Special Section

Causal Learning: Association Versus Computation

Anthony Dickinson[1]

Department of Experimental Psychology, University of Cambridge, Cambridge, England

Abstract

Causal learning enables humans and other animals not only to predict important events or outcomes, but also to control their occurrence in the service of needs and desires. Computational theories assume that causal judgments are based on an estimate of the contingency between a causal cue and an outcome. However, human causal learning exhibits many of the characteristics of the associative learning processes thought to underlie animal conditioning. One problem for associative theory arises from the finding that judgments of the causal power of a cue can be revalued retrospectively after learning episodes when that cue is not present. However, if retrieved representations of cues can support learning, retrospective revaluation is anticipated by modified versions of standard associative theories.

Keywords

learning; conditioning; associationism; causation

The ability to control our environment in the service of our needs and desires depends on learning about its causal structure. If we did not have knowledge of the consequences of our actions, our ability to gain valuable resources and avoid dangerous and deleterious situations would be seriously compromised. It is not surprising, therefore, that judgments of the causal power of an event, be it an action or a stimulus, in producing an outcome often accurately reflect the strength of the real causal relationship.

The effectiveness of a possible cause depends not only on the likelihood that the outcome follows the cause reliably, but also on the likelihood that the outcome happens in the absence of the cause. For example, if I suffer a skin rash relatively frequently after eating fish, I should attribute the allergy to this food only if I am free from the rash at times when I have not eaten the fish. If I am just as likely to have a rash on days when I have not eaten fish as on days when I have, there is no contingency between the fish and the allergy, and the fish is unlikely to be the causal agent. Rather, the allergy is most probably caused by some background feature of my environment, such the presence of Rolly, my dog.

Causal judgments are sensitive to cause-outcome contingencies. Indeed, these judgments often reflect accurately the difference between the probabilities of the outcome given the presence and given the absence of the putative cause. This finding has led to the idea that causal judgments are based on the computation of event contingencies (e.g., Cheng, 1997). However, the parallels between causal judgment and animal conditioning suggest that causal beliefs can be acquired by the associative learning processes thought to underlie conditioning (Dickinson, 2001).

CAUSAL INTERACTIONS

According to associative theories of conditioning, surprising outcomes enter into stronger associations than do predicted outcomes. I illustrate this effect of surprise with a recent study using a food-allergy scenario (Aitken, Larkin, & Dickinson, 2000). In this scenario, my colleagues and I asked participants to take the role of a food allergist attempting to determine which foods cause an allergic reaction in a hypothetical patient by observing the consequences of a number of meals eaten by the patient. Information about the food eaten by the patient in each meal and whether or not a reaction oc-

Table 1. *Cue-outcome contingencies*

	Training	
Contingency	Stage 1	Stage 2
Control	—	CY+
Blocking	X+	BX+
Preventive	X+	PX−
Super-learning	X+ and PX−	SP+
Generative (retrospective)	GX+	X−
Control (retrospective)	CY+	—
Preventive (retrospective)	PX−	X+
Neutral (retrospective)	NY−	Y−

Note. Each contingency involved two successive stages of training in which various food cues (B, C, P, N, S, X, and Y) were either paired with the allergic reaction as an outcome (+) or presented in the absence of this outcome (−). For example, CY+ indicates the participants received trials on each of which a compound of the food cues C and Y was paired with the allergic reaction, whereas X− shows that food cue X was presented in the absence of the allergic reaction for a number of trials.

curred following the meal was presented on a computer monitor.

In the control contingency (Table 1), the patient ate two foods, C and Y, in each meal, and these meals were followed by the allergic reaction, which (at least following the initial meals) would have been surprising or unexpected. Consequently, the participants should have learned a causal relationship between food C (and food Y) and the allergic reaction. In order to determine whether causal learning had occurred, after training with the CY meals, we asked the participants to rate how effective they thought food C was in causing the allergic reaction. As Figure 1 shows, food C received a positive causal rating, indicating that they thought this food was a generative cause of the allergic reaction.

The role of surprise in causal learning was demonstrated by the judgments for a second, blocking contingency. Again, compound meals, in this contingency consisting of foods B and X, were paired with the allergic reaction. The only difference from the control contingency was that the occurrence of the allergic reaction following these BX meals was rendered unsurprising, or predicted, by an initial stage of training in which food X was a sufficient cause of the allergic reaction. In this first stage, patients received meals consisting of food X alone, and each meal was followed by the allergic reaction (Table 1). Consequently, the occurrence of the outcome following the compound BX meals in Stage 2 should have been unsurprising because the allergic reaction was predicted by the presence of food X, and hence the participants should have learned little about the relationship between food B and the allergic reaction. In accord with this prediction, the causal power of food B was rated as very low (Fig. 1), and the initial training of food X is said to have blocked learning about food B.

In addition to predicting that causal learning should be blocked when the outcome is expected, associative theory anticipates that causal learning should be enhanced when the outcome is super-surprising. Rendering an outcome super-surprising depends on first training participants to view a cue as a preventive rather than a generative cause. To establish food P as a preventive cause in our allergy task, we first showed the participants a series of meals in which food X was paired with the allergic reaction. In the second stage of training, we presented food X in a compound meal with food P. The allergic reaction did not occur following the compound meals (Table 1), thereby establishing that food P acted as a prophylactic, or preventive, cause for the reaction caused by food X. The important feature of the PX meals is that the nonoccurrence of the outcome is surprising, and associative theory predicts that this form of surprise should result in cue P acquiring a negative association with the outcome (see Larkin, Aitken, & Dickinson, 1998). A negative association endows the cue with inhibitory properties in conditioning and represents that the cue prevents an expected outcome in causal learning. In our studies, this form of learning was manifest as a negative causal rating for food P (Fig. 1).

