Accounting for Occurrences: A New View of the Use of Contingency Information in Causal Judgment

Peter A. White
Cardiff University

When people make causal judgments from contingency information, a principal aim is to account for occurrences of the outcome. When 2 causes are under consideration, the capacity of either to account for occurrences is judged from how likely the cause is to be present when the outcome occurs and from the rate at which the outcome occurs when that cause alone is present, which gives an estimate of the strength of the cause. These propositions are formalized in a weighted averaging model, which successfully predicted several judgmental phenomena not predicted by other models of causal judgment. These include a tendency for judgment of one cause (A) to be reduced as the number of occurrences of when only the other one (B) increases and a tendency for A to receive higher judgments than B if A is better able to account for occurrences than B is even if B has a higher contingency with the outcome than A does. Overshadowing, a tendency for judgments of B to be depressed if A has a higher contingency, is weak or absent when B is better able to account for occurrences than A. Results of several experiments support these and related predictions derived from the accounting for occurrences hypothesis.

Keywords: causal judgment, contingency information, overshadowing

Suppose you are a doctor, faced with an array of patients some of whom have and some of whom do not have a particular symptom—let us say blue spots on their skin. You are considering the possibility that the symptom could be caused by a virus called **XZ virus**. You test each of the patients to see whether XZ virus is present in their blood or not. As a result of these tests, you end up with four frequencies: patients who have both XZ virus in their blood and blue spots, patients who have XZ virus but do not have blue spots, patients who have blue spots but do not have XZ virus, and patients who have neither. It is then possible to use this information to make a judgment about whether XZ causes blue spots or not. For example, if 10 patients fell into each of the four categories, it would seem unlikely that XZ virus had anything to do with the blue spots. If all the patients who had blue spots also had XZ virus in their blood, and all those who did not have blue spots did not have XZ virus in their blood, then XZ virus would seem a likely candidate as the cause.

Information of this kind is called **contingency information**. The degree of contingency between a candidate cause and an effect, that is, the degree to which the occurrence of the effect is associated with the presence and absence of the candidate cause, can be measured with the $\Delta P$ rule (Jenkins & Ward, 1965; McKenzie, 1994; Ward & Jenkins, 1965). $\Delta P$ is the probability of the effect occurring when the cause is present minus the probability of the effect occurring when the cause is absent. Values of $\Delta P$ vary from $-1$ (maximum negative contingency) to $+1$ (maximum positive contingency), with a value of zero indicating no contingency. This is an objective measure of contingency (McKenzie, 1994), but it is not a valid measure of causality. For example, a candidate cause may be associated with an effect not because the cause brings about the effect but because both of them are produced by some third factor. Infection with XZ virus and blue spots might be empirically associated because both are caused by exposure to a contaminated water supply.

Suppose now that you are considering two possible causes. For purposes of terminological conventions, call them Causes A and B. Now there are not four kinds of contingency information but eight: occurrences and nonoccurrences of the effect when both causes are present (hereafter AB+, and AB−, respectively), when only Cause A is present (A+ and A−), when only B is present (B+ and B−), and when both are absent (0+ and 0−). There are also four kinds of occurrence rates: AB occurrence rate (the proportion of instances in which the outcome occurs when both A and B are present), A occurrence rate (occurrence rate when A alone is present), B occurrence rate, and absent occurrence rate (occurrence rate when both causes are absent). How do people make causal judgments about two causes, given these kinds of information?

One possibility is that causal judgment is derived from the objective contingency between the cause and the outcome. There are two ways of computing the objective contingency when two causes are being considered. The contingency for one cause can be computed across all the available information, without considering whether the other cause is present or absent. This is the unconditional contingency, or $\Delta P_c$. Or the contingency for one cause can be computed while holding the presence or absence of the other cause constant. Thus, one can compute the contingency for Cause A conditional on Cause B being present, hereafter $\Delta P_A(B)$, or the contingency for A conditional on B being absent, $\Delta P_A(\neg B)$. These are conditional contingencies.
There is evidence supporting the hypothesis that causal judgments are based on conditional contingencies. Spellman (1996a) ran an experiment in which conditional contingencies were manipulated while the unconditional contingency was held constant and found that causal judgments varied in accordance with conditional contingency. In a second experiment, unconditional contingency was manipulated while conditional contingency was held constant, and there were no significant differences in causal judgment between conditions. The hypothesis that causal judgments are made in accordance with conditional contingencies has been used to predict and interpret a number of phenomena (Spellman, 1996a, 1996b; Spellman, Price, & Logan, 2001; Goedert & Spellman, 2005), including overshadowing, a tendency for judgments of one cause to decrease as the association between the other cause and the outcome increases (Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Baker, Vallée-Tourangeau, & Murphy, 2000; Price & Yates, 1993).

Explanatory accounts in which the computation of contingency plays a part, such as the conditional contingency hypothesis and the power PC theory (Cheng, 1997; Novick & Cheng, 2004; Spellman, 1996a), implicitly focus on the cause and ask what the overall pattern in the evidence implies for the status and power of the cause being judged. Information about nonoccurrences of the outcome contributes as much to this as information about occurrences of the outcome does. In the account proposed here, causal judgment focuses more on the outcome and assesses how well occurrences of the outcome can be accounted for by the cause being judged. This account amounts to a claim about psychological processes and their outputs. It is not claimed that it is a normative or objectively correct means of identifying causes; on the contrary, it generates predictions, such as effects of the mere prevalence of causes (White, 2004), that are unlikely to form any part of a normative account of causal inference. It is consistent with a view of humans as fallible but pragmatic causal judges.

In the absence of expectations, nonoccurrences of outcomes do not require causal explanation. If someone eats a meal and does not suffer an allergic reaction, no cause will be sought for the lack of reaction (unless a reaction was expected). Occurrences of outcomes, however, do drive causal judgment processes. If someone does have an allergic reaction after a meal, and there was no expectation either way, then it is likely that a search for a cause will ensue. I propose, therefore, that the main aim of causal judgment is accounting for occurrences and that the judgments made about each cause will depend on how well they are deemed to account for occurrences of the outcome.

Two kinds of information influence judgment of the extent to which a cause accounts for occurrences. One is the extent to which the cause is present when the outcome occurs. I shall call this occurrence judgment. Cause A cannot account for an occurrence of the outcome when Cause A is absent, so its judged capacity to account for occurrences is diminished by such instances. Thus, for Cause A, A+ instances tend to increase A’s judged capacity to produce the outcome, but B+ and 0+ instances tend to reduce it, because A cannot account for an outcome that occurs when it is absent.

The other kind of information is evidence concerning the strength of the cause as a generator of the outcome in question. To assess strength, people will give most weight to the most unambiguous information, which is information about what happens when the cause being judged is present and no other cause is explicitly identified as being present, that is, A+ and A− instances for Cause A and B+ and B− for Cause B. The occurrence rate is an empirical guide to the strength of the cause being judged. The higher the occurrence rate, the stronger the cause is estimated to be.

So far, then, A+ and A− are used for strength estimation, and A+, B+, and 0+ are used for occurrence judgment. However, AB+ and AB− instances would appear to be relevant as well. AB+ is occurrence information and therefore apparently relevant to occurrence judgment. AB+ and AB− are both information about instances where Cause A is present and therefore relevant in principle to strength estimation. The problem is that these kinds of information are relatively ambiguous. For AB+, it is not clear which of the two causes produced the outcome. But the outcome did occur, so there is a need to account for it. In the absence of disambiguating information, the simplest rule would be that each cause is attributed about half of the causal responsibility for producing the outcome. In effect, AB+ counts in the attempt to account for occurrences but with no more than half the weight of A+ information (for Cause A). By the same token, AB− counts in the attempt to estimate the strengths of the causes but with no more than half the weight of A− (for Cause A).

Evidence for this can be found in studies of cue interaction effects such as blocking (Chapman & Robbins, 1990; De Houwer & Beckers, 2002a; Dickinson, Shanks, & Evenden, 1984; Wasserman, Kao, Van Hamme, Katagiri, & Young, 1996) and superlearning (Aitken, Larkin, & Dickinson, 2000). These studies commonly present AB+ instances as a standard against which to assess the effects of a manipulation of other cues. In blocking, for example, C+ and CD+ instances are also presented. Judgments of Cues A and B are then compared with judgments of Cue D. It is usually found that Cues A and B receive equal and moderate causal ratings. For example, in Aitken et al. (2000, Experiment 1), Cues A and B received mean causal ratings of 4.0 and 4.8 (the scale extreme being 12). This can be compared to a mean of 10.3 for a different cue that was only ever presented on its own and with the outcome always occurring. This result supports the claim that AB+ instances result in equal causal responsibility being attributed to each cue and that AB+ instances carry no more than half the weight of A+ (or B+) instances.

