THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (12), 2388–2408

Psychology Press
Taylor & Francis Group

Weighting of contingency information in causal judgement: Evidence of hypothesis dependence and use of a positive-test strategy

David R. Mandel
DRDC Toronto, Toronto, Ontario, Canada, and University of Toronto, Toronto, Ontario, Canada

Oshin Vartanian
DRDC Toronto, Toronto, Ontario, Canada

Contingency is an important cue to causation. Research shows that people unequally weight the cells of a 2 × 2 contingency table as follows: cause-present/effect-present (A) > cause-present/effect-absent (B) > cause-absent/effect-present (C) > cause-absent/effect-absent (D). Although some models of causal judgement can accommodate that fact, most of them assume that the weighting of information is invariant as a function of whether one is assessing a hypothesized generative versus preventive relationship. An experiment was conducted that tested the hypothesis-independence assumption against the predictions of a novel weighted-positive-test-strategy account, which predicts hypothesis dependence in cell weighting. Supporting that account, judgements of hypothesized generative causes showed the standard A > B > C > D inequality, but judgements of hypothesized preventive causes showed the predicted B > A > D > C inequality. The findings reveal that cell weighting in causal judgement is both unequal and hypothesis dependent.

Keywords: Causation; Prevention; Contingency information; Hypothesis testing; Positive-test strategy.

Hypothesis testing and causal reasoning are deeply intertwined aspects of human cognition. Many of the hypotheses that people generate are about cause–effect relationships, and, likewise, many of the judgements that people make about cause–effect relationships follow from a hypothesis about the existence of such a relationship. Hypotheses, of course, can be stated in either directional or nondirectional form. For example, a researcher might want to test the nondirectional

Correspondence should be addressed to David R. Mandel, Thinking, Risk, and Intelligence Group, Adversarial Intent Section, DRDC Toronto, 1133 Sheppard Avenue West, P.O. Box 2000, Toronto, Ontario M3M 3B9, Canada. E-mail: David.Mandel@drdc-rddc.gc.ca

This research was funded by Natural Sciences and Engineering Council of Canada Discovery Grant No. 249537–2002 to the first author. Portions of the present research were presented at the Southern Ontario Behavioral Decision Research Conference in Toronto, Canada, in May 2006 and at the 28th Annual Meeting of the Cognitive Science Society in Vancouver, Canada, in August 2006. We thank Joshua Klayman, Derek Koehler, Robin Murphy, and three anonymous reviewers for their insightful comments on the present research. We also thank Aaron Gregg for programming the experiment and Ahna Kim for her research assistance.

http://www.psypress.com/qjep DOI:10.1080/17470210902794148

hypothesis that noise levels influence performance on a cognitive task or, more specifically, the directional hypothesis that noise levels and performance are inversely related. Most hypotheses about causal relationships are of the latter type—namely, about whether a candidate cause *generates* a particular effect (or type of effect) or whether it *prevents* an effect. Indeed, as Kelley (1973) proposed in a landmark paper, the causal attributions people draw are often informed by their schemas regarding the opposing effects of generative and preventive causes. Hence, the occurrence of a behavioural effect in the presence of a potent, external preventive cause is likely to lead to an attribution of an internal, generative cause of equal or greater potency. A similar theme is evident in literature on causal deduction, which has emphasized the tension between generative causes and disabling conditions (e.g., Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991), as well as in literature on causal representation, which has recently emphasized the importance of perceived force dynamics (Wolff, 2007).

In the present article, we examine how the direction of a hypothesis about causation influences the manner in which different sources of information pertinent to testing the hypothesis are given weight by individuals. Literature on hypothesis testing suggests that the weighting of information may be influenced by the nature of the hypothesis being assessed. Based on his classic "rule-discovery" research, Wason (1960, 1968) proposed that people seek out evidence in a manner that places greater weight on confirming one's hypothesis rather than on disconfirming it. However, in an insightful reinterpretation of the "confirmation bias" phenomenon, Klayman and Ha (1987) proposed that people do not necessarily seek out or give greater weight to confirmatory information. Rather, they prefer to test hypotheses by examining cases that conform to their stated hypothesis. For example, in testing the hypothesis that a hidden rule defining a target set of numeric triplets is "even numbers increasing by two", people tend to select hypothesis-conforming triplets such as 4, 6, 8 or 12, 14, 16. They seldom select nonconforming triplets such as 1, 2, 3 or 8, 6, 4. Although the former, which Klayman and Ha call *positive hypothesis tests*, are hypothesis conforming, they are not necessarily hypothesis confirming. Similarly, although the latter, which Klayman and Ha call *negative hypothesis tests*, are hypothesis nonconforming, they are not necessarily hypothesis disconfirming.[1]

Although the conceptual link between hypothesis testing and causal reasoning is strong, theories of causal reasoning have, thus far, had relatively little to say about the effect of hypotheses on the weighting of information in causal judgement. Most descriptive accounts of causal judgement, in fact, assume that the integration of information operates in a hypothesis-independent manner. The majority of information integration models focus on the integration of 2×2 contingency information derived from crossing binary cause and effect variables in which the levels of each variable are *present* and *absent*. The four types of conjunctions, or *cells* as they are often called, are typically represented in a contingency matrix labelled as in Table 1.

The delta rule, which has been used by some theorists as both a descriptive and a normative model of causal judgement (e.g., Cheng & Novick, 1992; Jenkins & Ward, 1965), assigns equal weight to the cells as follows:

$$\begin{aligned}\Delta P &= P(y|x) - P(y|\neg x) \\ &= [A \div (A+B)] - [C \div (C+D)]. \quad (1)\end{aligned}$$

The delta rule is the probability of the effect, y, given the presence of the putative cause, x, minus the probability of the effect given the absence of the putative cause. The delta rule thus captures the notion that the effect of a putative cause should be more likely to occur in its presence

[1] Indeed, the rule-discovery paradigm does not permit individuals to seek confirmatory information. They can merely decide whether to test an example that either conforms or does not conform to their hypothesis. Then, they will learn whether the evidence revealed confirms or disconfirms their hypothesis, but over that outcome they have no choice.

Table 1. *Contingency matrix for hypothesized cause and target effect variables*

	Target effect	
Hypothesized cause	*Present (y)*	*Absent (¬y)*
Present (x)	Cell A $A = freq(x \bullet y)$	Cell B $B = freq(x \bullet \neg y)$
Absent ($\neg x$)	Cell C $C = freq(\neg x \bullet y)$	Cell D $D = freq(\neg x \bullet \neg y)$

than in its absence. The delta rule, however, does not change the weighting of the four cells depending on whether the hypothesis is that x generates y or that x prevents y from occurring. The same hypothesis-independent characteristic is evident in many other information integration rules that have been examined including the weighted delta rule (Anderson & Sheu, 1995), which assigns greater weight to $P(y|x)$ than to $P(y|\neg x)$; the sum of the diagonals rule (Schustack & Sternberg, 1981), which sums the frequencies of cells B and C and subtracts them from the sum of the frequencies of cells A and D (see also Busemeyer, 1991, for a joint-probability variant of this rule); and the (self-explanatory) $A - B$ rule (Inhelder & Piaget, 1958; Mandel & Lehman, 1998).

Some models of causal judgement are ostensibly hypothesis dependent. For instance, Cheng's (1997) power PC model may appear to be sensitive to the direction of a causal reasoner's hypothesis because there are different integration rules for generative and preventive causation. However, a careful reading of Cheng's (1997; see also Buehner, Cheng, & Clifford, 2003) account reveals that the factor determining which variant of the model to apply is based not on the hypothesized direction of the causal relationship, but rather on the actual direction as determined by the sign of the ΔP coefficient. Therefore, the power PC model, too, is hypothesis independent, although it could presumably be adapted as a hypothesis-dependent model by shifting the basis for determining the integration rule from the actual direction of the relationship to the hypothesized direction (a modification that we test in the experiment reported later on).

Another example is provided by theorists (Anderson, 1990; Anderson & Sheu, 1995; Hattori & Oaksford, 2007; McKenzie & Mikkelson, 2007) who have proposed that causal reasoners prefer to test hypotheses about rare events rather than common ones because the former are deemed to be more informative (see also Klayman & Ha, 1987; Oaksford & Chater, 2003). According to this Bayesian view, normative integration of contingency information in causal judgement does not require equal weighting of the four cells. When hypotheses are formulated in terms of events that are rare, the often-observed overweighting of Cells A and B is justified—indeed, more so than equal weighting. However, as with Cheng's (1997) account, variations in the weighting of cells predicted by Bayesian accounts depend on the nature of evidence—in Bayesian accounts, on which cases are in fact rare as opposed to common—rather than being dependent on the hypothesis entertained. Indeed, according to some Bayesian proponents (e.g., Hattori & Oaksford, 207; McKenzie, Ferreira, Mikkelson, McDermott, & Skrable, 2001), the manner in which people formulate hypotheses is sensitive to the rarity of events. Thus, although Bayesian accounts, like the power PC model, permit differential weighting under different *evidential* conditions, they too provide a fundamentally hypothesis-independent description of information weighting in causal judgement, except in the narrow sense that hypotheses may be determined by perceptions of category rarity.

