Research Article

Two Heads Are Better Than One, but How Much?

Evidence That People's Use of Causal Integration Rules Does not Always Conform to Normative Standards

Miguel A. Vadillo,[1] Nerea Ortega-Castro,[2] Itxaso Barberia,[2] and A. G. Baker[3]

[1]University College London, UK, [2]Universidad de Deusto, Bilbao, Spain,
[3]McGill University, Montréal, Canada

Abstract. Many theories of causal learning and causal induction differ in their assumptions about how people combine the causal impact of several causes presented in compound. Some theories propose that when several causes are present, their joint causal impact is equal to the linear sum of the individual impact of each cause. However, some recent theories propose that the causal impact of several causes needs to be combined by means of a noisy-OR integration rule. In other words, the probability of the effect given several causes would be equal to the sum of the probability of the effect given each cause in isolation minus the overlap between those probabilities. In the present series of experiments, participants were given information about the causal impact of several causes and then they were asked what compounds of those causes they would prefer to use if they wanted to produce the effect. The results of these experiments suggest that participants actually use a variety of strategies, including not only the linear and the noisy-OR integration rules, but also averaging the impact of several causes.

Keywords: causal reasoning, integration rules, summation

It is difficult to find any single behavior in our daily life that does not require some sort of causal thinking. From the instant we wake up and turn off the alarm clock to the moment we brush our teeth before going to bed, all our activities are influenced by causal beliefs. However, it is still far from obvious how we acquire and use this vast amount of knowledge. Establishing with some certainty that a given event is the cause of another is a deeply complex cognitive process. Imagine, for example, that after eating a peach your throat suddenly feels irritated. You will probably suspect that the peach has provoked an allergic reaction. But the simple contiguity of eating the peach and the sore throat does not guarantee that there is an actual causal relation between them. There are a myriad of events that also took place just before your reaction and any of them could also be responsible for it (e.g., other foods you have eaten, the smoke from a cigarette, or, who knows, radiation from the sun, to name just a few).

Ideally, if you could discard the influence of all the other causal factors by eating a peach in a context free from all the other potentially relevant events, you would be much more confident that there is a true causal connection between the peach and the allergic reaction. Unfortunately, in most instances, such an ideal context simply does not exist. Moreover, it is often impossible to know what alternative causal factors need be taken into account. Fortunately, there is an indirect way to test the causal role of the peach. Although you often cannot eliminate all the other potential causes of the allergy, you can try to keep them constant while you test the causal impact of eating the peach. If, other things being equal, the probability of an allergic reaction is higher when you eat peaches than when you do not, then peaches are likely to make a causal contribution to your allergy. Most models of causal induction compute the causal strength of a potential candidate cause by subtracting the probability of the effect given the absence of the candidate cause from the probability of the effect given its presence. However, how this subtraction is done depends on assumptions about how the causal impact of multiple causes should be combined.

The simplest assumption about how several causes jointly determine the probability of the effect is that the impact of each cause should be integrated by simple linear addition. If, for example, we know that cause A produces an outcome 30% of the time and B produces the same outcome 20% of the time, then the notion of linear addition would argue that when both are present (and no other effective cause is present) the probability of the effect should be 50%. This is the law for combining mutually exclusive probabilities. Following this reasoning, isolating the role of a single cue from the effect of a larger collection of causes is based on a simple linear subtraction.

This assumption is made in many formal models of causal learning (Allan, 1980; Cheng & Novick, 1992; Jenkins & Ward, 1965; Rescorla & Wagner, 1972).

Although this mental operation is simple, it poses some normative problems if the causes are independent. We illustrate the nonnormative nature of simple linear addition by using a coin-tossing game as an example. Imagine a game in which two coins (i.e., candidate causes, Coin 1 and Coin 2) are tossed. To win (i.e., for the outcome to occur) either or both of the coins must come up heads. That is to say, a head on either coin represents the necessary generative condition to guarantee an effect on that trial. If both coins are tossed together, either or both of them might come up heads and hence meet the conditions for generating the desired effect (winning). When both coins are tossed, there are four possible permutations of heads and tails: both are heads; the first is heads and the second tails; the first is tails and the second heads; or both are tails. In the first three cases the outcome occurs because each includes at least one head, but in the final case the outcome does not happen – you lose. When only Coin 1 is tossed, we either get heads and win or we get tails and lose (we win on half of the tosses). One can see from the example that if Coin 2 is tossed along with Coin 1 we win on 3/4 of the tosses. If we follow the logic of the linear integration rule, the contribution of tossing Coin 2 to winning is the linear difference between the probability of winning if Coin 1 is tossed (.50) and the probability of winning when both are tossed (.75). Thus by the linear integration rule, the power of Coin 2 is .25. However we know that the actual likelihood of an outcome given the tossing of Coin 2 is .50. Thus the linear integration rule leads to problems of coherence when extrapolating the results of a causal inference to a novel causal context (Cheng, Novick, Liljeholm, & Ford, 2007; Liljeholm & Cheng, 2007). If a coin is tossed first its effectiveness is .50; if it is tossed second it is .25.

In order to solve these and other problems, Cheng (1997) suggested that, from a rational point of view, the combined power of several causes should be computed using a different integration rule. In our case it is the law of compound independent probabilities. It is clear from our example that when the second coin is tossed at least part of its effectiveness is masked because sometimes both it and the first coin meet the conditions for an outcome. Cheng suggested that people can control for these combined effective causes. Instead of using simple addition, she argued that generative causal powers should be combined by means of a noisy-OR integration rule. According to the noisy-OR rule, the probability of the effect if two potential causes, A and B, are present (and no other cause is present) can be computed as

$$p(e|\text{A\&B}) = q_A + q_B - (q_A \cdot q_B), \qquad (1)$$

where q_A and q_B are the generative causal powers of cause A and B, respectively. The power of a cause is simply the proportion of times it produces the outcome, in a hypothetical context in which no other effective causes of the same outcome exist (Cheng, 1997).