The accounting for occurrences hypothesis assigns specific functions to different kinds of contingency information, functions that support an activity of causal inference. However the foregoing discussion of those functions has some implications for the relative weights that the different kinds of information carry in the inferential activity. For this reason, it is possible to represent accounting for occurrences in terms of a weighted averaging model. A brief summary of the implications for cell weights follows.

A+ instances carry the most weight because they are involved in both strength estimation and occurrence judgment and are less ambiguous than AB+ instances. Each kind of occurrence information carries more weight than the corresponding kind of nonoccurrence information (e.g., A+ vs. A−) because nonoccurrence information is not involved in occurrence judgment. AB+ carries less weight than A+ (and B+ for Cause B), and AB− carries less weight than A−, because of the ambiguity due to the fact that both causes are present. B− and 0− are not involved in either strength estimation or in occurrence judgment, so they carry negligible weight. It is also likely that B+ carries more weight than 0+ does.
because it is more relevant to the assessment of Candidates A and B as competing accounts of the outcome. A weighted averaging model that represents these propositions about weights is shown in Equation 1:

\[
J_A = \frac{(.15 n_{AB} + ) - (.075 n_{AB} - ) + (.4 n_A + )}{(.15 n_A - ) + (.15 n_B + ) - (.075 n_0 + )} \\
J_B = \frac{(.15 n_{AB} + ) + (.075 n_{AB} - ) + (.4 n_A + )}{(.15 n_A - ) + (.15 n_B + ) + (.075 n_0 + )}
\]

Equation 1:

\[J(A) \text{ stands for judgment of Cause A and } n \text{ stands for the number of instances of a given kind. For judgment of Cause B, substitute } B+ \text{ for } A+, \text{ B}− \text{ for } A−, \text{ and } A+ \text{ for } B+. \text{ Weights are normalized to sum to 1, and the equation generates values in the range } -1 \text{ (prevent) to } +1 \text{ (cause). } AB+ \text{ and } A+ \text{ have positive valence, and the other four kinds of information in the model have negative valence. } B− \text{ and } 0− \text{ are not represented because under the accounting for occurrences hypothesis, they carry negligible weight. In a model of this kind, each kind of information makes its own contribution to the weighted average. Under this model, causal judgments are not derived from assessments of contingency but from assessments of the extent to which the cause accounts for occurrences, which is inferred from its estimated strength and the occurrence judgment. Any weights that do not violate the foregoing propositions may be used. For example, the weights are not meant to imply that } AB+ \text{ and } B+ \text{ must carry the same weight; there is just nothing in the propositions advanced here to indicate that one kind should carry more weight than the other.}

The aim of the experiments reported here was to test predictions that distinguish the accounting for occurrences hypothesis from the conditional contingency hypothesis. The predictions are briefly described here and are worked out in more detail in the account of each experiment.

1. According to the principle of accounting for occurrences, occurrences of the outcome in A+ instances can be accounted for by A because A is present, but they cannot be accounted for by B because B is absent. Therefore, A+ instances should tend to raise judgment of A and lower judgment of B.

White (2004) found support for this implication in several experiments. In these experiments the prevalence of one cause was manipulated while holding both the conditional contingency for that cause and the prevalence of the other cause constant. Increasing the prevalence of Cause A resulted in higher judgments of Cause A and lower judgments of Cause B, a phenomenon labeled the cause prevalence effect by White (2004). Under the present account, the cause prevalence effect occurs because increasing the prevalence of Cause A entails increasing the number of A+ instances. This increases judgment of A’s capacity to account for occurrences and reduces judgment of B’s capacity to do so, because B cannot account for outcomes that occur when it is absent. In a further series of studies by White (2005), when Cause B had a positive association with the effect, manipulations of the number of A+ instances had a significant effect on judgments of Cause B. Manipulations of the number of A− instances, however, had no significant effect on judgments of Cause B. These findings support the accounting for occurrences hypothesis because A+ instances alter B’s judged capacity to account for occurrences, but A− instances do not; nor are A− instances involved in assessing B’s strength, which is judged only from instances where B is present. This prediction is further tested here in Experiments 1, 2, and 3.

2. Under some circumstances, if the evidence indicates that Cause A accounts for more occurrences of the effect than Cause B does, Cause A should receive higher causal judgments than Cause B even though B may have a higher conditional contingency than A does. Such a difference is never predicted by the conditional contingency hypothesis, because ordinal differences in causal judgment should always conform to ordinal differences in conditional contingency between the candidates. This is tested in Experiments 1, 2, and 3.

3. When there are no nonoccurrences of the outcome in the presence of just one cause, manipulating the number of occurrences of the outcome in the presence of that cause alone does not alter the conditional contingency. Under the conditional contingency hypothesis, therefore, such manipulations should not have a significant effect on causal judgment. Such manipulations do, however, tend to affect judgment of the extent to which the cause can account for occurrences and to alter the weighted average computed by Equation 1. So the accounting for occurrences hypothesis predicts significant effects of such manipulations. This is tested in Experiments 2 and 3.

4. Instances of 0− should have no significant effect on judgment of either cause, because they do not contribute to either accounting for occurrences or assessing causal strength. Instances of 0+ cannot be accounted for by either cause, so some reduction in judgment of both causes as the number of 0+ instances increases is predicted. By contrast, if people make causal judgments by computing the contingency for one cause conditional on the absence of the other cause or by computing ΔP, for each cause, both 0+ and 0− instances should have a significant influence on causal judgment, because they both contribute to both kinds of contingency. If causal judgments are made by computing the contingency for one cause conditional on the presence of the other cause, then neither 0+ nor 0− should have a significant effect on causal judgment. This is tested in Experiment 4.

5. Under the present hypothesis, overshadowing is also an outcome of the attempt to account for occurrences. If this is the case, then overshadowing should be weakened or nonexistent when the number of occurrences of the outcome in the presence of the cause with the higher contingency is lower than the number of occurrences of the outcome in the presence of the cause with the lower contingency. This can be the case when the latter has greater prevalence than the former. This is tested in Experiment 5.

Experiment 1

This experiment tests Predictions 1 and 2. The design involved orthogonal manipulations of the prevalence of each cause. Table 1 shows that the conditional contingency for each cause did not vary across the four conditions and that the conditional contingency for Cause A was always higher than that for Cause B. Values generated by Equation 1 for each cause in each condition are also shown in Table 1. The predicted values are higher for A and lower for B when the prevalence of A is high than when it is low. The predicted values are higher for B and lower for A when the prevalence of B is high than when it is low. This is the case prevalence effect, Prediction 1. In the A low, B high condition, but not in any of the others, the predicted value for B is higher than that for A. The model therefore predicts that judgments of B
should be higher than judgments of A in this condition alone. This is the test of Prediction 2.

Method

Participants. Forty-one first-year undergraduate students of psychology participated in return for course credit.

Stimulus materials. The materials were composed of a questionnaire with an initial written set of instructions followed by four judgmental tasks (shown in Table 1).

The instructions told participants to imagine that they were a horticulturalist studying a garden plant called hosta. Normally this plant had plain green leaves, but sometimes the leaves had an attractive gold edge much prized by gardeners. Their task was to find out what determines whether plants have gold edges on their leaves or not. The suspicion is that chemicals found in the soil might influence this, either causing or preventing gold edges. To test this, they set up a number of pots each containing one hosta, and they control whether the chemicals they are testing are present or absent in the pots. They investigate two chemicals at a time. Once the pots are set up, they wait to see if each plant’s leaves have gold edges or not.

On each page following, the participant would see the results of a series of trials involving a different kind of hosta and two chemicals, different ones on each page. At the top of the page the kind of hosta and the chemicals would be identified by code letters.

Under this would be four columns of information. The first column identified individual plants by number. The second column identified one of the chemicals and told the participant whether that chemical was present or absent in the soil for each plant. The third column identified the other chemical and told the participant whether that chemical was present or absent in the soil for each plant. The fourth column indicated whether each plant developed gold edges on its leaves or not. Thus, by reading across a row, participants would see for each numbered plant whether the first chemical was present or absent, whether the second chemical was present or absent, and whether the plant developed gold edges on its leaves or not.

The instructions then told the participants that, at the bottom of each page, they would see the following two questions:

To what extent does [the first chemical] cause or prevent gold edges on the leaves?

To what extent does [the second chemical] cause or prevent gold edges on the leaves?