A third example of an account of causal judgement that ostensibly posits hypothesis dependence is White's (2003a, 2003b, 2004) evidence evaluation model, according to which people integrate contingency data from the four cells using the proportion-of-confirming-instances (pCI) rule:

$$p\text{CI} = (A + D) \div (A + B + C + D). \quad (2)$$

The evidence evaluation model posits that people think about information relative to a hypothesis and code it as supportive or nonsupportive. However, the pCI rule itself invariably assigns

equal weight to the four cells. As proof, note that the probability of disconfirming cases, 1 − pCI, must be equally predictive of causal judgements given that the two rules yield values that sum to a constant. Thus, the evidence evaluation model precludes the differential weighting of the four cells based on characteristics of the hypothesis, such as its direction. White (2003b, 2004) also introduced a weighted variant of the pCI rule in which the four cells are individually weighted (also see Catena, Maldonado, & Cándido, 1998, for a similar proposed rule). Although the weighted pCI rule could better accommodate predictions based on hypothesis dependence because of its weaker constraints (i.e., it subsumes the unweighted rule as a special case), at present, the evidence evaluation model has yet to offer specific predictions of how cell weights would vary as a function of hypothesis direction.

Griffiths and Tenenbaum (2005) proposed a model for the calculation of causal support, which is indeed hypothesis dependent. Causal support, in their account, is defined by the value of the log likelihood ratio for obtaining data from G_C over G_I, where G_C is a graphical model in which the putative cause and background conditions are causally linked to the target effect, and G_I is a graphical model in which only the background conditions are linked to the effect. Causal support cannot be computed analytically from the cell frequencies but can be estimated using the Monte Carlo method. Griffiths and Tenenbaum (2005) made no explicit predictions regarding how inequalities in cell weighting might be expected to vary as a function of the direction of the focal hypothesis. Nor is it evident from an examination of their support model what inequalities in cell weighting it might imply. In the experiment we subsequently report, we tested whether the support model can predict the observed pattern of cell weights as a function of whether the hypothesis was about generative or preventive causation.

In contrast to these previous accounts and models, we develop an account and accompanying model that makes clear predictions about the hypothesis-dependent nature of contingency cell weighting in causal judgement. We hypothesized that people will differentially weight the four cells in making causal judgements depending on whether they are asked to assess the strength of either a hypothesized generative cause–effect relationship or a hypothesized preventive cause–effect relationship. Moreover, we anticipated that the differences in weighting as a function of hypothesis will be explicable largely in terms of a positive-test strategy (Klayman & Ha, 1987). Before outlining our specific predictions regarding the rank ordering of weights assigned to the four cells under generative and preventive hypothesis test conditions, we first review an account of cell weighting in causal judgement proposed by Mandel and Lehman (1998), which was designed to explain a robust pattern of inequality in cell weighting observed in previous studies. We contrast the predictions of that account with those of the weighted positive-test strategy account, which we subsequently develop.

The positive-event and sufficiency biases (PSB) account

Using a variety of experimental methods, several studies have shown that people tend to unequally weight the four cells, such that Cell A is given the greatest weight followed by Cell B, Cell C, and Cell D in descending rank order (e.g., Anderson & Sheu, 1995; Kao & Wasserman, 1993; Mandel & Lehman, 1998; Perales, Catena, Shanks, & González, 2005; Schustack & Sternberg, 1981; Wasserman, Dorner, & Kao, 1990; for a meta-analysis, see Perales & Shanks, 2007). Because the cell weight inequality is expected (i.e., statistical) rather than strict (i.e., deterministic), it may be appropriately expressed as follows:

$$E(\omega_{\mathrm{A}}) > E(\omega_{\mathrm{B}}) > E(\omega_{\mathrm{C}}) > E(\omega_{\mathrm{D}}), \qquad (3)$$

where ω_i refers to the weight of the cell specified by the subscript, and the sign of the cells conveying disconfirmatory information is reversed (i.e., multiplied by −1).

Although several studies provide evidence of the cell weight inequality, and some researchers have attempted to parameterize information integration rules to better account for this inequality (e.g., Anderson & Sheu's, 1995, weighted ΔP rule; White's, 2003b, 2004, weighted *p*CI rule), few researchers have tried to explain why the inequality occurs in the first place. Mandel and Lehman (1998) provided an explanation of the cell weight inequality in terms of two information-processing biases: The first, *positive-event bias*, refers to the tendency to seek out and give greater weight to information about event occurrences than nonoccurrences (e.g., Newman, Wolff, & Hearst, 1980; Wason, 1959). Positive-event bias implies that the A cell will be weighted most heavily because the putative cause and effect are both positive (i.e., both are present), followed next by the B and C cells, each of which is composed of one positive event and one negative event, followed lastly by the D cell, which is composed of two negative events (i.e., both are absent). More formally, positive-event bias implies the following inequality:

$$E(\omega_A) > E(\omega_B) \approx E(\omega_C) > E(\omega_D). \quad (4)$$

The second, *sufficiency bias*, reflects a preference for testing whether a putative cause is sufficient rather than necessary to yield the target effect (e.g., Friedrich, 1993; Mandel, 2003; Mandel & Lehman, 1996; Wolff, 2007). It implies that people focus more on Cells A and B than Cells C and D because the former invariably inform assessments of $P(y|x)$, which probabilistically gauge the sufficiency of the putative cause regardless of whether the hypothesized relation is generative or preventive. By comparison, the contrast of Cells C and D invariably informs assessments of $P(y|\neg x)$, which probabilistically gauge the necessity of the putative cause. Hence, the following inequality is predicted on the basis of sufficiency bias:

$$E(\omega_A) \approx E(\omega_B) > E(\omega_C) \approx E(\omega_D). \quad (5)$$

Taken together, the two biases can account for the full pattern of inequality expressed in (3).

Because positive-event bias is hypothesis independent, Mandel and Lehman's (1998) positive-event and sufficiency biases (PSB) account predicts that people assign more weight to Cell A than Cell B regardless of whether the hypothesized causal relation is generative or preventive. Moreover, the PSB account proposes that the $E(\omega_A) > E(\omega_B)$ inequality will be greater if the focal hypothesis pertains to preventive causation (h^-) rather than generative causation (h^+). That is, the PSB account makes the following prediction:

$$[E(\omega_A) - E(\omega_B)]|h^- > [E(\omega_A) - E(\omega_B)]|h^+. \quad (6)$$

This prediction derives from the fact that, given h^-, A indexes the frequency of sufficiency violations in contrast with either B or C, both of which offer support for the hypothesis that x is sufficient to prevent y. In contrast, given h^+, B indexes the frequency of sufficiency violations in contrast with either A or D, both of which offer support for the hypothesis that x is sufficient to generate y. That is, the hypothesis, "x is sufficient to generate y", can only be disconfirmed by cases in the B cell, whereas the hypothesis, "x is sufficient to prevent y", can only be disconfirmed by cases in the A cell. Likewise, only cases in the C cell can disconfirm the hypothesis, "x is necessary to generate y", whereas only cases in the D cell can disconfirm the hypothesis, "x is necessary to prevent y".[2] The prediction that the weighting of the A cell should be even greater in hypothesis tests of prevention than in hypothesis tests of causation, however, has yet to be rigorously tested. Doing so would require manipulating the direction of the hypothesized causal relationship while

[2] This description focuses only on factual causal hypotheses in which the antecedent, x, is affirmed. An extension of this analysis to counterfactual causal hypotheses (e.g., "$\neg x$ might have been sufficient to prevent y") is given in Mandel (2005).

holding constant the set of cases presented to participants as evidence. The experiment reported later on conducted this critical test.

