K. J., NISBETT, R. E., & THAGARD, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
- HORWICH, P. (1982). *Probability and evidence*. Cambridge: Cambridge University Press.
- INHELDER, B., & PIAGET, J. (1958). *The growth of logical thinking: From childhood to adolescence*. New York: Basic Books.
- JENKINS, H. M., & WARD, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General & Applied*, **79**(1, Whole No. 594).
- JOHNSON-LAIRD, P. N. (1968a). The choice of the passive voice in a communicative task. *British Journal of Psychology*, **59**, 7-15.
- JOHNSON-LAIRD, P. N. (1968b). The interpretation of the passive voice. *Quarterly Journal of Experimental Psychology*, **20**, 69-73.
- KAHNEMAN, D., & TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263-291.
- KAHNEMAN, D., & TVERSKY, A. (1984). Choices, values, and frames. *American Psychologist*, **39**, 341-350.
- 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, & Cognition*, **19**, 1363-1386.
- KLAYMAN, J. (1995). Varieties of confirmation bias. In D. L. Medin (Series Ed.) & J. Busemeyer & R. Hastie (Vol. Eds.), *The psychology of learning and motivation* (Vol. 32, pp. 385-418). San Diego: Academic Press.
- KLAYMAN, J., & HA, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 211-228.
- KOTOVSKY, K., HAYES, J. R., & SIMON, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, **17**, 248-294.
- KÜHBERGER, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior & Human Decision Processes*, **75**, 23-55.
- LEVIN, I. P., & GAETH, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, **15**, 374-378.
- LEVIN, I. P., SCHNEIDER, S. L., & GAETH, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior & Human Decision Processes*, **76**, 149-188.
- LEVIN, I. P., SCHNITZLER, S. K., & THEE, S. L. (1988). Information framing effects in social and personal decisions. *Journal of Experimental Social Psychology*, **24**, 520-529.
- LEVIN, I. P., WASSERMAN, E. A., & KAO, S.-F. (1993). Multiple methods for examining biased information use in contingency judgments. *Organizational Behavior & Human Decision Processes*, **55**, 228-250.
- LIPE, M. G. (1990). A lens-model analysis of covariation research. *Journal of Behavioral Decision Making*, **3**, 47-59.
- MACKIE, J. L. (1963). The paradox of confirmation. *British Journal for the Philosophy of Science*, **13**, 265-277.
- MACKIE, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Oxford University Press, Clarendon Press.
- 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.
- MARTEAU, T. M. (1989). Framing of information: Its influence upon decisions of doctors and patients. *British Journal of Social Psychology*, **28**, 89-94.
- MCKENZIE, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, **26**, 209-239.
- MCKENZIE, C. R. M. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 771-792.
- MCKENZIE, C. R. M. (1999). (Non)complementary updating of belief in two hypotheses. *Memory & Cognition*, **27**, 152-165.