To make their judgment, they were to write a number from −100 to +100 beside each question. Writing −100 meant that the chemical in question very strongly prevents gold edges on the leaves. Writing zero meant that the chemical in question has no effect either way. Writing +100 meant that the chemical in question very strongly causes gold edges on the leaves. The more strongly they thought the chemical in question prevented gold edges on the leaves, the lower the number they should put. The more strongly they thought the chemical in question caused gold edges on the leaves, the higher the number they should put.

Each page of the questionnaire presented a separate judgmental task (hereafter task) conforming to this format but with different identifiers for hosta species and chemicals. There were four tasks in all. The first chemical, that is, the one information about which was in the second column, is here designated Cause A, and the second chemical, information about which was in the third column, is Cause B.

Design. Two variables were manipulated within subjects. The prevalence of Cause B (B prevalence) was manipulated with two values, low and high. As shown in Table 1, in the low condition there were 3 B+ instances and 1 B− instance. In the high condition there were 12 B+ and 4 B− instances. The other independent variable was the prevalence of Cause A (A prevalence), which was also manipulated with two values, low and high. As shown in Table 1, in the low condition there were 4 A+ instances and 1 A− instance, and in the high condition there were 12 A+ and 3 A− instances. Table 1 shows that both kinds of conditional contingency are constant across conditions for both causes and that the conditional contingency for Cause A is in all conditions higher than the corresponding kind of conditional contingency for B.

The nature of these manipulations entailed that the total number of instances per judgmental task varied from 17 to 39. Instances were randomly ordered within tasks, and order of tasks was randomized independently for each participant.
Procedure. Participants were tested individually in a small office. The questionnaire for this experiment was included among a set of materials for experiments on unrelated topics. Participants were told that they should ask questions if anything in the instructions was not clear. None had any questions about the materials for this experiment. Participants then proceeded through the tasks at their own pace. At the end of the session participants were given course credit and debriefed about the general aims of the research but not about the specific hypothesis being tested.

Results

Data were analyzed with a 2 (B prevalence, low vs. high) × 2 (A prevalence, low vs. high) within-subjects analysis of variance (ANOVA). Mean causal judgments are shown in Table 1.

Analysis of judgments of Cause A revealed that judgments of A were lower both when the prevalence of A was lower and when the prevalence of B was higher. These are the two components of the cause prevalence effect, so these results support Prediction 1. There was a significant effect of A prevalence, $F(1, 40) = 4.61$, \(MSE = 962.54\), \(p < .05\), with a higher mean at high prevalence (49.84) than at low prevalence (39.44). There was a significant effect of B prevalence, $F(1, 40) = 14.09$, \(MSE = 1,032.80\), \(p < .001\), with a higher mean at low prevalence (54.06) than at high prevalence (35.22). The interaction was not statistically significant.

Analysis of judgments of Cause B also yielded evidence for both components of the cause prevalence effect, supporting Prediction 1. Judgments of B were lower when prevalence of B was lower and when prevalence of A was higher. There was a significant effect of A prevalence, $F(1, 40) = 16.65$, \(MSE = 770.77\), \(p < .001\), with a higher mean at low prevalence (42.82) than at high prevalence (25.23). There was a significant effect of B prevalence, $F(1, 40) = 18.09$, \(MSE = 658.75\), \(p < .001\), with a higher mean at high prevalence (42.55) than at low prevalence (25.50). The interaction was not significant.

The analyses that test Prediction 2 involve comparisons between judgments of causes A and B within each of the four tasks. It was predicted that judgments of Cause B would be higher than judgments of Cause A in the A low, B high task. The means in Table 2 show a difference in the predicted direction, and this was found to be significant, $F(1, 40) = 8.11$, \(MSE = 1,415.56\), \(p < .01\). In the other three tasks, as predicted, judgments of Cause A appeared to be higher than those of Cause B. The difference was significant in the two tasks where the prevalence of B was low: when both causes had low prevalence, $F(1, 40) = 17.65$, \(MSE = 331.89\), \(p < .001\), and when B had low prevalence and A had high prevalence, $F(1, 40) = 28.11$, \(MSE = 1,179.59\), \(p < .001\). When both causes had high prevalence the trend was not significant, $F(1, 40) = 3.54$, \(MSE = 468.90\), \(p = .07\).

Significant cause prevalence effects were therefore found in judgments of both causes. This is predicted by the accounting for occurrences hypothesis, modeled in Equation 1, but not by the conditional contingency hypothesis. The mean judgment of Cause B was significantly higher than that of Cause A in one of the four conditions. This result was predicted by Equation 1. It is, however, contrary to the conditional contingency hypothesis because the conditional contingency for A was always higher than that for B. The correlation between values predicted by Equation 1 and observed means across the four conditions was +.99 for Cause A and +.99 for Cause B.

Discussion

One potential problem with the results of Experiment 1 is that judgments of Cause B were only higher than those of Cause A when B had high prevalence and A had low prevalence. Because the manipulation of prevalence was within subjects, some participants might notice the difference in prevalence between tasks. They might then give higher judgments to Cause B not because of the comparison with the number of A+ instances but because of the comparison with the number of B+ instances in other tasks. Thus, B might receive higher judgments when there were 12 B+ instances because of a comparison with other tasks in which there were 4 B+ instances, and this comparison might be enough to push judgment of B above judgment of A, when the prevalence of A is low. To address this problem, I designed Experiment 2 with a condition in which prevalence was low for both causes, and the conditional contingency was higher for Cause A than for Cause B, but there were still more B+ than A+ instances. If participants judge by comparing B+ instances across conditions, judgments of B should be lower than judgments of A in this task because the prevalence of B is low. But if participants judge by comparing B+ with A+ instances, then judgments of B should be higher than judgments of A, as the accounting for occurrences hypothesis predicts.

The aim of Experiment 2, therefore, was to run a further test of Prediction 2, but with different contingency information. The experiment was also designed to test Prediction 3, manipulating the number of A+ instances while holding the contingency for Cause A constant, as well as testing Prediction 1.

Experiment 2

Method

The participants were 40 first-year undergraduate students of psychology participating in return for course credit. None had participated in Experiment 1. The materials were similar to those of Experiment 1 except for differences in the design. The number of A+ instances was manipulated with two values, 2 and 12. The prevalence of Cause B was manipulated as in Experiment 1, with two values, low and high. These manipulations are shown in Table 2. As Table 2 shows, Equation 1 predicts that judgments of Cause B should be higher than judgments of Cause A in the two conditions where the prevalence of A is low. This is the test of Prediction 2. Equation 1 also predicts higher judgments of Cause A when there are 12 A+ instances than when there are 12, despite the fact that the conditional contingency for A does not change under this manipulation. This is the test of Prediction 3. The prevalence manipulation constitutes another test of Prediction 1. As Table 2 shows, the model predicts that judgments of B should be higher and judgments of A lower when the prevalence of B is high than when it is low. The manipulation of A+ instances also provides a test of Prediction 1: As Table 2 shows, judgments of B should be lower when A+ = 12 than when A+ = 2. All other details of method were similar to Experiment 1.
Results

Analysis of judgments of Cause A revealed support for Prediction 1 and also a cause prevalence effect, supporting Prediction 1. Judgments of Cause A varied depending on both the number of A+ instances and the prevalence of Cause B, despite the fact that the conditional contingency for A did not change. Specifically, there was a significant effect of A+, \( F(1, 40) = 25.37, MSE = 1,610.69, p < .001 \), with a higher mean at 12 occurrences (60.85) than at 2 occurrences (29.28). This result supports Prediction 3. There was a significant effect of B prevalence, \( F(1, 40) = 10.64, MSE = 796.93, p < .01 \), with a higher mean at low prevalence (52.26) than at high prevalence (37.88). This result supports Prediction 1. The interaction was not significant.

Analysis of judgments of Cause B revealed just one significant effect, a main effect of A+. As this is a prevalence manipulation, this result also replicates the cause prevalence effect, in support of Prediction 1, \( F(1, 40) = 19.11, MSE = 1,087.52, p < .001 \), with a higher mean at 2 occurrences (67.46) than at 12 occurrences (44.95). As Table 2 shows, an effect of the B prevalence manipulation was predicted. There was a trend in the predicted direction, but it fell short of statistical significance, \( F(1, 40) = 2.98, MSE = 482.66, p = .09 \). This could perhaps be regarded as marginal support for Prediction 1.