The weighted positive-test strategy (WPS) account

We pitted the PSB account's predictions against a novel weighted positive-test strategy (WPS) account.[3] The basic premise of the WPS account is that inequalities in cell weighting are explicable in terms of a positive-test strategy (Klayman & Ha, 1987; Levin, Wasserman, & Kao, 1993). In the present context, a positive-test strategy refers to the tendency to assign greater weight to information that conforms to the hypothesized cause (a *positive hypothesis test* or +H test) or the target effect (a *positive target test* or +T test) specified in a focal hypothesis. Cases in which the hypothesized generative or preventive cause, *x*, is present (cases in Cells A and B) provide the relevant data for conducting +H tests, whereas cases in which the hypothesized generative or preventive cause, *x*, is absent (cases in Cells C and D) provide the relevant data for conducting a *negative hypothesis test* (−H test). If the hypothesis is about a generative causal relationship (h^+), then the target event is *y*, and, accordingly, the A and C cells provide the relevant data for conducting +T tests. However, if the hypothesis is about a preventive causal relationship (h^-), then the target event is $\neg y$, and, accordingly, the B and D cells provide the relevant data for conducting +T tests. By comparison, a *negative target test* (−T test) relies on a contrast of Cells B and D given h^+ and on a contrast of Cells A and C given h^-. The case of h^- also illustrates that the WPS account differs from one based on a simple case of matching bias (Evans, 1998). A matching bias occurs when information is selected or weighted on the basis of whether it matches those events that are explicitly described in a hypothesis or query. In the present context, a matching-bias account would predict that both the hypotheses "*x* causes *y*" (h^+) and "*x* prevents *y*" (h^-) would elicit a predominant focus on the A cell because both explicitly describe the events *x* and *y*, the conjunction of which only the A cell represents.

A positive-test strategy that ascribes *equal* weight to +H and +T tests implies that people testing hypotheses about generative causation will weight the four cells in accordance with (4) because the A cell is relevant to both types of positive test, the B cell is relevant to +H tests but not to +T tests, the C cell is relevant to +T tests but not to +H tests, and the D cell is relevant to neither positive test. This can be restated more precisely as the following conditional inequality:

$$[E(\omega_A) > E(\omega_B) \approx E(\omega_C) > E(\omega_D)]|h^+. \quad (7)$$

However, unlike (7), a positive-test strategy that ascribes equal weight to +H and +T tests also implies that people testing hypotheses about a preventive causal relation, h^-, will tend to assign weight to the four cells as follows:

$$[E(\omega_B) > E(\omega_A) \approx E(\omega_D) > E(\omega_C)]|h^-. \quad (8)$$

This is predicted because the A cell is relevant to +H tests but not to +T tests, the B cell is relevant to both types of positive test, the C cell is relevant to neither type of positive test, and the D cell is relevant to +T tests but not to +H tests. Recently, Hattori and Oaksford (2007) presented two models that instantiate precisely this strategy and that both predict the conditional inequalities specified in (7) and (8). One model (Model 1 in their Table 2, p. 771), which they call the dual-factor heuristic, is a geometric mean of +H and +T. Another model, which they call the positive-test index (Model 15 in their Table 2, p. 771), is an arithmetic mean of +H and +T.

[3] In an earlier proceedings report (Mandel & Vartanian, 2006), we presented a slightly different account that focused on the joint influence of positive testing and confirmation bias. That account, however, is less parsimonious than the present account and does not lead to a key prediction: namely, that Cell D will receive greater weight than Cell C in tests of preventive causal relationships.

Table 2. *Frequency distributions of stimuli and factor levels for set size = 10*

A	*B*	*C*	*D*	Δ*P*	*Contingency*	*Dominance*
2	5	2	1	− .38	Negative	AB
1	2	5	2	− .38	Negative	CD
1	5	2	2	− .33	Negative	AB
2	2	5	1	− .33	Negative	CD
2	5	1	2	− .05	Near-zero	AB
2	1	5	2	− .05	Near-zero	CD
5	2	2	1	.05	Near-zero	AB
1	2	2	5	.05	Near-zero	CD
5	1	2	2	.33	Positive	AB
2	2	1	5	.33	Positive	CD
5	2	1	2	.38	Positive	AB
2	1	2	5	.38	Positive	CD

We propose, however, that it is unlikely that people would assign equal importance to hypothesis and target tests because, in the former case, the effect depends on the occurrence of the cause. Therefore, it is congruent with a basic feature of causal schemas—namely, that causes precede their effects (Einhorn & Hogarth, 1986; Tversky & Kahneman, 1980). In contrast, that intuition about the arrow of time in causal relations is turned on its head with +T tests because they examine the dependence of the cause on the presence of the effect. Tversky and Kahneman (1980) demonstrated that when participants were asked to compare the conditional probabilities of a cause given the occurrence of its effect and of an effect given the occurrence of its cause, and when these probabilities were of equal magnitude, most participants nevertheless assessed the probability that was congruent with their causal schema to be greater than the probability that was incongruent with the causal schema. They also showed that when the same event could be used as either a causal or diagnostic cue, the vast majority of participants used the cue in a manner consistent with the relevant causal schema in which effects depend on causes. On the basis of these studies, they argued that the fluency of causal thinking inhibits the diagnostic thinking. Using the present nomenclature, this means that H tests will tend to be weighted more heavily than *t* tests. Accordingly, we propose that people will tend to give more weight to +H tests than to +T tests in assessing causation via a positive-test strategy due to the congruence of the former, but not the latter, with people's causal schemas.

As a conservative test of the WPS account, we adopt the following simplifying assumptions: First, only positive tests receive weight in causal judgement. Second, +H tests receive greater weight than +T tests. Third, the two cells implicated in a particular test are assigned equal weight. Taken together, these assumptions may be expressed as follows:

$$E(\omega_{+\mathrm{H}}) > E(\omega_{+\mathrm{T}}) > E(\omega_{-\mathrm{H}}) \approx E(\omega_{-\mathrm{T}}) \approx 0, \quad (9)$$

where ω_i refers to the weight of the test strategy indicated by the subscript. In turn, (9) implies the following two predictions of hypothesis-dependent cell weighting: First, when testing hypotheses about generative causation, people will weight the four cells in accordance with the typically observed cell weight inequality:

$$[E(\omega_{\mathrm{A}}) > E(\omega_{\mathrm{B}}) > E(\omega_{\mathrm{C}}) > E(\omega_{\mathrm{D}})]|h^{+}. \quad (10)$$

Second, when testing hypotheses about preventive causation, people will weight the four cells in accordance with the following conditional prediction:

$$[E(\omega_{\mathrm{B}}) > E(\omega_{\mathrm{A}}) > E(\omega_{\mathrm{D}}) > E(\omega_{\mathrm{C}})]|h^{-}. \quad (11)$$

A study by Crocker (1982) provides preliminary support for these predictions. She asked participants to indicate which of the four cells were jointly necessary and sufficient to consider in assessing the contingency between two binary variables. If the hypothesis was stated in terms of a positive relation, participants assigned importance to the cells in the typical A > B > C > D pattern. In contrast, if the hypothesis was stated in terms of a negative relation, participants assigned importance in line with the B > A > D > C pattern predicted by the WPS account. Critically, however, Crocker (1982) only examined self-reported information use and did not ask her participants to make

either causal or contingency judgements. Thus, it is unclear whether Crocker (1982) would have found a comparable pattern of cell weight inequalities based on participants' judgements of causal strength. Indeed, given that the correspondence between beliefs regarding cell importance and cell weighting inferred from judgement was low in past studies (see Mandel & Lehman, 1998, Experiment 2; Wasserman et al., 1990, Experiment 3), her findings could, at best, be viewed as indicative of support for the WPS account. Moreover, unlike the present research, Crocker also did not explicitly manipulate the direction of the hypothesized relation. Rather, she varied whether the target outcome was a positive event (winning) or a negative event (losing). The present research, which relies on an analysis of causal judgement data and a clear manipulation of hypothesis direction, therefore provides a vital test of the PSB and WPS accounts.

The present research

We conducted an experiment designed to pit the predictions of these two explanatory accounts (as well as the information integration rules noted earlier that make hypothesis-dependent predictions) of cell weighting in causal induction. Participants were presented with contingency information and were then asked to rate the extent to which an antecedent event (viz., a viral or antiviral agent) either causes or prevents a consequent event (viz., an illness). That is, half our sample assessed hypotheses of the form "the viral agent causes the illness", and the other half assessed hypotheses of the form "the antiviral agent prevents the illness". As noted earlier, we predicted, in line with the WPS account, that participants in the cause condition will weight the cells in line with (10), and participants in the prevent condition will weight the cells in line with (11). Following Mandel and Lehman's (1998) feature analytic approach, we assessed support for the predictions of the alternative accounts by examining the degree to which the frequencies in the four cells covaried with participants' judgements in the cause and prevent conditions.