If the above assumptions are made it is not correct to assume that the unique contribution of a candidate cause to the probability of the effect may be found by subtracting the effect when it is present from that when it is absent. For example, if we calculated the causal power of Coin 2 by subtracting the probability of winning when only Coin 1 is tossed from the probability of winning when both Coin 1 and Coin 2 are tossed, we would underestimate the real contribution of Coin 2. Following this reasoning, Cheng (1997) proposed an alternative computational analysis of causal induction. According to her Power PC model, when some assumptions are met, the generative power of a candidate cause is given by the rule:

$$q = \frac{p(e|c) - p(e|\neg c)}{1 - p(e|\neg c)}, \qquad (2)$$

where $p(e|c)$ represents the probability of the effect given the cause and $p(e|\neg c)$ represents the probability of the effect given the absence of the cause. Equation 2 can be easily applied to our coin-tossing game. If the probability of winning is .50 when we only toss Coin 1 and .75 when we toss both coins, Equation 2 correctly yields a causal power of .50 for Coin 2. Equation 2 isolates the $p(e|c)$ in a hypothetical situation where no alternative causes of the effect exist. In our example, this would be the probability of getting a head in a situation in which only Coin 2 is tossed.

The idea that causal powers should be combined by means of a noisy-OR integration rule is particularly clear in the Power PC model, but it can be introduced in associative models as well (Danks, Griffiths, & Tenenbaum, 2003). Bayesian models of causal induction do not commit to a specific integration rule, but most formal versions of these models have favored the noisy-OR rule over the linear one (Griffiths & Tenenbaum, 2005, 2009). Precise probabilities aside, the logic of the noisy-OR rule is that the effectiveness of two independent causes, when they are combined, is somewhat more than the effectiveness of the stronger of the two alone and somewhat less than their sum. If people internalize this rule then one might expect that when asked to think about combining independent causes their estimates of the effectiveness of this combination should be higher than either of the individual causes and lower than their sum.

During the last decade, many experiments have tried to find out whether causal inferences spontaneously made by people conform to Equation 2. Unfortunately, this literature has failed to yield a clear and consistent pattern of results (cf., Allan, 2003; Barberia, Baetu, Sansa, & Baker, in press; Buehner, Cheng, & Clifford, 2003; Liljeholm & Cheng, 2007; Lober & Shanks, 2000; Perales & Shanks, 2003, 2007). Most of these experiments have relied on experimental preparations in which the participants received information about the probability of the effect in the presence of the candidate cause and the probability of the effect in the absence of that cause. In these preparations, the participants are expected to use this information to isolate the contribution of the candidate cause from the contribution of

all the background causes that might also influence the effect. Applied to our coin-tossing example, this is analogous to showing people the probability of getting heads when Coins 1 and 2 are tossed together and when Coin 1 is tossed by itself, and subsequently ask them to evaluate the contribution of Coin 2 to winning.

Quite surprisingly, none of those experiments has provided participants with information about the influence of several potential causes and asked them to combine that information so that they can predict the probability of the effect given a combination of causes. In our example, this would be equivalent to showing people the probability of getting heads when either Coin 1 or Coin 2 is tossed and then to ask them to predict the probability of getting heads if we toss both coins simultaneously. This would provide a more direct test of whether people spontaneously use the noisy-OR integration rule when directly asked to predict the effects of several causes. As we have just shown, the noisy-OR rule makes very specific predictions for these situations: When people combine the causal impact of several events, estimates of the combined effect should logically fall in between the value of the individual effects and their sum. Some studies in the associative learning tradition have addressed the issue of how people combine information about several causes when predicting events (Collins & Shanks, 2006; Glautier, Redhead, Thorwart, & Lachnit, 2010; Soto, Vogel, Castillo, & Wagner, 2009; Van Osselaer, Janiszewski, & Cunda, 2004), but most have focused on the predictions of several associative models and have not used experimental designs that would contrast the predictions made on the basis of the noisy-OR rule. The present series of experiments aimed at providing such a contrast.

Experiments 1A and 1B

In Experiments 1A and 1B, the participants were first given information about the impact of several causes of the same outcome. Then, they were presented with different combinations of these causes presented in compound and were asked to report which combinations they expected to be more effective. On the basis of the noisy-OR integration rule, the participants were expected to be indifferent to some choices but not to others. Table 1 depicts the information given to the participants about the effectiveness of five causes in each different experiment. Table 2 shows the predictions of the noisy-OR, linear summation, and an averaging rule for the test compounds in each experiment. Individual participants should prefer compounds with higher values and be indifferent to compounds that have the same value. Even though participants were always forced to choose one of the compounds, and therefore this indifference cannot be detected at the individual level, we should observe that choices distribute similarly between the two compounds at the group level. The table shows that if the participants, in all Experiments, use the noisy-OR they should prefer compound CE to AB and be indifferent to the choice between DE and AB.

Table 1. Design summary of the experiments

Cause	Experiment		
	1A, 2, 4A, 5A	1B, 3A, 3B	4B, 5B
A	.40	.60	.50
B	.40	.65	.50
C	.80	.95	1.00
D	.64	.80	.75
E	.00	.30	.00

Note. The numbers represent the probability of the outcome given each of the causes on each of the experiments.

For example, in Experiment 1A, participants were told that the probability of an effect was .40 in the presence of either A or B, .80 in the presence of C, .64 in the presence of D, and .00 in the presence of E. The cover story in these experiments was carefully chosen so that the participants could assume that the probability of the effect was .00 in the absence of any of these causes. On the basis of this information, participants were asked which compounds they would prefer in order to maximize the likelihood that they would produce the effect. Would the participants prefer a compound including A and B to one including C and E? If participants use a simple nonnormative linear summation strategy to make this decision, they should be indifferent, because the linear addition of A and B (.40 + .40) and the linear addition of C and E (.80 + .00) are exactly the same. However, if participants use a noisy-OR integration rule they should prefer the compound CE, whose ability to produce the effect is $[.80 + .00 - (.80 \times .00)] = .80$, to the compound AB, whose ability to produce the effect is $[.40 + .40 - (.40 \times .40)] = .64$. The participants were also asked to decide between the compound AB and the compound DE (.64 + .00). In this case, if the participants use a linear integration rule, they are expected to choose AB. However, if they use a noisy-OR integration rule they are expected to be indifferent to both compounds. The spe-

Table 2. Predictions made by several integration rules

Compound	Cause 1	Cause 2	Noisy-OR	Linear	Average
Experiments 1A, 2, 4A, 5A					
AB	.40	.40	.64	.80	.40
CE	.80	.00	.80	.80	.40
DE	.64	.00	.64	.64	.32
Experiments 1B, 3A, 3B					
AB	.60	.65	.86	1.25	.625
CE	.95	.30	.965	1.25	.625
DE	.80	.30	.86	1.10	.55
Experiments 4B, 5B					
AB	.50	.50	.75	1.00	.50
CE	1.00	.00	1.00	1.00	.50
DE	.75	.00	.75	.75	.375

cific predictions of the noisy-OR rule are represented in Figure 1.