The analyses that test Prediction 2 involve comparisons between judgments of Causes A and B within each of the four tasks. It was predicted that judgments of Cause B would be higher than judgments of Cause A when the prevalence of A was low. The means in Table 2 show substantial differences in line with this prediction, and a one-way ANOVA confirmed that these differences were significant: when both A and B had low prevalence, \( F(1, 40) = 24.73, MSE = 848.97, p < .001 \), and when A had low prevalence and B had high prevalence, \( F(1, 40) = 18.12, MSE = 2,226.94, p < .001 \). When the prevalence of A is high, Equation 1 predicts higher judgments for A than for B. This was found when the prevalence of B was low, \( F(1, 40) = 27.19, MSE = 679.56, p < .001 \), but not when the prevalence of B was high (\( F < 1 \)).

Discussion

The results of this experiment have confirmed and extended those of Experiment 1. In two tasks there were more B+ than A+ instances. In both tasks higher causal judgments were given to Cause B despite the fact that conditional contingencies were higher for A than for B. In one of these the prevalence of Cause B was low. This therefore rules out the hypothesis that participants were giving higher judgments to Cause B by comparing B+ instances across the four conditions as an explanation for the results of Experiment 1. Further evidence of the cause prevalence effect was also found, together with an effect on judgments of Cause A of manipulating the number of A+ instances. These phenomena are also predicted by the accounting for occurrences hypothesis but not by the conditional contingency hypothesis. The correlation between values predicted by Equation 1 and observed means across the four conditions was +.98 for Cause A and +.99 for Cause B.

Experiment 3

Although Experiments 1 and 2 found support for Prediction 2, the conditional contingency for the more prevalent cause was only marginally lower than that for the less prevalent cause. In Experiment 1 the respective conditional contingencies differed by .05, and in Experiment 2 the difference was .14. It is possible that the prediction holds only when the disparity in conditional contingencies is small. The aim of Experiment 3 was therefore to test Prediction 2 in a situation where the difference in conditional contingencies was greater than in Experiments 1 and 2. Table 3 shows that the respective conditional contingencies were in all six conditions higher for Cause A than for Cause B. However Equation 1 generates higher values for Cause B than for Cause A in all conditions.

Method

The participants were 37 first-year undergraduate students of psychology participating in return for course credit. None had participated in either of the previous experiments. The materials were similar to those of Experiment 1 except for the design. In this experiment both kinds of conditional contingency for Cause A were held constant across conditions, but the number of A+ instances was manipulated.
instances was manipulated, being either two or six. This provides a further test of Prediction 3. As Table 3 shows, Equation 1 predicts higher judgments of Cause A when $A^+ = 6$ than when $A^+ = 2$. The number of $B^-$ instances was manipulated with three values: two, four, and six. With this manipulation, the difference between the conditional contingency for Cause A and that for Cause B varied from .14 to .33. Table 3 shows that Equation 1 predicts higher judgments for B than for A in all conditions. Cell frequencies in all tasks are shown in Table 3. Simple effects analysis revealed that the effect of B was significant at $B = 2$ (57.03) was significantly higher than those for $B = 4$ (46.36) and $B = 6$ (38.77), which did not differ significantly. An effect of $B^-$ was not included in the list of predictions in the introduction. However it is consistent with the model in Equation 1, in showing that judgments of B are influenced by $B^-$ information.

These two main effects were qualified by a significant interaction, $F(2, 72) = 3.16, MSE = 673.81, p < .05$. The means are shown in Table 3. Simple effects analysis revealed that the effect of $A^+$ was statistically significant at $B^- = 4$ and at $B^- = 6$ but not at $B^- = 2$ ($F < 1$). The model in Equation 1 does not predict variations in effects of $A^+$ conditional on values of $B^-$, so this result, if reliable, might be worthy of further investigation.

The analyses that test Prediction 2 are comparisons between ratings of the two causes within each task, carried out with a one-way ANOVA. When there were two $A^+$ instances, ratings of B were significantly higher than ratings of A in all three tasks. When there were six $A^+$ instances, ratings of B were significantly higher than ratings of A only at the highest conditional contingency for B. However, there was no task in which A was rated significantly higher than B, even though in two tasks the conditional contingencies differed by .33. Details of the analyses now follow.

### Table 3

**Cell Frequencies, Conditional Contingencies, Predictions of the Accounting for Occurrences Model, and Mean Causal Judgments in Experiment 3**

<table>
<thead>
<tr>
<th>Condition</th>
<th>A2, B2</th>
<th>A2, B4</th>
<th>A2, B6</th>
<th>A6, B2</th>
<th>A6, B4</th>
<th>A6, B6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta P_A$ (-B)</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
</tr>
<tr>
<td>$\Delta P_A$ (B)</td>
<td>+.14</td>
<td>+.25</td>
<td>+.33</td>
<td>+.14</td>
<td>+.25</td>
<td>+.33</td>
</tr>
<tr>
<td>$\Delta P_B$ (-A)</td>
<td>+.86</td>
<td>+.75</td>
<td>+.67</td>
<td>+.86</td>
<td>+.75</td>
<td>+.67</td>
</tr>
<tr>
<td>$\Delta P_B$ (A)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Causal power**

<table>
<thead>
<tr>
<th>Variable</th>
<th>A2, B2</th>
<th>A2, B4</th>
<th>A2, B6</th>
<th>A6, B2</th>
<th>A6, B4</th>
<th>A6, B6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(A)$</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
</tr>
<tr>
<td>$p(B)$</td>
<td>+.86</td>
<td>+.75</td>
<td>+.67</td>
<td>+.86</td>
<td>+.75</td>
<td>+.67</td>
</tr>
</tbody>
</table>

**Weighted averaging model prediction**

<table>
<thead>
<tr>
<th>Condition</th>
<th>A2, B2</th>
<th>A2, B4</th>
<th>A2, B6</th>
<th>A6, B2</th>
<th>A6, B4</th>
<th>A6, B6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta P_A$ (-B)</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+1.0</td>
</tr>
<tr>
<td>$\Delta P_A$ (B)</td>
<td>+.14</td>
<td>+.25</td>
<td>+.33</td>
<td>+.14</td>
<td>+.25</td>
<td>+.33</td>
</tr>
<tr>
<td>$\Delta P_B$ (-A)</td>
<td>+.86</td>
<td>+.75</td>
<td>+.67</td>
<td>+.86</td>
<td>+.75</td>
<td>+.67</td>
</tr>
<tr>
<td>$\Delta P_B$ (A)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Mean causal judgment**

<table>
<thead>
<tr>
<th>Variable</th>
<th>A2, B2</th>
<th>A2, B4</th>
<th>A2, B6</th>
<th>A6, B2</th>
<th>A6, B4</th>
<th>A6, B6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>8.68</td>
<td>7.00</td>
<td>15.54</td>
<td>29.59</td>
<td>25.57</td>
<td>35.81</td>
</tr>
<tr>
<td>$B$</td>
<td>58.51</td>
<td>58.57</td>
<td>45.27</td>
<td>55.54</td>
<td>34.16</td>
<td>32.27</td>
</tr>
</tbody>
</table>

**Note.** A2 and A6 mean 2 and 6 occurrences of the outcome when A alone was present, respectively. B2, B4, and B6 mean 2, 4, and 6 nonoccurrences of the outcome when B alone was present, respectively. $p(A)$ and $p(B)$ are values of causal power for A and B, computed with the equation for generative power in the power PC theory. Means within columns not sharing the same subscript differ by $p < .05$ (ANOVA).
At A\(+\) = 2 and B\(\neg\) = 2, ratings of Cause B were significantly higher than ratings of Cause A, F(1, 36) = 24.57, MSE = 1,869.93, p < .001. At A\(+\) = 2 and B\(\neg\) = 4, ratings of Cause B were significantly higher than ratings of Cause A, F(1, 36) = 23.99, MSE = 2,050.68, p < .001. At A\(+\) = 2 and B\(\neg\) = 6, ratings of Cause B were significantly higher than ratings of Cause A, F(1, 36) = 6.55, MSE = 2,496.49, p < .05. At A\(+\) = 6 and B\(\neg\) = 2, ratings of Cause B were significantly higher than ratings of Cause A, F(1, 36) = 9.33, MSE = 1,334.61, p < .001. At A\(+\) = 6 and B\(\neg\) = 4, there was no significant difference, F(1, 36) = 0.59, MSE = 2,318.48. At A\(+\) = 6 and B\(\neg\) = 6, there was no significant difference, F(1, 36) = 0.13, MSE = 1,839.68.

Discussion

The results of this experiment have added to the support for Prediction 2 from the previous experiments. Even when \(\Delta P_A(-B) = +1\) and \(\Delta P_b(-A) = +.67\), causal ratings were higher for B than for A when there were two A+ instances. In this experiment there was no task in which ratings of A were significantly higher than those of B, even though the conditional contingency for A was always +1 and that for B was always less than +1. The correlation between values predicted by Equation 1 and observed means across the four conditions was + .93 for Cause A and + .89 for Cause B.