To increase the robustness of our test, we also manipulated the format in which contingency information was presented. Crossed with our manipulation of hypothesis, half the sample received information in trials format (i.e., one case at a time), and the other half received information in summary format (i.e., they were given the frequency for each of the four cells in the form of propositional statements; for a comparison of the effect of different summary format methods on causal judgement, see Vallée-Tourangeau, Payton, & Murphy, 2008). The trials and summary formats are widely used methods for presenting contingency information in noninterventional causal induction research—namely, in causal induction studies in which participants learn about contingencies through observation of discrete cases in which cause and effect are each present or absent rather than through direct intervention, as they would in the free-operant or continuous tasks (Hattori & Oaksford, 2007; Pearl, 2000; Sloman & Lagnado, 2005). Some research indicates that participants tend to adopt more complex information integration strategies, such as calculating ΔP, when information is presented in summary format (Kao & Wasserman, 1993; Ward & Jenkins, 1965). This effect has been attributed to a variety of factors, including reductions in memory demands (Kao & Wasserman, 1993; Shaklee & Mims, 1982; Yates & Curley, 1986), ambiguity concerning how event pairs are defined (Wasserman & Shaklee, 1984), and frequency estimation errors (Wasserman & Shaklee, 1984). Nevertheless, the cell weight inequality is evident in studies using both formats (cf. Anderson & Sheu, 1995; Kao & Wasserman, 1993; Mandel & Lehman, 1998; Schustack & Sternberg, 1981). Therefore, we hypothesized that support for the predictions of the WPS account would be robust across the two main discrete formats employed in causal induction research.

Method

Participants

Participants were 80 undergraduates from a Canadian university who received partial course credit.

Design

Participants were randomly assigned in equal numbers to one of four conditions in a 2 (hypothesis: cause, prevent) × 2 (format: summary, trials) factorial design. They judged the strength of the hypothesis that "*x* causes *y*" in the cause condition and that "*x* prevents *y*" in the prevent condition. Participants were presented with frequency totals for each of the four cells in the summary condition, and identical information was presented on a trial-by-trial basis in the trials condition. Three additional within-subject factors—set size (10 or 20 cases), contingency (moderately-negative, near-zero, or moderately-positive), and dominance (i.e., controlling for contingency, whether cases were predominantly in Cells A and B or Cells C and D)—were manipulated through the construction of the stimulus set presented to participants. These factors are described in detail later on in the *Stimuli* subsection.

Procedure

Participants in the cause condition were asked to imagine that they were a researcher trying to find the cause of a presently incurable illness and that their task was to test 24 distinct viral agents that might be causes of the illness. Participants read that 24 patient samples were drawn from the general population to test each virus. Participants in the prevent condition received a similar cover story but were told that their task was to test 24 distinct antiviral agents that might prevent the illness.

In the summary condition, each conjunction was described in a sentence followed by the corresponding cell frequency following the procedure used in Mandel and Lehman (1998, Experiment 1). For example, in the cause condition, Cell A was presented as "The virus was present, and the patient had myonic fever—*n*". Information order was counterbalanced such that half received the ABCD order, and half received the DCBA order. In the trials condition, information was presented one case at a time in an order randomly generated for each participant. For each case, two circles appeared on a computer screen. The left circle indicated that the patient either had or had not been exposed to the hypothesized causal agent by stating inside the circle either "(ANTI)VIRUS PRESENT" or "(ANTI)VIRUS ABSENT", respectively. Similarly, the right circle indicated that the patient either had or did not have the illness by stating either "MYONIC FEVER PRESENT" or "MYONIC FEVER ABSENT", respectively. In addition, circles conveying presence were coloured green, and circles conveying absence were coloured red. Participants advanced through the trials self-paced by mouse clicking a button until the judgement task appeared. Participants in the cause condition were asked to rate the degree to which the relevant virus brought about the illness by indicating whether the virus was: *not at all a cause* (0), *a weak cause* (1), *a moderate cause* (2), or *a strong cause* (3). Participants in the prevent condition were asked to rate the degree to which the relevant antivirus prevented the illness by indicating whether the virus was: *not at all preventive* (0), *weakly preventive* (1), *moderately preventive* (2), or *strongly preventive* (3).[4]

Stimuli

The 24 "samples" presented to participants were drawn from Mandel and Lehman (1998, Experiment 1). The stimulus set is well designed for assessing differential cell weighting because each cell has the same mean and variance ($M = 3.75$, $SD = 2.74$). Moreover, the correlation between cell frequencies is constant across cell pairs and near zero ($r = -.04$). In other words, the manipulation of cell frequencies is virtually orthogonal. Cell importance can therefore be estimated by examining the absolute value of the Fisher-transformed correlation between the frequencies of a given cell and participants' causal judgements.

Table 2 shows the cell frequencies and corresponding ΔP coefficients for the twelve 10-case

[4] Following the judgement task, participants completed another task unrelated to the hypotheses tested in this article.

samples, along with the levels of contingency and dominance to which they correspond. The same twelve patterns were also reproduced as 20-case samples. Our interest in doing so was primarily to increase the number of observations for our correlational analyses of cell weighting, although it also allowed us to assess whether there was a set size effect on causal judgements. As Table 2 shows, half of the samples had negative ΔP coefficients, and the other half had mirror-image positive ΔP coefficients. We established the stimulus set in this manner so that the direction of the actual contingencies presented to participants would be orthogonal to the direction of the hypothesis they were asked to assess. As is evident from observing the magnitudes of the ΔP coefficients, the stimuli can, however, be summarized in terms of a 3 (contingency: moderately negative, near zero, and moderately positive) $\times$ 2 (set size: 10, 20) $\times$ 2 (dominance: AB, CD) within-subjects design. As Table 2 shows, contingency is moderately negative when ΔP equals $-.33$ or $-.38$, near zero when ΔP equals $\pm .05$, and moderately positive when ΔP equals .33 or .38. Each specific value of ΔP was instantiated in two stimuli, one of which has the majority of cases in Cells A and B (AB) and the other of which has the majority of cases in Cells C and D (CD). This factor, as noted earlier, is called dominance. We use the factors of set size, dominance, and contingency, however, only in our preliminary analyses in order to characterize the relation between stimulus characteristics and participants' causal judgements. These factors are not central to tests of our key hypotheses.

Results

Variance partitioning of raw judgements

We began by conducting an exploratory analysis of the effects of our five independent variables on participants' raw judgements. Recall that two of these variables (i.e., hypothesis and format) were manipulated between subjects, and the remaining three (i.e., set size, contingency, and dominance) were manipulated within subjects. Thus, we conducted a five-way mixed-model analysis of variance (ANOVA) on the judgement data. We applied the Bonferroni correction and report only those effects for which the corrected alpha level is less than .05 (we therefore do not report uncorrected p values). We found significant main effects of hypothesis, $F(2, 75) = 20.78$, $\eta_p^2 = .36$, set size, $F(2, 75) = 10.86$, $\eta_p^2 = .23$, and dominance, $F(2, 75) = 33.21$, $\eta_p^2 = .47$. We also found the following significant interaction effects: Hypothesis $\times$ Contingency, $F(4, 73) = 55.21$, $\eta_p^2 = .75$, Hypothesis $\times$ Dominance, $F(2, 75) = 15.76$, $\eta_p^2 = .30$, Hypothesis $\times$ Contingency $\times$ Set Size, $F(4, 73) = 6.60$, $\eta_p^2 = .27$, and Hypothesis $\times$ Contingency $\times$ Dominance, $F(4, 73) = 39.26$, $\eta_p^2 = .68$. The means and standard deviations of the judgement ratings as a function of the four factors that yielded significant effects (all but format) are shown in Table 3.[5]

Note that all of the significant main and second-order interaction effects were qualified by at least one of the two significant three-way interactions observed in this analysis. Both of the latter represent moderation of the Hypothesis $\times$ Contingency interaction effect by a third variable, which in one case is set size and in the other case is dominance. We therefore examined how these two factors moderated that common interaction effect. First, note that a Hypothesis $\times$ Contingency crossover interaction effect is expected on the basis that participants' mean judgements in the cause condition would increase as contingency varied from negative to positive and, conversely, that their mean judgements in the prevent condition would decrease as contingency varied from negative to positive. Based on past research showing a set-size effect of small magnitude in which causal judgements become more extreme as set size is increased controlling for ΔP (e.g., Anderson & Sheu, 1995; Shanks, 1985), we anticipate that the slopes in the two-way

[5] Given that we did not have other predictions to test involving format, we aggregate over format in Table 3.