Experiment 1B used different probabilities than Experiment 1A (see Table 1), but the pattern of choices predicted by the noisy-OR rule is the same as that of Experiment 1A: In all cases the participants were expected to prefer CE to AB and to be indifferent to the AB-DE choice. The main reason why each experiment used different probability sets is to make sure that the results are not dependent on specific details of the parameters used. For example, in Experiment 1A one of the causes, E, had a .00 effectiveness, which might have induced participants to avoid any compound that contained E. Similarly, another cause, D, has a .64 effectiveness, which may be a highly unconventional number. Therefore, in Experiment 1B we used only multiples of five, which are more conventional and commonly reported values, and we changed the probability associated with cause E from .00 to .30. Experiment 1B also included compounds whose linearly-summed values would yield probabilities with values larger than 1.00. For instance, if cause A produces the effect with a .60 probability and B does so with a .65 probability, then if the participants are asked to compute the probability of the effect given A and B and they apply a linear integration rule, then they might expect the effect to occur with a probability of 1.25, which is an impossible probability. The goal was to investigate if this modification would increase the odds of observing behavior more consistent with the noisy-OR rule, because it avoids these inconsistencies.

Method

Participants

All the participants included in Experiments 1A and 1B were psychology students from the University of Deusto. Twenty-six of them volunteered for Experiment 1A and 57 for Experiment 1B.

Procedure and Design

All the materials were presented on a single sheet of paper. The Appendix shows an example of the materials used for the participants in one of the conditions, translated from the Spanish original. The participants were asked to imagine that the cosmetics industry had just discovered substances that could change eye color to pink and that customers were interested in buying products that included these. Then, they were given information about the efficacy of each of the substances (Alpha, Beta, Gamma, Delta, and Omega). For example, they were told that when substance Alpha was injected, eyes turned pink 40% of the time. The substances played the role of causes A–E in Table 1. The assignment of substance names, and hence probabilities of an outcome, to causes was partially counterbalanced, following a latin-square design. The order in which the substances appeared on the sheet of paper was also counterbalanced. After receiving this information, participants

were told that different companies were selling products containing several combinations of these substances and they were asked to choose which product they would buy if they wished to change their eye color to pink. Specifically, participants were asked whether they would prefer a product containing substances A and B to one containing C and E, and whether they would prefer a product containing A and B to another containing D and E. The order of these two questions and the order in which both compounds were presented in each question (AB first vs. CE/DE first) were counterbalanced across participants. The materials and procedure used in each of the three experiments were exactly the same, except for the probabilities of changing the eye color shown in Table 1.

Results and Discussion

The proportion of choices of Compound AB over Compounds CE and DE from Experiments 1A and 1B (and the other four Experiments) are shown in Figure 1. The figure also indicates the proportion of preferences for AB over CE and DE predicted by the noisy-OR rule. Instances where these proportions differ from chance ($p = .50$) by the binomial test are marked with an asterisk. Quite clearly these proportions do not conform to the predictions of the noisy-OR rule. For both experiments, it was predicted that people should prefer CE to AB. However, this pattern was not found in either experiment. The proportion of participants choosing either AB or CE did not differ in Experiment 1A (binomial test, $p = .327$) and, contrary to the prediction of the noisy-OR rule, a reliably high proportion of participants chose AB over CE in Experiment 1B. Similarly, it was predicted that participants should have no particular preference when choosing between AB and DE. However, they showed a clear preference for AB in both experiments. These results are more consistent with the linear summation rule that predicts the participants should be indifferent to AB or CE but prefer AB to DE.

Experiment 2

The previous experiments were conducted with psychology students, who receive training in probability theory and statistics as part of their formal education. In spite of this they did not seem to use the more normative noisy-OR rule. Nonetheless, it is possible that the results are not representative of the performance that one would find in the general, untrained population. To address this, we replicated Experiment 1A with middle-school age students.

Method

Participants, Procedure, and Design

Thirty-eight school students participated in Experiment 2. All the participants were 12 or 13 years old. The procedural

details and materials were exactly the same as in Experiments 1A–1B. Experiment 2 used the probabilities of Experiment 1A (see Table 1).

Results and Discussion

As can be seen in Figure 1, the results of Experiment 2 were very similar to those of Experiment 1A, which was based in the same probability set. Again participants did not prefer CE to AB. The modest preference for AB over CE did not reach statistical significance (binomial test, $p = .143$). However, participants preferred the compound AB to the compound DE. So, again, these results are at odds with the predictions of the noisy-OR integration rule and are consistent with linear summation. This implies that the results of the previous experiments cannot be attributed to the peculiarities of psychology students.

Experiments 3A and 3B

In the previous experiments, the information was provided to the participants as probabilities. However, people commonly fail to use information presented as probabilities or percentages correctly (e.g., Gigerenzer & Hoffrage, 1995; Sloman, Over, Slovak, & Stibel, 2003). Therefore, in Experiment 3A we presented information in a frequency format (e.g., we told participants that when cause A was present the effect occurred in 80 out of 200 occasions, instead of telling them that produced the effect 40% of the time). In Experiment 3B we used pie charts to present the probabilities in a graphical manner.

    Additionally, we also wanted to ensure that our results would generalize to other scenarios. Therefore we used a new cover story for Experiments 3A and 3B. Testing the generality of the previous results is not only important for methodological reasons, but also to rule out alternative explanations. Although the results of our previous experiments are in principle inconsistent with the predictions of the noisy-OR rule, they can be accommodated by assuming that the participants believed the two elements might interact in producing the effect. In other words, the presence of one substance would alter the ability of the other to produce the effect, so that the effect of the compound might be different than the sum of its elements. In the following experiments we used different materials to ensure that the results of Experiments 1–2 were not due to specific details of that particular cover story that might suggest potential interactions between causes.

Method

Participants, Procedure, and Design

Fifty-three undergraduate psychology students volunteered to participate in Experiment 3A and 34 volunteered for

Experiment 3B. In both experiments the set of probabilities from Experiment 1B was again used (see Table 1). The new cover story was inspired by a short science-fiction story (Chiang, 2002) that we assumed most participants would be unfamiliar with. The participants were asked to imagine that scientists have discovered a treatment that makes people insensitive to physical beauty. Some people demand this treatment because they want to be more sensitive to other people's inner beauty than to their physical appearance. Participants were given information about the effectiveness of five different proteins that could produce this effect. In Experiment 3A this probabilistic information was given in a frequency format (e.g., "120 out of 200 people injected with protein Alfa stop perceiving physical beauty"). In Experiment 3B, the information was given by means of pie charts. After seeing this information, the participants were asked to say which compounds they would choose, if they wanted to stop being influenced by the physical beauty of other people.