It could be argued that the lower prevalence of Cause A means that there is less opportunity to learn about the relation between A and the outcome. Ratings of A might therefore be relatively low because the causal judgment of A is at an early stage of acquisition and should not be regarded as asymptotic. This argument could hold when there were just two A+ instances. However even when there were six A+ instances, B was still rated significantly higher than A when there were two B− instances. If people assess contingency for A only from instances where B is absent, in this task there were 10 relevant instances: six A+ and four 0− instances. This should be enough for a judgment that is close to asymptote. Studies of the acquisition of causal judgment have tended to show that much of the learning about a cause takes place in the first few trials and that the rate of acquisition of causal judgment tends to taper off after that (Shanks, 1985, 1987; Shanks, Lopez, Darby, & Dickinson, 1996; Wasserman, et al., 1996; White, 2000). Previous experiments that have compared different total numbers of instances for otherwise identical contingencies have not found significant effects (Anderson & Sheu, 1995; Lober & Shanks, 2000; Mandel & Lehman, 1998; White, 2003b). The only exceptions involved a small number of trials on one side of the comparison, eight in Mercier and Parr (1996) and six in Anderson and Sheu (1995). These findings indicate that it is unlikely that the differences found in the present experiments could be attributed to insufficient opportunity to learn about the contingency for Cause A, at least when there were six A+ instances.

If a participant judged in accordance with either \(\Delta P_A(-B)\) or \(\Delta P_b(-A)\), their judgments of Cause A would be at or close to the scale extreme of +100 in all six tasks. This was found for 4 participants, 2 of whom each gave six judgments of 100, and the other 2 of whom each gave five judgments of 100 and one of 90. Only three other ratings of 100 were recorded in the entire sample. These 4 participants constitute 11% of the sample. This is similar to the figure of 10% judging in accordance with \(\Delta P\) reported by Kao and Wasserman (1993). It is not possible to say that the present 4 participants were judging in accordance with either conditional or unconditional contingency because judging just from the occurrence rate when A alone was present would yield the same result. It is noteworthy, however, that removing these 4 participants from the data results in mean causal judgments of Cause A that were less than zero in two tasks. When A\(+\) = 2 and B\(\neg\) = 2, the mean judgment was −2.39, and when A\(+\) = 2 and B\(\neg\) = 4, the mean judgment was −4.27. This refutes any claim that judgments of A are in an early stage of acquisition in those tasks, because two A+ instances and no A− instances would never result in a negative judgment (see, e.g., Shanks, 1987). All forms of contingency, conditional and unconditional, yield positive values for Cause A in these cases. The only plausible explanation is that judgments of Cause A were depressed by the comparison with the number of B+ instances.

Experiment 4

Experiment 4 was designed to test Prediction 4: Although there could be an effect of 0+ instances, there should be no significant effect of 0− instances on judgments of either cause. Table 4 shows that the design included similar manipulations of 0+ frequency and 0− frequency.

Method

Thirty-eight first-year undergraduate students of psychology participated in return for course credit. None had taken part in any of the previous experiments. The materials were similar to those of Experiment 1 except for the design. The number of 0+ instances was manipulated with two values: two and eight. The number of 0− instances was manipulated with two values: two and eight. The number of AB− instances was manipulated with two values: zero and four. Cell frequencies in all conditions are shown in Table 4. All other details of method were similar to Experiment 1.

Results

The main findings gave partial support for Prediction 4: There was a significant effect of 0+ in judgments of Cause A but not in judgments of Cause B, and there was no significant effect of 0− in judgments of either cause.

Judgments for both causes were analyzed separately with a 2 within (0+, 2 vs. 8) × 2 within (0−, 2 vs. 8) × 2 within (AB+, 0 vs. 4) ANOVA. Means are reported in Table 4.

Analysis of judgments of Cause A revealed a significant main effect of 0+, F(1, 37) = 9.57, MSE = 1,422.68, p < .01, with a higher mean at the value of two (36.31) than at the value of eight (22.93). There was a significant effect of AB−, F(1, 37) = 14.01, MSE = 1,275.86, p < .001, with a higher mean at the value of zero (37.29) than at the value of four (21.95). This result does not relate to any of the predictions for this experiment, though it is consistent with Equation 1 in showing that AB− does carry some weight in causal judgment. There were no other significant results. In particular, there was no significant effect of 0− (F < 1).

Analysis of judgments of Cause B yielded just one significant result, a main effect of AB−, F(1, 37) = 10.26, MSE = 825.86, p < .01, with a higher mean at the value of zero (16.09) than at the
value of four (5.53). This is also consistent with Equation 1 in showing that AB− instances carry some weight in causal judgment. There was no significant effect of 0− (F < 1).

The correlation between values predicted by Equation 1 and observed means across the four conditions was + .90 for Cause A and + .72 for Cause B. The latter value is lower than those obtained in the previous experiments. This may be explained by the fact that the range of predicted values for Cause B is both low and small, ranging only from .02 to .24. This would be likely to magnify effects of error variance that would tend to reduce the correlation.

Discussion

As predicted, there was no significant effect of the 0− manipulation. The effect of the 0+ manipulation was significant for Cause A but not for Cause B. The latter result could itself be a consequence of the need to account for occurrences. The predictions for B shown in Table 4 tend to be close to zero, especially when there are eight 0+ instances. But giving a judgment of zero to Cause B is problematic because it implies that B is not a cause. Despite this, B is the only explicit candidate that can account for the four B+ instances in each condition. The need to account for those occurrences means that there will be resistance to judgments of the cause being pushed below zero. In effect, the closer to zero judgment of the cause gets, the less weight the disconfirmatory information carries. This tendency could have occurred in Experiment 1 as well. In the A high, B low condition the predicted value for B was − 0.04, but the mean causal judgment was 16.15 (see Table 1). In this case, assigning a negative judgment to B leaves three B+ instances unaccounted for. Making a judgment of B that is greater than zero leaves it with some capacity to account for these occurrences. Thus, disconfirmatory information such as 0+ may be partially discounted when there is a need for positive judgment to account for occurrences of the outcome when only the cause being judged is present.

These results count against the possibility that causal judgments might be based on conditional contingencies. If causal judgments were based on the contingency conditional on the absence of the other cause, then both 0+ and 0− should have had an influence on causal judgment. The finding (in judgments of Cause A) that 0+ had a significant effect but 0− did not disconfirms this expectation. Instances of 0+ contribute to a conditional probability that can be computed only by taking 0− instances into account as well. If there is no significant effect of 0−, then that conditional probability is not being computed, and the influence of 0+ on causal judgment is of some other kind. If causal judgments were based on the contingency conditional on the presence of the other cause, then there should have been no significant effect of either 0+ or 0−. Obtaining a significant result for one manipulation and not for the other is difficult to reconcile with the conditional contingency hypothesis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>2.2,0</th>
<th>2.2,4</th>
<th>2.8,0</th>
<th>2.8,4</th>
<th>8.2,0</th>
<th>8.2,4</th>
<th>8.8,0</th>
<th>8.8,4</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB+</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>AB−</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>A+</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>A−</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B+</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>B−</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0+</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>0−</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

No. of instances

<table>
<thead>
<tr>
<th>Conditional contingency</th>
<th>ΔP, A(− B)</th>
<th>+ .5</th>
<th>+ .5</th>
<th>+ .8</th>
<th>+ .8</th>
<th>+ .2</th>
<th>+ .2</th>
<th>+ .5</th>
<th>+ .5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔP, A(B)</td>
<td>+ .5</td>
<td>0.0</td>
<td>+ .5</td>
<td>0.0</td>
<td>+ .5</td>
<td>0.0</td>
<td>+ .5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>ΔP, B(− A)</td>
<td>0.0</td>
<td>0.0</td>
<td>+ .3</td>
<td>+ .3</td>
<td>– .3</td>
<td>– .3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>ΔP, B(A)</td>
<td>0.0</td>
<td>– .5</td>
<td>0.0</td>
<td>– .5</td>
<td>0.0</td>
<td>– .5</td>
<td>0.0</td>
<td>– .5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Causal power

<table>
<thead>
<tr>
<th>p(A)</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
<th>+ 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(B)</td>
<td>0.0</td>
<td>0.0</td>
<td>+ .37</td>
<td>+ .37</td>
<td>+ .37</td>
<td>+ .37</td>
<td>+ .37</td>
<td>+ .37</td>
<td>+ .37</td>
<td>+ .37</td>
</tr>
</tbody>
</table>

Weighted averaging model prediction

<table>
<thead>
<tr>
<th>Mean causal judgment</th>
<th>A</th>
<th>47.74</th>
<th>30.70</th>
<th>37.66</th>
<th>29.16</th>
<th>33.55</th>
<th>10.87</th>
<th>30.21</th>
<th>17.08</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>13.57</td>
<td>8.97</td>
<td>24.38</td>
<td>5.08</td>
<td>9.31</td>
<td>2.63</td>
<td>17.11</td>
<td>5.45</td>
</tr>
</tbody>
</table>

Note. Conditions are identified respectively as 0+ (2 or 8), 0− (2 or 8), AB− (0 or 4). p(A) and p(B) are values of causal power for A and B, computed with the equation for generative power (for positive values of ΔP) or the equation for preventive power (for negative values of ΔP) in the power PC theory.
Experiment 5

This experiment was designed to test Prediction 5. Overshadowing should be weak or absent when the number of occurrences of the outcome in the presence of the cause with the higher contingency is lower than the number of occurrences of the outcome in the presence of the cause with the lower contingency.