Table 3. *Judgement ratings as a function of set size, dominance, contingency, and hypothesis*

			Hypothesis			
			Cause		*Prevent*	
Set size	*Dominance*	*Contingency*	*M*	*SD*	*M*	*SD*
10	AB	Negative	0.44	0.52	2.04	0.75
10	AB	Near-zero	1.24	0.52	1.25	0.61
10	AB	Positive	2.00	0.69	0.50	0.69
10	CD	Negative	0.65	0.84	1.28	0.77
10	CD	Near-zero	0.60	0.66	0.94	0.66
10	CD	Positive	0.93	0.76	0.89	0.84
20	AB	Negative	0.50	0.59	2.30	0.60
20	AB	Near-zero	1.45	0.64	1.30	0.65
20	AB	Positive	2.39	0.54	0.50	0.77
20	CD	Negative	0.78	0.81	1.54	0.68
20	CD	Near-zero	0.79	0.70	0.93	0.54
20	CD	Positive	1.34	0.61	0.80	0.69

interaction would be steeper as set size increases from 10 to 20 cases per stimulus. Figure 1, which plots mean judgements as a function of hypothesis, contingency, and set size, reveals that this prediction was in fact supported.

Turning then to moderating effect of dominance on the Hypothesis × Contingency interaction effect, the PSB and WPS accounts, in fact, both make the same prediction, which is that the steepness of the slopes will be greater in the AB condition than in the CD condition given that both accounts posit (for different reasons) that Cells A and B tend to be weighted more heavily than Cells C and D. Figure 2, which plots mean judgements as a function of hypothesis, contingency, and dominance, confirms this prediction. Indeed, it is evident that when the majority of cases were in the C and D cells, judgements were largely insensitive to the actual degree of contingency. Taken together, this preliminary analysis reveals a meaningful pattern of findings. We now proceed to more subtle tests designed to pit the predictions of the WPS account against its main competitors—the PSB account as well as information integration models, such as the positive test index and the dual-factor heuristic (Hattori & Oaksford, 2007), which are based on an unweighted positive-test strategy.

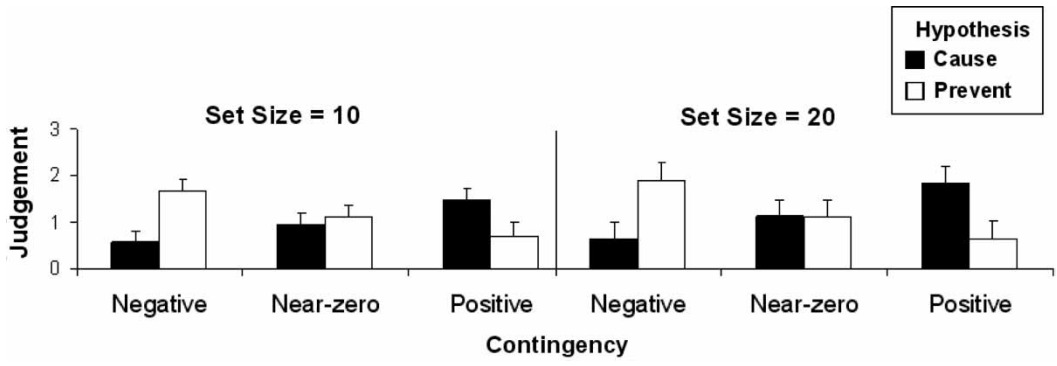

Figure 1. *Mean judgement as a function of hypothesis, contingency, and set size (T bars represent standard error of the mean).*

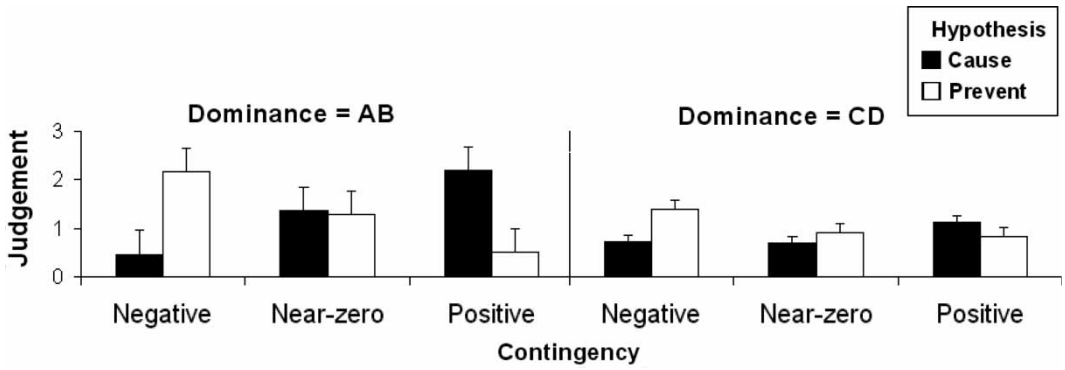

Figure 2. *Mean judgement as a function of hypothesis, contingency, and dominance (T bars represent standard error of the mean).*

Hypothesis dependence in cell weighting

Nomothetic analysis. In order to pit the WPS account against its competitors and test for hypothesis dependence in cell weighting, we calculated cell weights following Mandel and Lehman's (1998) feature-analytic procedure. That is, each participant's judgements were correlated with each of the four cell frequencies. The correlations were then normalized using the Fisher transformation, and the signs of the correlations involving the disconfirmatory cells (Cells B and C in the cause condition and Cells A and D in the prevent condition) were reversed (i.e., multiplied by −1). We analysed these derived cell weights using a three-way (Hypothesis × Format × Cell) mixed-model ANOVA, in which cell was a repeated measures factor. This analysis revealed significant main effects of format, $F(1, 76) = 8.99$, $p < .005$, $\eta_p^2 = .11$, and cell, $F(3, 228) = 74.77$, $p < .001$, $\eta_p^2 = .50$. Overall, the cells were weighted more heavily in the summary condition ($M = 0.36$, $SD = 0.10$) than in the trials condition ($M = 0.26$, $SD = 0.18$). That is, the cells were better predictors of participants' judgements in the summary condition than in the trials condition. However, format did not interact with the other factors. As shown in the bottom row of Table 4, the main effect of cell replicated the typical A > B > C > D inequality. This effect, however, was qualified by an interaction with hypothesis, $F(3, 228) = 33.87$, $p < .001$, $\eta_p^2 = .31$. As Table 4 shows, the pattern of weighting as a function of hypothesis and cell strongly supports the WPS account and is inconsistent with the PSB account. Whereas the A > B > C > D cell inequality captured by (10) was observed in the cause condition, the B > A > D > C cell inequality captured by (11) was observed in the prevent condition. Moreover, each of these six inequalities, all of which are predicted by the WPS account (i.e., A > B, B > C, and C > D cell inequalities in

Table 4. *Cell weights as a function of hypothesis and cell*

	Cell							
	A		*B*		*C*		*D*	
Hypothesis	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Cause	0.83	0.35	0.27	0.18	0.14	0.27	0.00	0.22
Prevent	0.37	0.28	0.65	0.39	0.03	0.28	0.14	0.20
Overall	0.61	0.40	0.47	0.36	0.09	0.29	0.06	0.22

Note: Within rows, all means differ significantly ($ps < .02$) by repeated contrasts for the simple-effect analysis of cell.

the cause condition and the B > A, A > D, and D > C cell inequalities in the prevent condition) was shown to be significant by an alpha-error inflation-correcting procedure (see note to Table 4). Thus, the current test also revealed that the WPS account provides a better account of the hypothesis-dependent nature of cell weighting in causal judgement than do models such as the positive-test index and dual-factor heuristic, which are both unweighted variants of a positive-test strategy that predict (7) and (8) instead of (10) and (11), as observed.

Idiographic analysis. The preceding nomothetic analyses showed that the WPS account does a good job of predicting participants' cell weighting as a function of hypothesis tested, *on average*. We also conducted idiographic analyses to assess whether the WPS account also does a good job of predicting participants' cell weighting as a function of hypothesis tested, *in general*. For each participant, we coded whether Cell A received greater, equal, or less weight than Cell B, and we repeated the procedure for the contrast of Cells B and C and Cells C and D. As Table 5 shows, invariably, the majority of participants' weightings conformed to the relational pattern predicted by the WPS account. Of particular interest is the weighting of Cells A and B and Cells C and D in the prevent condition because the predictions of the two accounts differ. Supporting the WPS account, a greater percentage of participants in the prevent condition showed the $\omega_B > \omega_A$ inequality predicted by the WPS account than the $\omega_A > \omega_B$ inequality predicted by the PSB account, $\chi^2(1, N = 40) = 10.00$, $p < .003$. And, a greater percentage of participants in the prevent condition showed the $\omega_D > \omega_C$ inequality predicted by the WPS account than the $\omega_C > \omega_D$ inequality predicted by the PSB account, $\chi^2(1, N = 40) = 6.40$, $p < .02$. It is also evident that a majority of participants showed the $\omega_B > \omega_C$ inequality in both the cause and prevent conditions as the WPS predicts. This pattern of inequality is inconsistent with unweighted variants of the positive-test strategy, such as the dual-factor heuristic and positive-test index, which assigns equal weight to Cells B and C regardless of the direction of the hypothesized causal relationship. In summary, both the nomothetic and idiographic analyses we conducted revealed precisely the type of hypothesis dependence in cell weighting that the WPS account predicts.