Results and Discussion

The pattern of results, depicted in Figure 1, is very similar to that of previous experiments. Contrary to the predictions of the noisy-OR rule, rather than showing indifference, the majority of participants in both experiments reliably preferred the compound AB to the compound DE. In Experiment 3A a reliable majority preferred AB to CE, but in Experiment 3B they showed no clear preference for either compound at the group level (binomial test, $p = .392$). These results strengthen the idea that the results observed in Experiments 1 and 2 are not attributable to either peculiarities of the cover story or to the precise way in which causal information was presented to participants.

Experiments 4A and 4B

In all the previous experiments, the results disagreed systematically with the predictions of the noisy-OR integration rule. If anything, they are more consistent with alternative integration rules such as linear summation, according to which the causal impact of several factors is equal to the simple sum of the causal impact of the elements. The results are also quite consistent with the averaging strategy, shown in the right-most column of Table 2, according to which participants might estimate that the causal impact of a compound is the average of the causal impact of the elements. Although the binary decision test we have used through Experiments 1–3 was well suited for contrasting the predictions of the noisy-OR integration rule, it might not be the best method for an exploratory analysis of alternative integration rules that might be used by the participants. Therefore, in addition to asking participants to choose between AB and CE and between AB and DE, in Experiments 4A and 4B we also requested ratings of the probability of the effect given AB, CE, and DE.

Additionally, in Experiment 4B we asked participants to decide between AB and C and between AB and D. As can be seen in Table 1, in Experiment 4B, element E lacks any power to produce the effect (i.e., the probability of the outcome in the presence of E is zero). Therefore, removing it from any compound should have no impact on choices: The participants' behavior should be largely similar when deciding between AB and CE and when deciding between AB and C. However, if removing this component has an effect on the pattern of choices, this might provide an interesting insight into the strategy used by the participants. The linear and the noisy-OR rules predict no effect of adding the neutral compound E, but if a participant relies on an averaging strategy then adding a noneffective cause (such as E) to a relevant cause (such as C or D) should reduce the perceived causal impact. In this case, preference for C and D should be higher when they are presented in isolation than when they are presented in compound with E.

Method

Participants and Apparatus

Forty-eight psychology students from McGill University participated in Experiment 4A and 58 psychology students from the University of Deusto took part in Experiment 4B. Three participants in Experiment 4A and two in Experiment 4B wrote open responses (e.g., "some value between 0 and 40") in the space provided for their probability ratings. The data from these participants were removed from the analyses.

Design and Procedure

The pink-eyes cover story used in Experiments 1 and 2 was used in Experiments 4A and 4B. As shown in Table 1, the participants in Experiment 4A were exposed to the set of probabilities used in Experiments 1A and 2A. For Experiment 4B, a new probability set was used and is summarized in Table 1. These probabilities were chosen so that, as in the previous experiments, preference of CE over AB could be taken as an evidence for the noisy-OR integration rule. As mentioned in the introduction to this set of experiments, Experiment 4B also differed from the previous ones in that participants were not only asked to choose between AB and CE, on the one hand, and AB and DE, on the other, but also between AB and each of the elements C and D. As in the previous experiments, the order in which information about the different causes was given and the order of the choices were partially counterbalanced following a latin-square procedure.

Unlike the previous experiments, after making their choices, participants from both Experiments 4A and 4B were also asked to rate the probability of the effect given the combinations AB, CE, and DE by means of the following question: "If you inject substances A and B into 100 people, how many will get the pink eyes?" These extra questions were added to further explore which strategy

might be responsible for the pattern of choices observed. The order in which these three judgments were requested was also counterbalanced following a latin-square design.

Results and Discussion

The choices made by participants in Experiments 4A and 4B are shown in Figures 1 and 2, respectively. As can be seen, the pattern of preference of AB over CE and DE in Experiment 4A replicates the results of previous experiments. Again a larger proportion of participants preferred AB to DE but, in spite of a trend in favor of CE, there was no reliable difference in the proportions choosing AB or CE. The same pattern of results concerning the AB, CE, and DE compounds was found in Experiment 4B. However, a different pattern was observed when participants chose between AB and the elements C and D now presented alone. Participants tended to choose C and D over AB more when they were presented on their own than when they were presented in compound with E. This unexpected effect of removing cause E from the CE and DE compounds is not consistent with the noisy-OR integration rule but is also not consistent with the simpler, linear integration rule. However, as discussed above, it is consistent with an averaging strategy.

Participants also estimated the number of times out of 100 that each causal configuration would turn the eyes pink. Figure 3 plots the number of participants who made various estimates for configurations AB, CE, and DE. The descriptive statistics (mean and standard error) for each condition are shown in each graph. As explained in the Introduction, on the basis of the noisy-OR integration rule, we would expect participants' ratings to fall somewhere in between the individual effects of each of the causes and their sum. However, only a small number of participants gave ratings within this range. The frequency graphs for the AB compound show that the most frequent responses correspond either to the linear sum of the probability of the effect given A and the probability of the effect given B or the arithmetic mean of the two probabilities. Ratings for CE and DE are less informative because the responses predicted by the linear and the noisy-OR rules are exactly the same. However, it is still interesting to note that many participants in those conditions provided responses equal to the arithmetic mean of the probabilities of the elements. This pattern of judgments further suggests that the reduction in preference for AB over CE and DE caused by removing element E from the compounds may be due to the fact that some participants rely on an averaging strategy.