Method

Thirty-nine first-year undergraduate students of psychology participated in return for course credit. None had taken part in any of the previous experiments. The materials were similar to those of Experiment 1 except for the design. $\Delta P_c(A(-B))$ was manipulated by altering frequencies of $A+$ and $A-$ instances. In one condition there were 10 $A+$ and 2 $A-$ instances, and in the other there were 2 $A+$ and 10 $A-$ instances. As shown in Table 5, given that the absolute occurrence rate was zero, this yields conditional contingencies of +.83 and +.17, respectively. $\Delta P_c(B(-A))$ was held constant at +.5, intermediate between the two contingencies for $A$. This is an overshadowing manipulation, and the usual prediction would be that judgments of $B$ should be depressed in the condition where $A$ has the higher contingency (Price & Yates, 1993). The prevalence of $B$ was manipulated, such that in one condition there were 6 $B+$ and 6 $B-$ instances, and in another condition there were 12 of each. The latter number is greater than the number of $A+$ instances in the high contingency condition for $A$ (10). Because of this, $B$ is still better able to account for occurrences of the outcome than $A$ is, even though the conditional contingency for $B$ is lower than that for $A$.

This is shown in the model predictions in Table 5. When the prevalence of $B$ is low, the model predicts an overshadowing effect with values of .43 when $\Delta P_c(A(-B))$ is low and .11 when $\Delta P_c(A(-B))$ is high. When the prevalence of $B$ is high, little or no overshadowing effect is predicted because the values for $B$ are .44 when $\Delta P_c(A(-B))$ is low and .38 when $\Delta P_c(A(-B))$ is high. Cell frequencies in all conditions are shown in Table 5. All other details of method were similar to Experiment 1.

Results

As predicted, there was a significant overshadowing effect when the prevalence of $B$ was low, but there was no significant overshadowing effect when the prevalence of $B$ was high. Unexpectedly, there was no significant effect of the prevalence manipulation on judgments of either, and in fact at the low value of contingency for $A$ there was a reversal of the usual cause prevalence effect in judgments of $B$.

Data for both causes were analyzed with a 2 within (A contingency, low vs. high) × 2 within (B prevalence, low vs. high) ANOVA. Means are reported in Table 5.

This analysis of judgments of $A$ revealed just one significant result, a main effect of the $A$ contingency manipulation, $F(1, 38) = 115.40, MSE = 380.15, p < .001$, with a higher mean at low contingency (55.14) than at high contingency (29.22). This amounts to a manipulation check, showing that participants were sensitive to the manipulation of information about $A$, as predicted by both the conditional contingency hypothesis and by the accounting for occurrences hypothesis (see Table 5).

The analysis of judgments of $B$ yielded a significant effect of $B$ contingency, $F(1, 38) = 6.80, MSE = 624.27, p < .05$, with a higher mean at low contingency (55.14) than at high contingency (44.71). This is in the direction of an overshadowing effect. It was qualified by a significant interaction with $B$ prevalence, $F(1, 38) = 8.12, MSE = 339.60, p < .01$. Simple effects analysis revealed a significant effect of $A$ contingency at low $B$ prevalence, $F(1, 38) = 13.56, MSE = 510.70, p < .01$, but no significant effect of $A$ contingency at high $B$ prevalence, $F(1, 38) = 0.18, MSE = 453.17$. The means involved in these comparisons are in Table 5. These results show, as predicted, an overshadowing effect when the prevalence of $B$ was low but not when the prevalence of $B$ was high.

There was no significant effect of $B$ prevalence on judgments of $A$ at high $A$ contingency, $F(1, 38) = 0.70, MSE = 322.89$. However there was an unexpected reversal of the usual cause prevalence effect at low $A$ contingency, $F(1, 38) = 12.01, MSE = 291.97, p < .01$. The means involved in these comparisons are also those reported in Table 5. The main effect of $B$ prevalence was not significant, $F(1, 38) = 3.54, MSE = 275.25, p = .07$.

Discussion

As predicted, no significant overshadowing effect was found when $B$ had high prevalence. In this condition there were more $B+$
than A+ instances, even when A had high conditional contingency. B is therefore better able to account for occurrences of the outcome than A is, and, according to the accounting for occurrences hypothesis, this weakens or eliminates effects of changes in A’s contingency on judgments of B. The usual overshadowing effect was found, however, when B’s prevalence was low. In this condition, when A’s contingency was high there were more A+ than B+ instances, whereas the opposite was the case when A’s contingency was low. According to the accounting for occurrences hypothesis, this is what determines the occurrence of an overshadowing effect.

Unusually, however, there was no significant main effect of prevalence in this experiment. In fact, when A’s contingency was low, there was a statistically significant reversal of the cause prevalence effect in judgments of B. This is unprecedented in research on the cause prevalence effect. There were 11 tests of the cause prevalence effect in White (2004), 10 of which yielded statistically significant tendencies in support of the effect, with the exception showing a nonsignificant trend in the direction of cause prevalence. Experiments 1–3 in the present article included 8 tests of the cause prevalence effect (Prediction 1) and 7 of them yielded statistically significant support for it, and in the 8th there was a nonsignificant trend in the predicted direction.

There is no obvious difference between the present study and the others that might account for a reversal of the cause prevalence effect. One noteworthy feature of the comparison that yielded the reversal is that there were more B+ (6 and 12) than A+ instances (2) in both conditions. The difference between 6 and 12 might not be meaningful, given that both are substantially greater than 2. In support of this, in Experiment 2 an identical manipulation resulted in no significant difference in judgments between 6 and 12 B+ instances when there were 2 A+ instances. There was a significant difference, however, when there were 12 A+ instances (see Table 2). This would not explain a reversal of the effect, but it might explain an absence of the effect, and the reverse tendency observed could be regarded as a statistical aberration, the reliability of which would need to be confirmed before too much effort was expended on the attempt to interpret it.

General Discussion

It has been proposed that a principal aim of causal judgment is to account for occurrences of the outcome. Capacity to account for occurrences is based on the strength of the cause as estimated by the occurrence rate when the cause is present and by occurrence judgment, which is the extent to which the cause is present when the outcome occurs. Under this hypothesis, people compute neither any form of contingency nor the conditional probabilities from which contingencies are computed. Instead, the inferential activity of accounting for occurrences can be represented by a weighted averaging model.

The model shown in Equation 1 successfully predicted several findings: the cause prevalence effect (Experiments 1, 2, and 3); the tendency for the cause prevalence effect to occur under manipulations of occurrences and not of nonoccurrences (Experiments 2 and 3); a tendency to give higher judgments to a cause that is better able to account for occurrences than another cause, even when the latter has a higher conditional contingency than the former (Experiments 2 and 3); a significant effect of changing the number of occurrences in the presence of one cause even though the conditional contingency does not change (Experiments 2 and 3); and a significant effect of 0+ instances, particularly when other available information supports a high causal judgment, but no significant effect of 0− instances (Experiment 4). It was also predicted that overshadowing would be weak or nonexistent if the cause, the contingency of which was manipulated, was less able to account for occurrences of the outcome even at its higher level of contingency. Experiment 5 found no significant overshadowing when this was the case, but overshadowing did occur when both causes were equally prevalent. This experiment had an additional peculiar result in the form of an unexplained reversal of the usual cause prevalence effect in one condition, but this effect is independent of the overshadowing test and does not provide grounds for doubting the result of the overshadowing test. Overall, then, the results supported the predictions derived from the accounting for occurrences hypothesis as formalized in Equation 1, and they did not support those derived from the hypothesis that causal judgments follow from the computation of conditional contingencies.