Table 5. *Percentage of participants' cell weight relations as a function of hypothesis*

	Hypothesis	
Cell weight relation	*Cause*	*Prevent*
AB contrast		
$<M> \mid \omega_A \mid > \mid \omega_B \mid$	90.0[a,b,c]	25.0[b]
$<M> \mid \omega_A \mid = \mid \omega_B \mid$	0.0	0.0
$<M> \mid \omega_A \mid < \mid \omega_B \mid$	10.0	75.0[a,c]
BC contrast		
$<M> \mid \omega_B \mid > \mid \omega_C \mid$	65.0[a,b]	92.5[a,b]
$<M> \mid \omega_B \mid = \mid \omega_C \mid$	2.5[c]	0.0[c]
$<M> \mid \omega_B \mid < \mid \omega_C \mid$	32.5	7.5
CD contrast		
$<M> \mid \omega_C \mid > \mid \omega_D \mid$	67.5[a,b,c]	30.0[b]
$<M> \mid \omega_C \mid = \mid \omega_D \mid$	2.5	0.0
$<M> \mid \omega_C \mid < \mid \omega_D \mid$	30.0	70.0[a,c]

Note: Superscripts indicate the cell weight relation predicted by the WPS account ([a]), the PSB account ([b]), and unweighted positive-test models ([c]).

Model discrimination

Our final set of analyses tested whether two other models of causal induction that exhibit hypothesis dependence—namely, a modified version of Cheng's (1997) power PC model and Griffiths and Tenenbaum's (2005) causal support model—could explain the observed hypothesis-dependent pattern of inequality in cell weighting. As noted earlier, the power PC model, strictly speaking, is not hypothesis dependent because the calculation of causal power depends on the observed value of ΔP (which, of course, is based on evidence) rather than the hypothesized direction of the relevant causal relationship. However, it seems not only plausible but also more likely that the strategy a causal reasoner would use to compute causal power would depend on his or her belief or hypothesis about the direction of the putative causal relationship than on its actual direction. With this modification to the theory, the calculation of causal power becomes hypothesis dependent and, accordingly,

can be pitted against WPS as a predictive account of cell weighting. Causal power, *PW*, was thus calculated as follows: for hypothesized generative relationships, $PW = \Delta P/[1 - P(y|\neg x)]$; for hypothesized preventive relationships, $PW = \Delta P/P(y|\neg x)$. The second model pitted against WPS as an account of cell weighting was Griffiths and Tenenbaum's (2005) support model, which, as noted earlier, is hypothesis dependent but opaque regarding predictions about cell weighting. We calculated causal support using the Matlab program for Monte Carlo simulation that they provide. The resulting support values then were transformed using the power transformation and optimization-for-fit technique prescribed by Griffiths and Tenenbaum (2005, Footnote 2) and used by others (e.g., Perales & Shanks, 2007).[6] Finally, the WPS account was instantiated as the following conditional probability model, in accordance with (9):

$$0.67P(\mathrm{T}^+|\mathrm{H}^+) + 0.33P(\mathrm{H}^+|\mathrm{T}^+) \qquad (12)$$

where T^+ stands for instances in which the target effect is present, and H^+ stands for instances in which the hypothesized cause is present.

To pit the power, support, and WPS models against each other as predictive models of cell weighting, we correlated the values of the models with each of the four cell frequencies as a function of hypothesis. Figure 3 plots these correlations along with the correlations between mean causal judgements and the cell frequencies. As in the preceding analyses, we reversed the sign of the correlations for the disconfirmatory cells. Figure 3 shows that only the WPS model offered predictions that perfectly matched the rank order of cell weights observed in participants' mean judgements in both the cause and prevent conditions. In contrast, the modified causal power model and the causal support model incorrectly predicted the rank order of the cell weights in both the cause and the prevent conditions. The WPS model also matched the observed data by showing a zero correlation for Cell D in the cause condition (i.e., $\omega_D \approx 0|b^+$) and for Cell C in the prevent condition (i.e., $\omega_C \approx 0|b^-$)—predictions that neither of the two competing models generated. To directly assess the fit of the three models, we correlated the eight cell weights (i.e., four for cause and four for prevent conditions) derived from participants' causal judgements with the models' predicted cell weights. For these analyses, we did not reverse the sign of the correlations for the disconfirmatory cells because that would have restricted the range of the variables. The correlations were .011 for the modified causal power model, .887 ($p < .005$) for the causal support model, and .993 ($p < .001$) for the WPS model.

Discussion

Our research demonstrated three key findings: First, consistent with past studies, Cells A and B were more influential in judgements of probable cause than were Cells C and D. This was evident from our analyses of participants' raw judgements as well as our analyses of their derived cell weights. Second, cell weighting exhibited hypothesis dependence as predicted by the WPS account. That is, on average, the diminishing rank order of the four cells was A, B, C, and D if the focal hypothesis concerned a generative causal relationship, whereas the diminishing rank order was B, A, D, and C if the focal hypothesis concerned a preventive causal relationship. This pattern of findings is fully consistent with what Crocker (1982) found in her study of people's beliefs about cell importance. Third, the majority of participants weighted the four cells in a manner consistent with the WPS account, confirming that the hypothesis dependence observed in our nomothetic analyses was not a consequence of a limited set of extreme cases. The key findings were evident using two commonly used methods

[6] The transformation is of the form $j = sign(support)abs(support)^k$, where k was searched in intervals of .05 to maximize the correlation between support and judgements. For hypothesized generative causation, $k = .95$ ($r = .78$); for hypothesized preventive causation, $k = 1.0$ ($r = .88$).

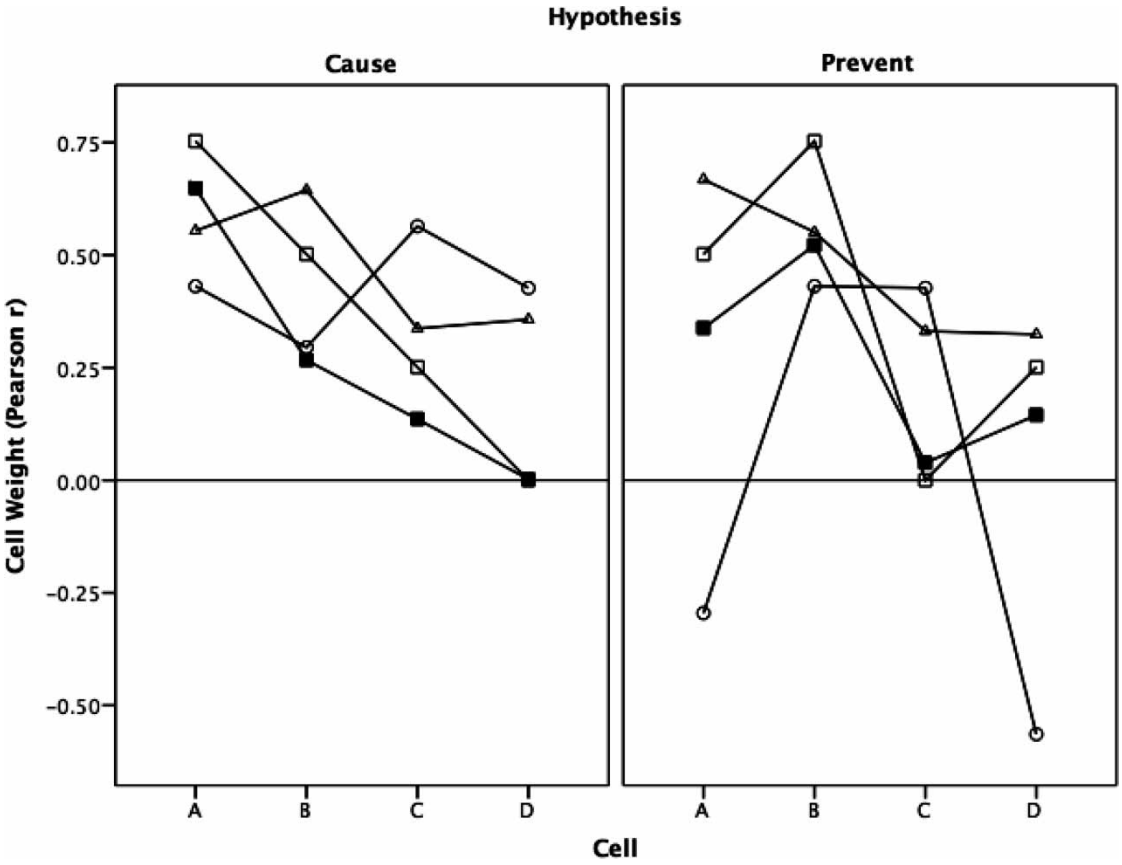

Figure 3. *Observed and predicted cell weights as a function of hypothesis and rule (■ = mean judgement; □ = weighted positive-test strategy, WPS; ○ = causal power; and △ = causal support; the sign of the values for disconfirmatory cells was reversed).*

for presenting contingency information (i.e., the summary and trials formats). Taken together, the findings indicate that the WPS account provides a better explanatory account of cell weighting in judgements of probable cause than the PSB account or its unweighted positive-test counterparts, such as the dual-factor heuristic and the positive-test index (Hattori & Oaksford, 2007). The findings also clearly reveal that the WPS account, when instantiated as a conditional probability model, outperforms other hypothesis-dependent competitors such as the modified power PC model (Cheng, 1997) and the causal support model (Griffiths & Tenenbaum, 2005).