Experiments 5A and 5B

In Experiments 5A and 5B we implemented a number of modifications that we expected would promote behavior consistent with the noisy-OR rule. First, we used two cover stories that precluded the perception of any potential rational interaction between causes. In Experiment 5A, par-
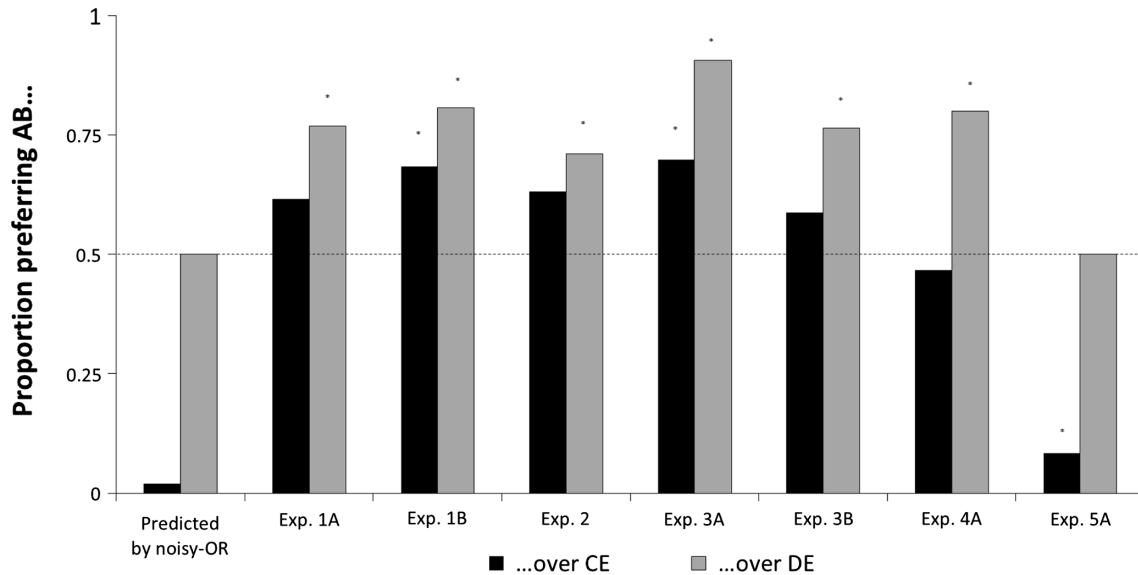
Figure 1. Pattern of preferences predicted by the noisy-OR rule and actual proportion of participants preferring the compound AB over CE and over DE in Experiments 1A–4A and Experiment 5A. The 0.50 axis represents indifference between AB and the other options. Proportions above 0.50 represent a preference for AB over the other options. Asterisks are placed upon values that are significantly different from .50 in a binomial test with $\alpha = .05$.

ticipants were asked to imagine that they were playing a coin-tossing game similar to that described in the Introduction. They were given the probabilities of obtaining heads in several biased coins and then were asked which combinations of coins would provide the higher probabilities of winning (i.e., getting a head in at least one of the coins). In Experiment 5B, participants were alerted about several independent potential dangers of living in a fictitious planet. Then they were asked to choose among several routes for a mission on that planet. Each route involved a different combination of dangers. Most importantly, according to the
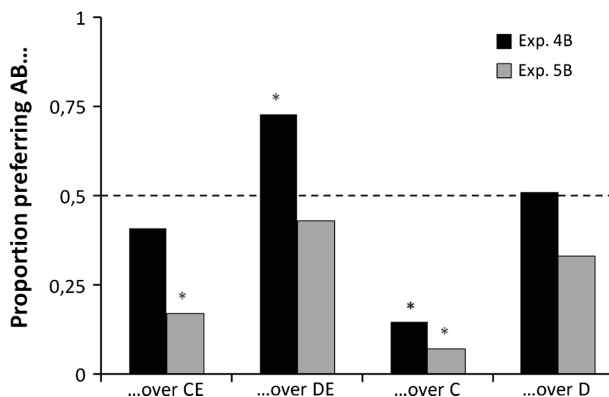
description that we gave to participants, each of these dangerous events could produce the death by means of completely different mechanisms that should not interact with one another. Therefore, participants should have no reason to assume the causes might interact.

As in Experiments 4A and 4B, participants not only had to choose between different combinations of causes, but they also had to provide probability ratings of each of the combinations. Furthermore, in Experiment 5B we presented the information using a frequency format that has been previously used in studies that found support for the noisy-OR integration rule (e.g., Liljeholm & Cheng, 2009). We also used different probability sets in both experiments (see Table 1).

Method

Participants and Apparatus

Forty-nine psychology students from McGill University participated in Experiment 5A and 30 engineering and medicine students from the University of the Basque Country took part in Experiment 5B. All of them performed the experiment on paper and pencil questionnaires. One participant in Experiment 5A failed to provide probability rating so was removed from the final analyses.

Design and Procedure

Two new cover stories were used in Experiments 5A and 5B. In Experiment 5A participants were asked to imagine

Figure 2. Proportion of participants preferring the compound AB over CE and over DE in Experiments 4B and 5B. To make the comparison with Experiment 4B easier, preferences were reverse-scored in Experiment 5B (see the main text for further details). Asterisks are placed upon values that are significantly different from .50 with $\alpha = .05$.
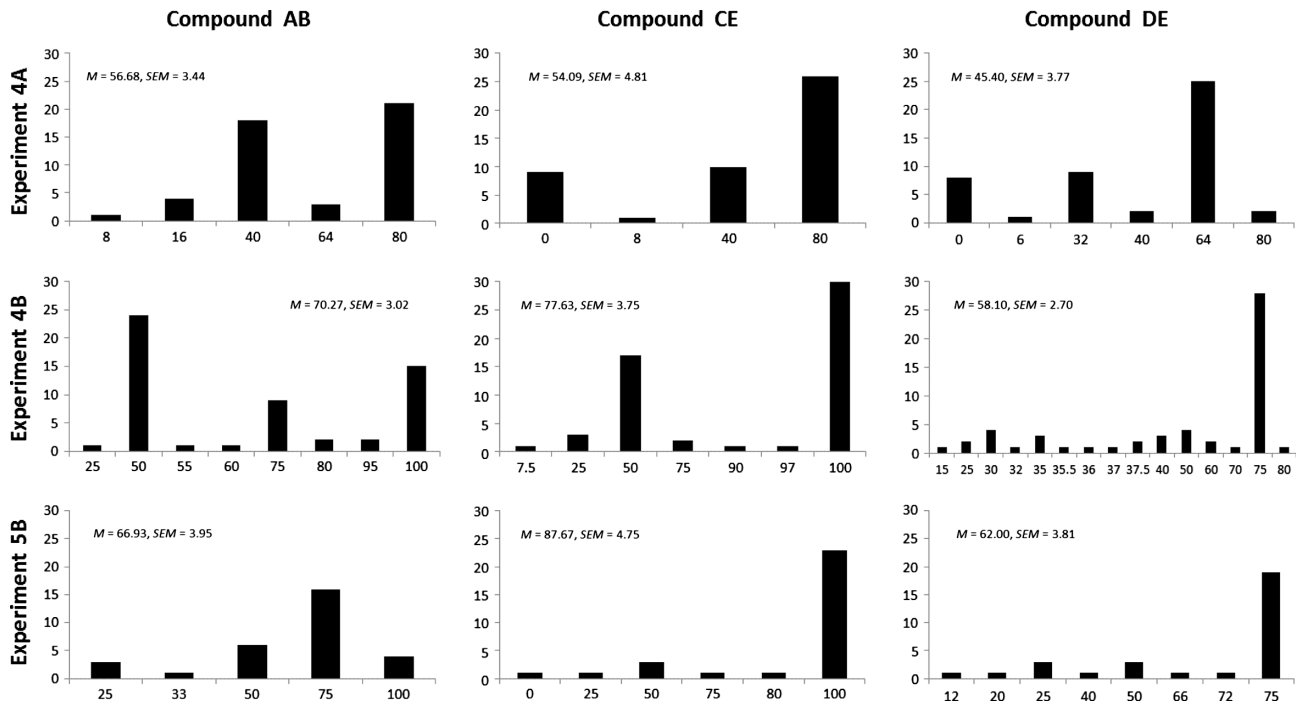
Figure 3. Frequencies of responses to the probability ratings requested to the participants in Experiments 4A, 4B, and 5B. The *x*-axis represents specific values of judgments given by participants and the *y*-axis the frequency of those values in our sample.