The results of the study by Spellman (1996a) supported the conditional contingency hypothesis, but they are also predicted by the accounting for occurrences hypothesis. The design of Experiment 1 from Spellman (1996a) is shown in Table 6. In this experiment there were two causal candidates: blue and red liquids. There were three conditions, designed so that conditional contingencies varied as shown in Table 6, but unconditional contingencies were held constant at zero for blue and +.5 for red. Table 6 shows that mean judgments of blue declined across conditions in accordance with the conditional contingency. Judgment of red was, unexpectedly, significantly higher in Condition 3 than in Condition 1, and Condition 1 was not significantly different from Condition 2, despite the difference in conditional contingencies.

Table 6 also shows the predictions generated by Equation 1. These too predict the observed decline in judgments of blue across

| Table 6 | Frequencies of Different Kinds of Instances in the Three Conditions of Spellman (1996a), Experiment 1 |
|---------|---------------------------------|-----------------|-----------------|-----------------|
| Variable | Condition 1 | Condition 2 | Condition 3 |
| Bloom Not Bloom Not Bloom Not Bloom Not |
| No. of instances | 5 0 7.5 2.5 10 5 | 5 10 2.5 7.5 0 5 | 10 5 2.5 7.5 5 10 |
| Conditional contingency | | | |
| ΔP_blue(−red) | +.33 | 0.0 | −.33 |
| ΔP_blue(red) | +.33 | 0.0 | −.33 |
| ΔP_red(−blue) | +.67 | +.5 | +.67 |
| ΔP_red(blue) | +.67 | +.5 | +.67 |
| Weighted averaging model prediction | | | |
| Blue | −.04 | −.11 | −.20 |
| Red | +.52 | +.57 | +.79 |
| Mean causal judgment | | | |
| Blue | 11.7 | −11.3 | −61.8 |
| Red | 60.1 | 51.9 | 83.5 |
conditions. They also predict that judgment of red should be higher in Condition 3 than in the other two conditions, which was found. Equation 1 also predicts higher judgment in Condition 2 than in Condition 1, which was not found. Both the conditional contingency hypothesis and the accounting for occurrences hypothesis therefore predict the results for judgments of blue. Neither model exactly predicts the results for judgment of red, but the accounting for occurrences hypothesis does predict the higher mean found in Condition 3, which was not predicted by the conditional contingency hypothesis. Overall, then, the accounting for occurrences hypothesis accounts for the results at least as well as the conditional contingency hypothesis does.

Other Theories and Models

There are four ways of making causal judgments that involve the computation of contingencies: unconditional contingencies, contingency for one cause conditional on the absence of the other cause, contingency for one cause conditional on the presence of the other cause, and computation of simple causal power according to the power PC theory (Cheng, 1997). The computational procedure of the power PC theory modifies the contingency for the cause being judged conditional on the absence of the other cause, with information about the absent occurrence rate (0+ and 0− information). Specifically, ΔP for the cause being judged is divided by one minus the absent occurrence rate.

Leaving aside unconditional contingencies for a moment, the other three ways cannot account for the cause prevalence effect because manipulations of cause prevalence do not alter the conditional contingencies. They are unable to account for effects, on judgments of a cause, of changing the number of occurrences in the presence of that cause when the contingency does not change (Experiments 2 and 3). They are unable to explain the finding that the occurrence of an overshadowing effect depends on the relative prevalence of the two causes (Experiment 5). Illustrating the failings of the power PC theory in this respect, in Experiment 1 the absent occurrence rate is zero in all four conditions. This means that the simple causal power for each cause is equal to its contingency conditional on the absence of the other cause. As Table 1 shows, this does not vary across conditions and is therefore unable to explain the significant differences found in that experiment.

In the study by Spellman (1996a), both forms of conditional contingency had the same values in all conditions. In the present experiments, as the tables show, the values were usually different, so each form has to be evaluated separately. The hypothesis that people compute contingency for one cause conditional on the absence of the other cause cannot explain why 0+ instances had a significant effect on causal judgment but 0− instances did not (Experiment 4). Both kinds of occurrence contribute equally to the computation of the conditional contingency, so if only one of them is influential, it must be used in some other way. The hypothesis that people compute contingency for one cause conditional on the presence of the other cause cannot explain the significant effect of 0+ instances (Experiment 4), because those events do not enter into that conditional contingency. It is unable to explain the influence of the manipulation of numbers of occurrences in the presence of one cause on judgments of that cause for the same reason (Experiments 2 and 3). It is also unable to explain the cause prevalence effect, because manipulations of cause prevalence do not alter either form of conditional contingency.

Turning now to unconditional contingencies, cause prevalence manipulations for one cause (A) do alter the unconditional contingency for the other cause (B). To illustrate, in Experiment 1, when the prevalence of Cause A is low, ΔP_{A} = +.52 when the prevalence of B is low and +.29 when the prevalence of B is high. This would predict a higher causal judgment for Cause A in the former task than in the latter, consistent with the results. However White (2004) found that unconditional contingencies for both causes consistently decreased as the prevalence of one cause increased, whereas judgments tended to increase for the cause, the prevalence of which increased, and to decrease for the other cause. Also, at low conditional contingencies (+.25), unconditional contingencies predict the reverse of a cause prevalence effect, and this was not found (White, 2004). In the present research, unconditional contingencies do not predict the overshadowing effect found at low prevalence of Cause B in Experiment 5. Nor do unconditional contingencies predict the effects on judgments of Cause A of altering the number of A+ instances in Experiment 2, because this manipulation did not alter the unconditional contingency for Cause A. On the other hand, unconditional contingencies predict effects of both 0+ and 0−, because these contribute to the computation of unconditional contingency for both causes. However, there was no significant effect of 0− in Experiment 4. Furthermore, the results of the study by Spellman (1996a) count against the hypothesis that causal judgments are based on unconditional contingencies.

When there are two possible causes, an additional possibility is that they may be judged to interact in some way. When participants are not allowed to judge interactive causal influence, as in the present research, they may distribute their judgment of the interaction between their judgments of the two individual causes. In this way, judgmental tendencies could reflect the influence of a covert judgment that the causes interact with each other. There are a priori many kinds of interaction that might take place. A might facilitate the effect of B while B has no effect on A, the combination of the two causes might pass a threshold for production of the outcome that is not passed by either cause on its own, the two causes might inhibit each other, and so on. It is clear, however, that the two causes cannot interact with each other when one of them is not present. This implies that an effect of B+ on judgments of A cannot reflect a judgment that A and B interact with each other: A and B cannot be interacting in the case of B+ because A is absent, and they cannot be interacting in the case of A+ instances because B is absent. The cause prevalence effect therefore cannot reflect an implicit judgment that the two causes are interacting with each other. People could judge that an interaction is going on when both A and B are present, and they could use other kinds of contingency information to assess the likelihood of this (Novick & Cheng, 2004). But such an effect would be independent of the cause prevalence effect and of the other effects investigated here.

Novick and Cheng (2004) proposed that people make interactive causal influence judgments in a normatively appropriate way, extending the principles of the power PC theory (Cheng, 1997). The equation used to compute interactive causal influence depends on two things: whether values of ΔP for A and B are both generative, both preventive, or one of each, and on whether the two-way interaction contrast, a version of ΔP for the conjunction of A and B, is positive or negative. I illustrate with the judgmental
tasks from Experiment 1 (Table 1). ΔP is computed for each cause conditional on the absence of the other cause. This gives values of +.8 for Cause A and +.75 for Cause B in all four conditions. The interactive contrast, computed with Equation 2 from Novick and Cheng (2004) is −.55 in all conditions. This means that the empirical equation for assessing interactive causal influence is Case 2 from Figure 6 of Novick and Cheng (2004). This gives a value of −.05 for all four conditions.

The extension of the power PC theory to interactive causal influence is therefore unable to explain the results because the interactive causal influence is the same, and close to zero, in all four conditions. In general, the empirical procedures for computing both simple and interactive causal powers all employ the various conditional probabilities. If the conditional probabilities do not vary across conditions, as is the case in Experiment 1 (and Experiment 2), all the computational procedures predict that judgments will not differ across conditions. The present results therefore have disconfirmatory import for the power PC theory and indicate that causal judgment does not involve the computation of conditional probabilities.

Recently a number of authors have proposed that causal judgment can be modeled in accordance with principles of Bayesian inference (e.g., Griffiths & Tenenbaum, 2005; McKenzie & Mikkelsen, 2007). Thus far, these accounts have been confined to the case where a single explicit cause is under consideration, and it is not clear how they could be extended to the two-cause case. Griffiths and Tenenbaum (2005) treated the discovery of causal structure as distinguishing between two possible causal graphs, one in which only B (the background or set of all other possible causes) is connected to the effect and one in which both B and C1 (the cause under consideration) are connected. The account also distinguishes between the structure of the system and the strength of a given link.