- 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). Oxford: Blackwell.
- McKENZIE, C. R. M., & AMIN, M. B. (2002). When wrong predictions provide more support than right ones. *Psychonomic Bulletin & Review*, *9*, 821-828.
- McKENZIE, C. R. M., FERREIRA, V. S., MIKKELSEN, L. A., McDERMOTT, K. J., & SKRABLE, R. P. (2001). Do conditional hypotheses target rare events? *Organizational Behavior & Human Decision Processes*, *85*, 291-309.
- 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. (in press). A Bayesian view of covariation assessment. *Cognitive Psychology*.
- McKENZIE, C. R. M., & NELSON, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, *10*, 596-602.
- McKENZIE, C. R. M., WIXTED, J. T., NOELLE, D. C., & GYURJAN, G. (2001). Relation between confidence in yes-no and forced-choice tasks. *Journal of Experimental Psychology: General*, *130*, 140-155.
- MEYEROWITZ, B. E., & CHAIKEN, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality & Social Psychology*, *52*, 500-510.
- MOXEY, L. M., & SANFORD, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, *14*, 237-255.
- MYNATT, C. R., DOHERTY, M. E., & TWENEY, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85-95.
- NICKERSON, R. S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking & Reasoning*, *2*, 1-31.
- NICKERSON, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220.
- OAKSFORD, M., & CHATER, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608-631.
- OAKSFORD, M., & CHATER, N. (1996). Rational explanation of the selection task. *Psychological Review*, *103*, 381-391.
- OAKSFORD, M., & CHATER, N. (2003). Optimal data selection: Revision, review, and re-evaluation. *Psychonomic Bulletin & Review*, *10*, 289-318.
- OAKSFORD, M., CHATER, N., & GRAINGER, B. (1999). Probabilistic effects in data selection. *Thinking & Reasoning*, *5*, 193-243.
- OAKSFORD, M., CHATER, N., GRAINGER, B., & LARKIN, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 441-458.
- SANFORD, A. J., FAY, N., STEWART, A., & MOXEY, L. (2002). Perspective in statements of quantity, with implications for consumer psychology. *Psychological Science*, *13*, 130-134.
- SCHUSTACK, M. W., & STERNBERG, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, *110*, 101-120.
- SCHWARZ, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Erlbaum.
- SHAKLEE, H., & TUCKER, D. (1980). A rule analysis of judgments of covariation between events. *Memory & Cognition*, *8*, 459-467.
- SHER, S., & McKENZIE, C. R. M. (2004). *Information leakage from logically equivalent frames*. Manuscript submitted for publication.
- SMEDSLUND, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, *4*, 165-173.
- SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- SNYDER, M. (1981). Seek and ye shall find: Testing hypotheses about other people. In E. T. Higgins, C. P. Heiman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium on personality and social psychology* (pp. 277-303). Hillsdale, NJ: Erlbaum.
- SNYDER, M., & SWANN, W. B., JR. (1978). Hypothesis testing in social interaction. *Journal of Personality & Social Psychology*, *36*, 1202-1212.
- TVERSKY, A., & KAHNEMAN, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.
- TVERSKY, A., & KAHNEMAN, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, S251-S278.
- WARD, W. C., & JENKINS, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, *19*, 231-241.
- WASON, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.
- WASSERMAN, E. A., DORNER, W. W., & KAO, S.-F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 509-521.
- WILSON, D. K., KAPLAN, R. M., & SCHNEIDERMAN, L. J. (1987). Framing of decisions and selections of alternatives in health care. *Social Behaviour*, *2*, 51-59.

## NOTES

1. Two other phenomena in higher-order cognition that might also be interpreted as framing effects are (1) different rates of logically correct responses in reasoning tasks depending on task content (e.g., in Wason's selection task; see Holland, Holyoak, Nisbett, & Thagard, 1986, pp. 266-273) and (2) differential solution rates for problem isomorphs differing only in terms of their cover stories (see, e.g., Kotovsky, Hayes, & Simon, 1985). These phenomena are not addressed in this article. However, it is worth noting that (1) Oaksford and Chater's (1994) account of a wide variety of results in Wason's selection task is clearly in general agreement with the approach taken in this article and (2) unlike the potential framing effects noted above, framing effects in the present article occur when the same content and the same cover story are used.

2. After calculating the cell values for the independent case ( $\rho = 0$ ), cell values for a different value of  $\rho$  can be calculated by multiplying that  $\rho$  value by its denominator (the square root of the product of the four marginal probabilities; see text for calculating  $\varphi$ ). Relative to the values in the independent case, add this product to the A and D values and subtract this product from the B and C values. Using the example in the text,  $.5 \times (.1^2 \times .9^2)^{1/2} = .045$ . Thus, the Cell A value for  $\rho = .5$  is  $.01 + .045 = .055$ , and the Cell D value is  $.81 + .045 = .855$ . Subtracting the product from the Cell B and Cell C values for the independent case yields  $.09 - .045 = .045$  for both cells. Calculating  $\rho$  with these new cell values yields 0.5. The new cell values are the only ones that will yield the desired level of  $\rho$  with these marginal probabilities.

3. The following is one way to think about whether the presence of properties is generally rarer than their absence: Imagine two terms, "X" and "not-X" (e.g., red things and nonred things, accountants and nonaccountants), where there is no simple, nonnegated term for "not-X." Which would be the larger category, X or not-X? The claim here is that not-X will be the larger category in the vast majority of cases.

4. Despite some similarities, hypothesis testing and covariation assessment tasks differ in potentially important ways. One is that the levels of the variables in hypothesis testing are often symmetric (e.g., introvert/extravert), whereas in covariation assessment they are traditionally asymmetric (e.g., treatment/no treatment). In addition, the task instructions are different. In hypothesis testing, participants are often asked to test "If X, then Y" statements, whereas in covariation assessment, participants are asked to assess a relationship between variables.

5. Some readers might object to applying a deductive logical analysis here because the hypotheses are likely to be interpreted as probabilistic rather than deterministic. However, to my knowledge there is also no straightforward inductive logical analysis that will render coherent the conclusions that (1)  $X1 \& Y1$  provides stronger support than  $X2 \& Y2$  when  $X1 \rightarrow Y1$  is tested, and (2)  $X2 \& Y2$  provides stronger support than  $X1 \& Y1$  when  $Y2 \rightarrow X2$  is tested.