that they were playing a coin-tossing game with three different biased coins. In this game, they would win if they obtained at least one head. The three coins produced heads 40%, 80%, and 64% of the times, respectively. Then they were asked to choose whether they would prefer to toss twice the coin with the 40% chance of winning (and win if they got at least a head in one of the tosses) or to toss just once the coin with the 80% chance of winning. Similarly, they had to choose whether they preferred tossing the 40% coin twice or tossing the 64% just once. After making these decisions, participants were asked to rate the probability of winning if they tossed the 40% coin twice.

In Experiment 5B we used a different science fiction scenario. Participants were asked to imagine that they were scientists in charge for colonizing an alien planet. They had been ordered to install an atmosphere purification system that would allow humans to live on the planet. They were also told that the routes to the ideal candidate locations for the purification system were dangerous. They were given information about five specific problems that they might encounter. They might: Approach a methane ocean, which might cause death by burning; step onto mercury quicksand, which might drown them; swallow alien lichen, which might poison them; fall into a cave, causing death by impact; and, finally, be exposed to electric storms, which might cause death by electrocution. These dangers were mentioned in random order and were randomly assigned to play the role of causes A–E in Table 1. On a second sheet they were given detailed information about the probability that each of these events might cause death. Specifically, for each event, they were told how many of the last 24

explorers who had faced those situations had died. This information was shown by means of frequency charts in which each explorer who had survived was represented as a smiley face and each explorer who had died was represented as a skull. In these charts, all the smiley faces were grouped together as were all the skulls (for a similar procedure, see Liljeholm & Cheng, 2009). The final sheet of the questionnaire asked whether the participants would prefer to use a route where they would face dangers A and B to one in which they would face dangers C and E. Similarly, they were asked to choose between AB and DE, between AB and C, and between AB and D. Unlike all the previous experiments, because participants here had to avoid a negative outcome, they should choose the less effective combination of causes (instead of the more effective). Finally, as in Experiments 4A and 4B, participants were asked to rate the probability of dying if they faced situations AB, situations CE, and situations DE. As in previous experiments, these questions were asked in a frequency format (e.g., "Out of 100 new explorers taking a route that exposed them to X and Y, how many would die?"). We expected that framing this test question in terms of frequencies would promote more normative responses (Gigerenzer & Hoffrage, 1995; Sloman et al., 2003).

Results and Discussion

In contrast to the rest of the results represented in Figure 1, the pattern of choices made by participants in Experiment 5A was surprisingly consistent with the predictions of the
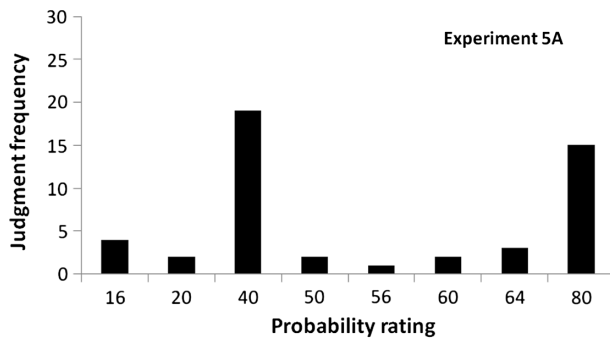
Figure 4. Frequencies of responses to the probability ratings requested to the participants in Experiment 5A.

noisy-OR rule: They preferred to toss a single coin with an 80% chance of winning over twice tossing a coin with a 40% chance of winning. They were, however, indifferent to the choice between tossing the 40% coin twice or the 64% coin once. Their numerical ratings of the probability of winning if they tossed the 40% coin twice are represented in Figure 4. These ratings are in stark contrast with the pattern of decisions. Following the noisy-OR rule, we would expect most ratings to fall within the 40–80 interval. However, few participants gave responses within that range. Instead, most of them responded either 40 or 80. These numerical responses are consistent with the use of an averaging or a linear summation rule, respectively.

The preference responses for Experiment 5B are depicted in Figure 2. To facilitate comparison with the previous experiments (in which the outcome was a desired event) the choices made by participants in Experiment 5B (in which the outcome was not desired) were reverse-scored. As can be seen in Figure 2, participants did not think that the compound of two causes with a 50% efficacy was more dangerous than the compound of a 75% effective cause and a 0% effective cause. However, they thought the combination of the 100% effective cause and the 0% effective cause was more dangerous than the combination of the two 50% effective causes. This pattern of results is consistent with the noisy-OR rule. Unlike in Experiment 4B (also depicted in Figure 2) the pattern of choices was not influenced by the presence or absence of the 0% effective cause, suggesting that in this experiment participants did not average the effect of several causes. Figure 3 shows the ratings of the probability of the effect given the combinations AB, CE, and DE. The first of them provides the most valuable insight into the strategies used by participants. As can be seen, most participants thought that the probability of the effect given two 50% effective causes was 75%. Again, this is consistent with the noisy-OR integration rule. Note, however, that one third of participants gave ratings equal to either 50 or 100. These are the values from the averaging and linear integration rules, respectively. Moreover four of the participants gave ratings that were outside the 50–100 range. Therefore even in this experiment only a bare majority of participants gave ratings consistent with the noisy-OR rule.