It might be thought that extending this approach to the case of two possible causes, C1 and C2, involves four graphs rather than two: structures where B only, B and C1 only, B and C2 only, or all three, respectively, are connected to the effect. But in fact there is an extra structural layer on top of this, concerned with the link between the two possible causes. As it is not clear a priori which of these could be the effect, there are four possibilities: no link, C1 affects C2 only, C2 affects C1 only, and both causes affect each other. These possibilities are superimposed on the preceding ones, making a total of 16 different graphs. And this is without considering the nature of the interaction that might take place: C1 might enable C2 to produce the outcome, so that the link between C2 and the outcome is not there unless C1 is part of the system; C1 might inhibit C2; the two causes might interact to produce an effect that neither of them produces alone, and so on. This illustrates a general problem for the Bayesian approach. Extending the model to the two-cause case would involve a combinatorial explosion that is computationally intractable. As McKenzie and Mikkelsen (2007, p. 35) wrote, “Bayesian models are notorious for their enormous complexity when applied to real-world problems, making them unlikely candidates for models of psychological processes.”

The Bayesian inference account by McKenzie and Mikkelsen (2007) seeks to interpret the unequal cell weights found in many previous experiments as adaptive, given assumptions about the rarity of things being present. They showed that the weights can be changed when things being present are explicitly presented as, or believed to be, common. They too considered only the single-cause case. There are three main problems with extending their model to the two-cause case. One is that the assumption about the rarity of things being present is a free parameter (so long as it is less than .5), which gives opportunities to fit data to the model by adjusting the assumed value. The second is that it is not clear how the model would deal with instances where both causes are present. If cell weights are set only by rarity, one would expect these cells to carry the most weight because the conjunction of A and B must be rarer than either A or B occurring alone (if A and B are distributed independently of each other). But this implied difference in cell weights is contradicted by the results of Spellman’s (1996a) experiments. The third problem is that on any assumption about rarity other than zero, the cells equivalent to cell D, namely B− and 0− (for judging Cause A) should carry nonzero weight. The results of Experiment 4 and of the study by White (2005) do not support this, although nonsignificant results carry limited disconfirmatory weight. The Bayesian approach to causal judgment holds great promise, but it is not yet clear what form a Bayesian account of judgment with two explicit causal candidates would take.

The fact that instances occur individually in the instance list procedure means that judgment could be acquired by some form of associative learning process operating during visual scanning of the instances in the list. Any associative learning model would readily explain one finding from the present research: That judgments of a cause were higher when there were 12 instances than when there were 2, in the absence of any change in conditional contingency (Experiment 2), would be predicted on the grounds that 12 instances provide more opportunity for learning the causal relation than 2 do, making the association between cause and outcome stronger.

The associative learning model that has been most investigated in application to human causal judgment is the Rescorla–Wagner model (Allan, 1993; De Houwer & Beckers, 2002b; Rescorla & Wagner, 1972; Shanks, 1995; Tangen & Allan, 2003). White (2005) showed that the model fails to predict some of the phenomena of the cause prevalence effect. When there are no instances where both causes are present, the model predicts that B+ and B− will have no effect on judgment of A, and vice versa (Tangen & Allan, 2003). This prediction was disconfirmed by the results of two experiments reported by White (2005), where effects of B+ on judgments of A did occur despite the fact that there were no instances where both causes were present together. When there are instances where both causes are present, the model predicts that both B+ and B− will affect judgment of A (Tangen & Allan, 2003). This prediction was also disconfirmed by several findings in White (2005) that B− had no significant effect on judgments of A. Therefore the Rescorla–Wagner model does not account for the evidence relating to the cause prevalence effect.

If any other associative learning model is to account for these results, it has to predict effects on judgments of A of instances where B alone is present (or vice versa), even when there are no instances where both causes are present; not only that, but it should also predict effects of B+ and no effects of B−. It would then be required to explain the other findings reported here, such as the finding that the occurrence of overshadowing depends on the relative prevalence of the two causes.
The cause prevalence effect can be predicted on the hypothesis that participants are making diagnostic judgments, that is, reasoning from effect (e.g., symptoms) to cause (e.g., disease) instead of from cause to effect. A typical task is predicting which of two diseases a patient has from information about symptoms. On any given trial information about two symptoms is presented, one of which is a perfect predictor of one of the diseases and the other is an imperfect predictor (I) because it is associated with both diseases. One of the diseases is common in the stimulus materials and the other rare, typically with a ratio of 3:1. When presented with I on its own, participants tend to judge that the common disease is present. This is a standard base rate effect (Medin & Edelson, 1988). If participants adopted the diagnostic direction of reasoning, then, giving higher ratings to the more prevalent cause could be interpreted as an instance of the standard base rate effect.

It is not likely, however, that judgments made in the present research were diagnostic judgments. The initial instructions contained a possible ambiguity in this respect. Participants were told that their task was to find out what determines whether plants have gold edges on their leaves or not. This could mean that they were to judge the strength of different possible causes or that they were to evaluate the extent to which the evidence favored one possible cause over another, an interpretation that could encourage a diagnostic direction of reasoning. Against that, the wording of the causal judgment question was unambiguously causal and not diagnostic: “To what extent does [the first chemical] cause or prevent gold edges on the leaves?” Participants were faced with this question every time they made a judgment, so it was more salient than the sentence from the early part of the instructions. The instructions on how to make the judgment were also unambiguously causal, as shown in the method section of Experiment 1.

In addition, the cause prevalence effect was initially found with tasks where the measure could not be interpreted as a request for diagnostic judgment. These were medical scenarios involving judging what was causing either headaches or rashes of blue spots (White, 2004, 2005). The initial instructions did not include the ambiguous statement that was present in the instructions used here (see, e.g., White, 2004, p. 966). Moreover it has been shown that cue interaction effects such as overshadowing do not occur with the diagnostic direction of reasoning (Waldmann, 2000; Waldmann and Holyoak, 1992). If participants had adopted the diagnostic direction of reasoning to make their judgment in the present research, the overshadowing effect reported in Experiment 5 would not have occurred. It is therefore likely that the findings reported here are phenomena of causal judgment.

The One-Cause Case

The accounting for occurrences hypothesis can be extended to the case where just one cause is under consideration. As in the two-cause case, A+ instances contribute to both accounting for occurrences and assessing strength. A− instances contribute only to assessing strength and should therefore carry less weight than A+ instances. Instances of 0+ contribute only to accounting for occurrences and should therefore also carry less weight than A+ instances. It is not clear whether A− instances should carry more weight than 0+ instances. Instances of 0− contribute to neither and should therefore carry negligible weight.

As far as these predictions go, they fit with the results of studies of cell weights with one causal candidate. These studies have consistently found the order A+ > A− > 0+ > 0−, though the difference between A− and 0+ has not always been significant (Anderson & Sheu, 1995; Kao & Wasserman, 1993; Levin, Wasserman, & Kao, 1993; Mandel & Lehman, 1998; Wasserman, Dorner, & Kao, 1990; White, 2003a). However, under the present account, 0− instances carry no weight, and the results of the present experiments are consistent with that. But 0− information does carry some weight when a single cause is being judged, though less than the other three kinds. One possible explanation for this is that the format of judgment with a single cause induces some participants to adopt a different approach in which they compare occurrence rates when the cause is present and absent, respectively. This may be more likely to happen when just one cause is being judged, particularly with summary presentation where all the information is available at once in a statistical summary, because the information is simpler (four kinds of information as opposed to eight when there are two causes). Kao and Wasserman (1993) reported that approximately 10% of their participants appeared to be making judgments in accordance with ΔP. These participants gave approximately equal weight to all four cells, so the limited weight given to 0− information could be explained as the average of ΔP judges giving substantial weight to that cell and the rest giving no weight to it.

This is not a complete explanation of the weight given to 0−, however. White (2000) found that different individuals used 0− information in different ways: Some gave it no weight at all, others treated it as confirmatory, others treated it as disconfirmatory, and others treated it as either confirmatory or disconfirmatory depending on whether the previous judgment had been below or above zero, respectively. The overall weight given to 0− emerges from the combination and prevalence of these different tendencies. It therefore appears that the meaning of 0− information varies between individuals in ways that have yet to be elucidated. Further research may reveal whether these meanings contribute to the attempt to account for occurrences or whether other factors additional to accounting for occurrences are involved. Such research might also help to explain why 0− information is influential in the single cause case but apparently not when two causes are being judged.

References


Received February 14, 2007
Revision received October 3, 2007
Accepted October 4, 2007