We note, however, that our findings do not imply that the information-processing biases highlighted by the PSB account play no role in causal judgement. Given that people preferentially express statements in the affirmative (Clark & Clark, 1977) and reason more effectively from affirmative premises (Evans & Handley, 1999), positive testing, especially +H tests, will often coincide with a positive-event bias. And, the predictions of cell weighting on the basis of these tendencies do in fact overlap in some respects. For instance, both accounts predict that Cells A and B will be weighted more heavily than Cells C and D, as was found in the present research. Indeed, in the exposition of the PSB account, Mandel and Lehman (1998) employ some of Klayman and Ha's (1987) terminology and refer to Klayman and Ha's positive-test strategy as

another example of a more general "positive bias". Their operational definition of the positive–negative distinction, however, is clearly based on the positive-event notion and, in contrast to the WPS account, leads to a prediction of hypothesis-independent cell weighting.

Although our findings, overall, provide stronger support for the WPS account than for the PSB account, we did observe one finding that does seem to be better accounted for by the PSB account—namely, the main effect of cell on cell weights, which revealed the standard cell weight inequality. As Table 4 shows, Cell A was weighted more strongly in the cause condition than Cell B was weighted in the prevent condition. And, Cell A was weighted more strongly in the prevent condition than Cell B was weighted in the cause condition. This pattern of findings is consistent with the notion of a positive-event bias. Taken together, the findings suggest that while causal reasoners use a weighted positive-test strategy in reaching their judgements, they also show signs of susceptibility to the positive-event bias.

Positive testing, as others have noted (e.g., Friedrich, 1993; Klayman & Ha, 1987), may also reflect a stronger interest in sufficiency than necessity, and some studies suggest that people's concepts of causation tend to place greater weight on a sufficiency criterion rather than a necessity criterion (e.g., Goldvarg & Johnson-Laird, 2001; Mandel, 2003; Mandel & Lehman, 1998; Wolff, 2007). In a related vein, some research indicates that counterfactual conditionals of the "if only *x*, then *y*" variety tend to represent factors seen as having been sufficient to prevent an unwanted outcome rather than those seen as having been necessary to have caused its occurrence (Mandel & Lehman, 1996; for a review, see Mandel, 2005). Indeed, under some assumptive conditions, accounts of information weighting based on notions of positive testing and sufficiency bias lead to identical predictions. Notably, if one assumes that target tests are given no weight in causal reasoning, then positive-test and sufficiency accounts lead to the same prediction. Namely, both accounts predict that Cells A and B would be given more weight than Cells C and D because contrasts of the former constitute positive tests that provide sufficiency information, whereas contrasts of the latter constitute negative tests that provide necessity information. Note that given the strict assumption that target tests receive no weight, neither account can accommodate the fact that Cell B is often weighted more heavily than Cell C in judgements of hypothesized generative causation. In the present account, that assumption was relaxed. Given the importance of temporal order as a cue to causation, we assume that hypothesis tests are given more weight than target tests. However, we also assume that the weight assigned to the latter is greater than zero. Under these assumptive conditions, accounts based on positive testing and sufficiency bias clearly diverge.

For example, if we assume that (a) positive tests receive double the weight of negative tests, (b) sufficiency tests receive double the weight of necessity tests, and (c) hypothesis tests receive double the weight of target tests, then the sufficiency-bias account predicts the following standardized weights:

$$\omega_A|h^+ = .28,$$
$$\omega_B|h^+ = .33,$$
$$\omega_C|h^+ = .17,$$
$$\omega_D|h^+ = .22,$$

and

$$\omega_A|h^- = .33,$$
$$\omega_B|h^- = .28,$$
$$\omega_C|h^- = .22,$$
$$\omega_D|h^- = .17.$$

As is evident, the sufficiency-bias account (like Griffiths & Tenenbaum's, 2005, causal support model) predicts the same pattern of inequality in cell weighting for hypothesized generative relations, h^+, that the WPS account predicts for hypothesized preventive relations, h^-. Similarly, the sufficiency-bias account predicts the same pattern of inequality in cell weighting for

hypothesized preventive relations, b^-, that the WPS account predicts for hypothesized generative relations, b^+. In the present research, the rank order of mean cell weights corresponded perfectly with the predictions of the WPS account and clearly did not support the alternative, sufficiency-bias account discussed here. In short, barring some very restrictive assumptions, the WPS account not only outperforms the PSB account, it also outperforms a sufficiency-bias account that, like the WPS account, assumes a stronger focus on hypothesis testing than on target testing.

The literature also suggests other reasons why people may adopt a positive-test strategy. First, when hypotheses are communicated to hypothesis testers, positive testing may reflect a conversational implicature of relevance (Hilton & Slugoski, 2000). For instance, the explicit form of the hypothesis communicated by the experimenter (e.g., "*x* causes *y*") may be taken to indicate what events require the greatest scrutiny (i.e., *x* cases). Note that a conversational implicature of relevance also offers an explanation for the matching bias (Evans, 1998), which, as discussed earlier, refers to the tendency to select information that matches events that are explicit in a query, proposition, or rule. Indeed, the matching bias might appear to provide a reasonable account of the cell weighting in causal judgement because several studies have found that Cell A receives the greatest weight, and most queries about causation explicitly refer to the *x* and *y* conjunction that defines that cell (e.g., "to what extent does *x* cause *y*?"). The present findings, however, are not well explained by a matching bias for two reasons. First, matching bias does not account for systematic inequalities among Cells B and C. Second, a matching-bias account does not predict hypothesis-dependent weighting of information. Specifically, and contrary to what we observed in the present research, it predicts that Cell A will receive the greatest weight regardless of whether the focal hypothesis concerns generative or preventive causation.

Another explanation is that positive tests may be commonly employed because they are likely to be correlated with the selection of rare events. That is, because people tend to express hypotheses in terms of rare events (McKenzie et al., 2001), positive tests will also tend to hone in on the selection of rare events. Several theorists have proposed that testing hypotheses about rare events is more informative than testing hypotheses about common events (e.g., see Anderson & Sheu, 1995; Hattori & Oaksford, 2007; Klayman & Ha, 1987; McKenzie & Mikkelson, 2000, 2007; Oaksford & Chater, 2003), which suggests that positive testing may represent an adaptive learning strategy. By extension, one might argue that the inequalities in cell weights observed in the present research reflect adaptive biases. To test this intriguing idea, we reexamined cell weighting separately for the AB and CD levels of dominance. Contrary to the "rarity hypothesis", each cell was assigned more weight when it was common (grand $M = 0.38$) than when it was rare (grand $M = 0.08$), $F(1, 75) = 42.40$, $p < .001$, $\eta_p^2 = .36$. Thus, in our research, reliance on positive testing was strongest precisely when it was least informative. Of course, advocates of the rarity hypothesis might argue that our participants may not have perceived the distribution of events in a manner congruent with their actual distribution and that it is the perception of rarity that matters. We agree that it would be instructive to test the hypothesis using subjective estimates of relative frequency in future research. However, we would also argue that there is little reason to suspect that our participants were off the mark. Indeed, our findings in the summary condition paralleled those in the trials condition, and, clearly, it would be difficult in the former condition to err regarding which events were relatively rare versus common. Accordingly, at present, we recommend caution in offering claims about the normative status of observed relations in weighting of contingency information in causal judgement tasks.