General Discussion

The results of the present series of experiments are more consistent with the idea that people use a diversity of integration rules than with the assumption that there must be a single, monolithic rule (either the linear or the noisy-OR one) that is systematically adopted in all cases. On the one hand, the preference choices of participants in Experiments 1A, 1B, 2, 3A, 3B, and 4A agreed with the predictions of the linear rule (see Figure 1). On the other hand, choices in Experiments 5A and 5B were consistent with the predictions of the noisy-OR rule (see Figures 1 and 2). However the participants' ratings of the probabilities weaken this conclusion. The probability ratings provided in Experiments 4A–4B and 5A–5B show a complex pattern (see Figures 3 and 4). With only the exception of Experiment 5B, most of the ratings suggested that participants were using either a linear integration rule or, alternatively, an averaging heuristic. Note that the averaging strategy can account for all the preference choices that can be explained by the linear rule, plus the negative effect of adding a 0% effective cause to a compound found in Experiment 4B. Furthermore, even in the experiments that showed the pattern of choices predicted by the noisy-OR rule, the probability ratings of many participants still seemed to be based on an averaging strategy.

The diversity of strategies used by participants in the present series of experiments converges with the results of previous studies on summation in causal learning and reasoning. Although none of those experiments was specifically designed to contrast the predictions of the noisy-OR integration rule with those of the linear rule, their results show that summation is highly unstable and that it can change dramatically depending on minor procedural details. For instance, Glautier et al. (2010) found that increasing the similarity between two causes by means of adding a common feature reduces and even reverses summation. Similarly, Van Osselaer et al. (2004) found that seemingly minor alterations in the procedure for collecting judgments or presenting the information could completely abolish summation. Note that this reduction in the size of summation is what one would expect on the basis of an averaging strategy: If participants average the causal power of each factor, then the whole need be better than either of its parts.

Most interestingly, some experiments show that even within a single experimental task and with the same materials, participants might sometimes use different integration rules. Waldmann (2007) found that subtle manipulations could influence whether participants combined causes using an averaging or an additive (linear or noisy-OR) rule. However, when participants had to subtract instead of combine causal powers, their responses tended to be consistent with an additive rule most of the time, even when the same materials and instructions were used. This is consistent with the divergences that we found in some of our experiments between the pattern of choices and the probability ratings. For instance, in Experiment 5A, participants' choices were consistent with the noisy-OR rule. But their probability rat-

ings were more consistent with either the averaging or the linear rules.

The results of our experiments pose problems for models relying exclusively on either the linear rule or on the noisy-OR rule. Those models can only accommodate our results by assuming that, in spite of our efforts to prevent it, participants assumed potential interactions between causes. From this perspective, theories that rely on the linear integration rule could explain behavior that seems consistent with the noisy-OR integration rule by assuming that participants believed there were interactions. Alternatively, the theories that rely on the noisy-OR rule might explain behavior consistent with the linear rule by making exactly the opposite assumption. Based on this divergent prediction, an interesting idea for future research would be to test whether participants are more likely to believe there are interactions in scenarios that promote behavior consistent with the linear rule or in scenarios that promote behavior consistent with the noisy-OR rule. Given that we did not ask whether participants assumed interactions between causes, our data are not suited to test this prediction. However, it is interesting to note that we observed decisions and judgments were more consistent with the noisy-OR rule in the two experiments (5A and 5B) that most strongly emphasized the independence or lack of interaction between causes. In principle, this supports the predictions of the noisy-OR rule. As explained above, Experiments 1–4 used cover stories in which drugs and chemical substances were used as potential causes of physiological and psychological effects. Perhaps the participants' familiarity with the, sometimes complex, effects of drugs might have encouraged them to perceive potential interactions between the causes. This assumption, in turn, would be responsible for any behavior departing noticeably from the predictions of the noisy-OR rule. Although plausible, this assumption does not account well for previous experimental findings, given that experiments seemingly consistent with the noisy-OR rule also used cover stories involving chemicals and drugs that are fairly similar to the materials used in our Experiments 1–4 (e.g., Liljeholm & Cheng, 2007, 2009). In light of these discrepancies, we suggest that future experiments contrasting these two integration rules include a direct measure of whether or not participants believe there to be potential interactions.

Stressing the noninteractivity of causes might not be the only factor improving the normativity of judgments and choices in our last experiments. Quite interestingly, the only experiment that yielded results perfectly consistent with the predictions of the noisy-OR rule (Experiment 5B) not only emphasized that each cause acted by means of different and independent mechanisms, but also included an important change in the way information was presented. Specifically, while the other experiments presented the information by means of probabilities, frequencies, or pie charts, in Experiment 5B the information was presented in frequency charts in which each instance of the effect was represented as an individual item among other instances where the effect was absent. Quite interestingly, this information format has been used extensively in other experiments that found support for rational models relying on the noisy-OR rule (e.g.,

Buehner et al., 2003; Liljeholm & Cheng, 2009; Novick & Cheng, 2004). Given that the information format is known to have an effect on causal judgments (Shanks, 1991; Vallée-Tourangeau, Payton & Murphy, 2008; Ward & Jenkins, 1965), this coincidence might deserve further attention. It is possible that presenting the information in numeric format, as is frequently done in many causal learning experiments, invites participants to combine the information using simple algebraic operations that, despite their intuitive appeal, yield nonnormative behavior. Unfortunately, our Experiment 5B differed from previous experiments in other procedural details, apart from the presentation format, and therefore it remains unclear to what extent the information format, rather than any of the other differences, is responsible for these results. However, in light of the prevalence of using this method of presenting information in many of the experiments supporting the noisy-OR rule, an in-depth exploration of the effects of information formats on the normativity of causal judgment and decision-making seems a promising idea for future research.

In the Introduction we mentioned that some associative models rely on the linear integration rule (Rescorla & Wagner, 1972) while some others rely on the noisy-OR rule (Danks et al., 2003). Interestingly, there is a third category of models that, even without committing specifically to any rule, make predictions consistent with an averaging strategy. In contrast to so-called elemental models, configural models (Kruschke, 1992; Pearce, 1987, 1994) assume that participants do not encode causes individually, but treat every combination of causes as a completely different entity. Very interestingly, this idea implicitly implements the assumption that causes interact, given that from this perspective the effect of a combination of causes need not be the same as the sum of the effects of its elements. In configural models, the extent to which participants generalize what they know about a single cause to a novel combination of causes depends on the similarity between the element and the compound. In the case of our experiments, according to the rule proposed by Pearce (1987, 1994), the similarity between A and AB would be .50, because half of the elements in AB are present in A. For the same reasons, the similarity between B and AB would also be .50. If the associative strength of A is .40 and that of B is .40 as well, as in many of our experiments, then the associative strength that generalizes to the compound AB is $(.40 \times .50) + (.40 \times .50) = .40$. In other words, these models predict that the associative strength of the compound AB should be equal to the average of the associative strengths of A and B. Therefore, the generalization rules used by these models could potentially explain why many participants showed a pattern of responses consistent with an averaging rule.

Quite interestingly, just as we found that the behavior of participants can be consistent with different integration rules, research on associative learning shows that people and other animals sometimes behave as predicted by elemental models and sometimes behave as predicted by configural models (for a review, see Melchers, Shanks, & Lachnit, 2008). Therefore, instead of deciding which of

these models fits better to the data, researchers are now trying to find out which conditions promote elemental or configural learning and to develop formal models that can incorporate both processing strategies (Schmajuk & DiCarlo, 1992; Wagner, 2003). In light of the diversity of strategies used by participants in our experiments and in the previous reports on summation, we think that the area of causal learning and reasoning would benefit from a similar approach: Instead of deciding whether it is the linear or the noisy-OR rules that underlies causal learning, it seems more sensible to explore when and how participants behave according to those or other rules and to develop models that can incorporate a diversity of strategies.

Acknowledgments

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*, 147–149.

Allan, L. G. (2003). Assessing power PC. *Learning & Behavior, 31*, 192–204.

Barberia, I., Baetu, I., Sansa, J., & Baker, A. G. (in press). When is a cause the "same"? Incoherent generalization across contexts. *The Quarterly Journal of Experimental Psychology*. Advance online publication. doi:10.1080/17470218.2013. 804102

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.

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.

Cheng, P. W., Novick, L. R., Liljeholm, M., & Ford, C. (2007). Explaining four psychological asymmetries in causal reasoning: Implications of causal assumptions for coherence. In M. O'Rourke (Ed.), *Topics in contemporary philosophy, Vol. 4. Explanation and causation* (pp. 1–32). Cambridge, MA: MIT Press.

Chiang, T. (2002). *Stories of your life and others*. New York, NY: Tor Books.

Collins, D. J., & Shanks, D. R. (2006). Summation in causal learning: Elemental processing or configural generalization? *The Quarterly Journal of Experimental Psychology, 59*, 1524–1534.

Danks, D., Griffiths, T. L., & Tenenbaum, J. B. (2003). Dynamical causal learning. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems* (Vol. 15, pp. 67–74). Cambridge, MA: MIT Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.

Glautier, S., Redhead, E., Thorwart, A., & Lachnit, H. (2010). Reduced summation with common features in causal judgments. *Experimental Psychology, 57*, 252–259.

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

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661–716.

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

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Liljeholm, M., & Cheng, P. W. (2007). When is a cause the "same"? Coherent generalization across contexts. *Psychological Science, 18*, 1014–1021.

Liljeholm, M., & Cheng, P. W. (2009). The influence of virtual sample size on confidence and causal-strength judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 157–172.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107*, 195–212.

Melchers, K. G., Shanks, D. R., & Lachnit, H. (2008). Stimulus coding in human associative learning: Flexible representations of parts and wholes. *Behavioural Processes, 77*, 413–427.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111*, 455–485.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review, 94*, 61–73.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review, 101*, 587–607.

Perales, J. C., & Shanks, D. R. (2003). Normative and descriptive accounts of the influence of power and contingency on causal judgment. *The Quarterly Journal of Experimental Psychology, 56A*, 977–1007.

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

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appelton-Century-Crofts.

Schmajuk, N. A., & DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. *Psychological Review, 99*, 268–305.

Shanks, D. R. (1991). On similarities between causal judgments in experienced and described situations. *Psychological Science, 2*, 341–350.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91*, 296–309.

Soto, F. A., Vogel, E. H., Castillo, R. D., & Wagner, A. R. (2009). Generality of the summation effect in human causal learning. *The Quarterly Journal of Experimental Psychology, 62*, 877–889.

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.

Van Osselaer, S. J. J., Janiszewski, C., & Cunda, M. Jr. (2004). Stimulus generalization in two associative learning processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 626–638.

Wagner, A. R. (2003). Context-sensitive elemental theory. *The Quarterly Journal of Experimental Psychology, 56B*, 7–29.

Waldmann, M. R. (2007). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science, 31*, 233–256.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 19*, 231–241.

Miguel A. Vadillo

Division of Psychology and Language Sciences
University College London
26 Bedford Way
London WC1H 0AH
UK
Tel. +44 20 7679 5364
E-mail m.vadillo@ucl.ac.uk

Appendix

Please, answer the questions in the order they are written. Once you have answered one question, please do not come back to this question later to change your answer. We are interested in your first answers.

Try to imagine the following situation. The cosmetics industry has recently been revolutionized by the discovery of a number of natural substances that can change the color of the people's eyes. Interestingly, the color most demanded by consumers is not any of the normal ones (blue, green, brown...), but pink. Unfortunately, there are just a few known substances that produce this eye color.

For the time being, scientists have identified only five substances that may cause the pink eyes. The results of these studies are as follows:

– If you inject substance Alpha, eye color turns pink 40% of the time.
– If you inject substance Beta, eye color turns pink 40% of the time.
– If you inject substance Gamma, eye color turns pink 80% of the time.
– If you inject substance Delta, eye color turns pink 64% of the time.
– If you inject substance Omega, eye color turns pink 0% of the time.

There are several cosmetic companies that sell products to change the color of the eyes. Most of these products are based on different compounds of one or several of these substances. Imagine that you also want to change the color of your eyes and answer the following questions accordingly:

1. If there is a product that contains a compound of the Alpha and Beta substances, and another that contains a compound of the Gamma and Omega substances, which one would you choose? (Please circle your choice)
Alpha and Beta / Gamma and Omega

2. If there is a product that contains a compound of the Alpha and Beta substances, and another that contains a compound of the Delta and Omega substances, which one would you choose? (Please circle your choice)
Alpha and Beta / Delta and Omega