Implications for information integration models of causal judgement

The findings of the present research call into question the psychological plausibility of a large proportion of the extant descriptive models of information integration in causal judgement. As we reviewed earlier, a wide variety of competing

information integration models of causal judgement nevertheless share in common the feature that they predict hypothesis independence. That is, all else being equal, they predict that changing a hypothesis from "*x* causes *y*" to "*x* prevents *y*" should not alter the weights assigned to the four cells. The present findings, however, indicate that hypothesis independence is a psychologically implausible attribute. Rather, the findings support the view that people are guided by hypotheses in their information searches and that the weight they implicitly assign to different types of information is reflective of the types of inferential tests they deem most relevant. In support of the WPS account, the pattern of observed cell weight inequalities as a function of hypothesis indicated that those inferential tests give priority to positive tests over negative tests, and especially to $+\mathrm{H}$ tests, which cohere with people's causal schemas in which effects depend on causes.

Because test strategies are defined in relation to an expressed hypothesis rather than in relation to particular event conjunctions, any information integration model that attempts to instantiate this attribute must also define that computational process in relation to a focal hypothesis. The conditional WPS model (12) tested earlier is, of course, a special case of the following generalized weighted test strategy model:

$$J = \omega_{+\mathrm{H}} P(\mathrm{T}^{+}|\mathrm{H}^{+}) + \omega_{+\mathrm{T}} P(\mathrm{H}^{+}|\mathrm{T}^{+}) + \omega_{-\mathrm{H}} P(\mathrm{T}^{-}|\mathrm{H}^{-}) + \omega_{-\mathrm{T}} P(\mathrm{H}^{-}|\mathrm{T}^{-}), \quad (13)$$

in which $\omega_{+\mathrm{H}} = 1$, $\omega_{+\mathrm{T}} = 0.5$, and $\omega_{-\mathrm{H}} = \omega_{-\mathrm{T}} = 0$, in accordance with (9). The generalized model, however, could also accommodate less restrictive assumptions. For example, assume that negative tests have some weight, albeit less than positive tests, and that hypothesis tests are favoured over target tests. Then, the model's weights would show the following pattern of inequality: $\omega_{+\mathrm{H}} > \omega_{+\mathrm{T}} > \omega_{-\mathrm{H}} > \omega_{-\mathrm{T}}$. Although the present research was primarily concerned with developing an account of cell weighting, future research could examine other attributes, such as combination method, to determine how best to instantiate the feature of hypothesis dependence in a psychologically plausible, descriptive model (for a description of a technique for separating weighting and combination assessments, see Mandel & Lehman, 1998).

Finally, we wish to draw the reader's attention to the conceptual link between the present findings and studies that have demonstrated primacy effects of information access on causal or covariational judgement (Dennis & Ahn, 2001; Yates & Curley, 1986; but cf. López, Shanks, Almaraz, & Fernández, 1998; Wasserman, Kao, Van Hamme, Katagiri, & Young, 1996). In one series of experiments, Dennis and Ahn (2001) presented information to participants in a blocked trials format. Although the overall contingency across trials was zero, the first block of trials provided greater support for either a generative or preventive causal relationship, whereas the second block provided greater support for the opposing hypothesis. Participants were given a nondirectional response scale that, in contrast to the present research, did not indicate a directional hypothesis to participants. The findings revealed a primacy effect, such that when the data in the initial block favoured a generative relationship, participants' causal ratings were significantly greater (i.e., more positive) than when the initial block favoured a preventive relationship.

We concur with Dennis and Ahn (2001) that this type of primacy effect is probably attributable to an inferential process whereby early information is used to establish a working directional hypothesis, and subsequent information is interpreted in light of that hypothesis. This interpretation is also supported by Catena, Maldonado, Perales, and Cándido (2008) who found that experimentally manipulated prior beliefs about two co-occurring putative causes influenced the strength of their perceived relationship with a target effect, even though the postbelief contingency data provided the same evidential strength for each putative cause. Although the findings of Dennis and Ahn (2001) and Catena et al. (2008) strongly suggest that hypotheses influence the evaluation of evidence, they do not directly demonstrate the effect of hypotheses per se on judgement while

controlling for evidence. Nor do the previous studies examine how prior beliefs or evidence influence the weighting of certain types of information in judgement, such as the four cells in the 2×2 case. In this regard, the present research builds on this recent work by showing that the manipulation of the direction of a hypothesized causal relationship has a direct effect on the manner in which contingency information is weighted in causal judgement.

Original manuscript received 19 June 2007
Accepted revision received 10 December 2008
First published online 17 April 2009

REFERENCES

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition*, *23*, 510–524.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1119–1140.

Busemeyer, J. R. (1991). Intuitive statistical estimation. In N. H. Anderson (Ed.), *Contributions to information integration theory* (Vol. 1, pp. 187–215). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Catena, A., Maldonado, A., & Cándido, A. (1998). The effect of frequency of judgment and the type of trials on covariation learning. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 481–495.

Catena, A., Maldonado, A., Perales, J. C., & Cándido, A. (2008). Interaction between previous beliefs and cue predictive value in covariation-based induction. *Acta Psychologica*, *128*, 339–349.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*, 365–382.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin*, *8*, 214–220.

Cummins, D. D. (1995). Naive theories of causal deduction. *Memory & Cognition*, *23*, 646–658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274–282.

Dennis, M. J., & Ahn, W.-K. (2001). Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships. *Memory & Cognition*, *29*, 152–164.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 3–19.

Evans, J. St. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking & Reasoning*, *4*, 45–82.

Evans, J. St. B. T., & Handley, S. J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology*, *52A*, 739–769.

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, *100*, 298–319.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565–610.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 334–384.

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, *31*, 765–814.

Hilton, D. J., & Slugoski, B. R. (2000). Judgment and decision making in social context: Discourse processes and rational inference. In T. Connolly, H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (2nd ed., pp. 651–676). Cambridge, UK: Cambridge University Press.

Inhelder, P., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Jenkins, H. M., & Ward, W. C. (1965). Judgments of contingency between responses and outcomes. *Psychological Monographs*, *79*, 1–17.

Kao, S.-F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *19*, 1363–1386.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*, 107–128.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Levin, I. P., Wasserman, E. A., & Kao, S. F. (1993). Multiple methods for examining biased information use in contingency judgments. *Organizational Behavior and Human Decision Processes*, *55*, 228–250.

López, F. J., Shanks, D. R., Almaraz, J., & Fernández, P. (1998). Effects of trial order on contingency judgments: A comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 672–694.

Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General*, *137*, 419–434.

Mandel, D. R. (2005). Counterfactual and causal explanation: From early theoretical views to new frontiers. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 11–27). New York: Routledge.

Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology*, *71*, 450–463.

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.

Mandel, D. R., & Vartanian, O. (2006). Is the weighting of contingency data contingent on the hypothesis assessed? In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (p. 2652). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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 and 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. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*, 33–61.

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.

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

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Perales, J. C., Catena, A., Shanks, D. R., & González, J. A. (2005). Dissociation between judgments and outcome-expectancy measures in covariation learning: A signal detection theory approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1105–1120.

Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin and Review*, *14*, 577–596.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, *110*, 101–120.

Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *2*, 208–224.

Shanks, D. R. (1985). Continuous monitoring of human contingency judgment across trials. *Memory & Cognition*, *13*, 158–167.

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science*, *29*, 5–39.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 49–72). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Vallée-Tourangeau, F., Payton, T., & Murphy, R. A. (2008). The impact of presentation format on causal inferences. *European Journal of Cognitive Psychology*, *20*, 177–194.

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. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *11*, 92–107.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *32*, 129–140.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281.

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, and Cognition*, *16*, 509–521.

Wasserman, E. A., Kao, S.-F., Van Hamme, L. J., Katagiri, M., & Young, M. E. (1996). Causation and association. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *Causal learning: The psychology of learning and motivation* (Vol. 34, pp. 208–264). San Diego, CA: Academic Press.

Wasserman, E. A., & Shaklee, H. (1984). Judging response–outcome relations: The role of response–outcome contingency, outcome probability, and method of information presentation. *Memory & Cognition*, *12*, 270–286.

White, P. A. (2003a). Causal judgment as evaluation of evidence: The use of confirmatory and disconfirmatory information. *Quarterly Journal of Experimental Psychology*, *55A*, 97–114.

White, P. A. (2003b). Making causal judgments from the proportion of confirming instances: The *p*CI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 710–727.

White, P. A. (2004). Causal judgment from contingency information: A systematic test of the *p*CI rule. *Memory & Cognition*, *32*, 353–368.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136*, 82–111.

Yates, J. F., & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement. *Acta Psychologica*, *62*, 293–302.