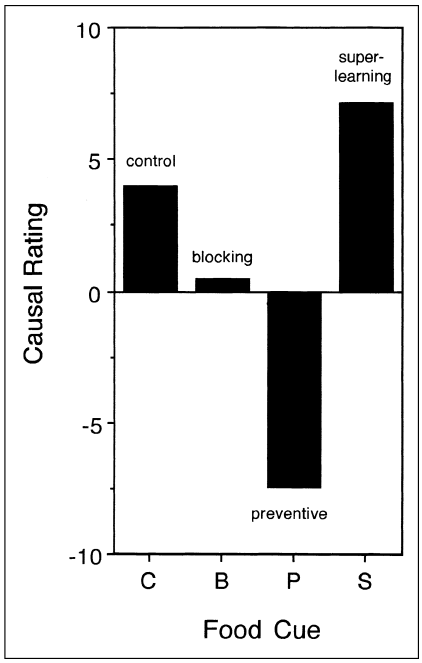

Fig. 1. Mean causal ratings for food cues C, B, P, and S trained under the control, blocking, preventive, and super-learning contingencies (see Table 1), respectively. Positive and negative ratings represent generative and preventive causal relationships, respectively.

Preventive causal learning provides a vehicle for rendering the occurrence of an outcome super-surprising. Associative theory predicts that a super-surprising outcome should enhance generative causal learning above the normal or control level, or, in other words, produce super-learning (Aitken et al., 2000). Having established food P as a preventive cause in Stage 1 by training with an X+ and PX− contingency, we assessed this prediction by giving the hypothetical patient meals consisting of a new food, S, along with the preventive food P and following each of these meals with the allergic reaction (Table 1, super-learning contingency). The presence of food P predicted that this outcome should have been prevented, so its occurrence was super-surprising. In accord with the prediction of associative theory, the participants demonstrated super-learning: Food S yielded the highest causal rating of all the cues (Fig. 1).

RETROSPECTIVE REVALUATION

Blocking, preventive learning, and super-learning all have analogies in animal conditioning and, taken together, provide evidence for the role of associative processes in causal learning. In its standard forms, however, associative theory places implausible constraints on causal inference. Consider a case in which the allergist observes first that a compound GX meal is followed by an allergic reaction and then, in a second stage, that food X by itself fails to generate the outcome (Table 1, retrospective generative contingency). The rational inference from this contingency is that food G alone is a sufficient, generative cause of the outcome, and yet this simple inference defeats standard associative theory. Although the initial compound GX training endows food G with a moderate associative strength, the subsequent X-alone trials, which demonstrate that food X by itself is not sufficient to cause the allergic reaction, should have no impact on the associative strength of food G. Standard associative theory assumes that the associative strength of a cue can be changed only on learning episodes in which that cue is present.

My colleagues and I investigated whether causal judgments conform to the rational analysis of the generative retrospective contingency rather than to the prediction of standard associative theory (Larkin et al., 1998). We compared the causal ratings for food G following training on this retrospective generative contingency with those for food C, which had been trained according to a retrospective control contingency (Table 1). In the first stage of training for food C, it received the same number of pairings with the allergic reaction as did food G, but in compound meals with food Y, which, unlike food X, received no training in the second stage. Whereas the rational inference is that food G has greater generative causal power than food C, associative theory predicts similar ratings for the two foods. The observed ratings (Fig. 2, top panel) favor an inference-based account: The experience with food X in the absence of the outcome produced a retrospective enhancement of the causal status of generative food G above the level of the control food C.

Not only can generative causes be established retrospectively, but so can preventive causes. Another group of participants received exposure to two compound meals, PX and NY, neither of which was followed by the allergic reaction. The absence of the outcome after each meal could have been due to two reasons. The first possibility is that neither of the foods was a generative cause of the allergic reaction. Alternatively, one of the foods may have been a generative cause and the other a preventive cause, so that in combination no reaction occurred. The second stage of training disambiguated these alternatives: In the retrospective preventive contingency, after training on the compound meal of PX, food X alone was followed by the outcome. In the neutral contingency, after training on the compound meal of NY, food Y was presented in the absence of the reaction (Table 1). At issue was whether the generative X training in the second stage retrospectively endowed food P with preventive status relative to the neutral food N. Although the magnitude of the effect was small, such retrospective revaluation occurred reliably (Fig. 2, top panel).

LEARNING ABOUT ABSENT CUES

Although retrospective revaluation favors the computational theory of causal inference, this conclusion is based on the assumption that there is no learning about absent cues within associative theory. However, Burke and I have argued that participants not only learn associatively about the causal relationships between the food cues and the allergic reaction, but also learn associatively about the composition of the compound meals, by forming within-compound associations between the constituent foods (Dickinson & Burke, 1996). For example, according to this account, a within-compound association was formed between foods G and X during exposure to the compound GX meal. As a consequence, when food X was subsequently presented alone in the second stage of the retrospective generative contingency (Table 1), it should have retrieved, or activated, a represen-

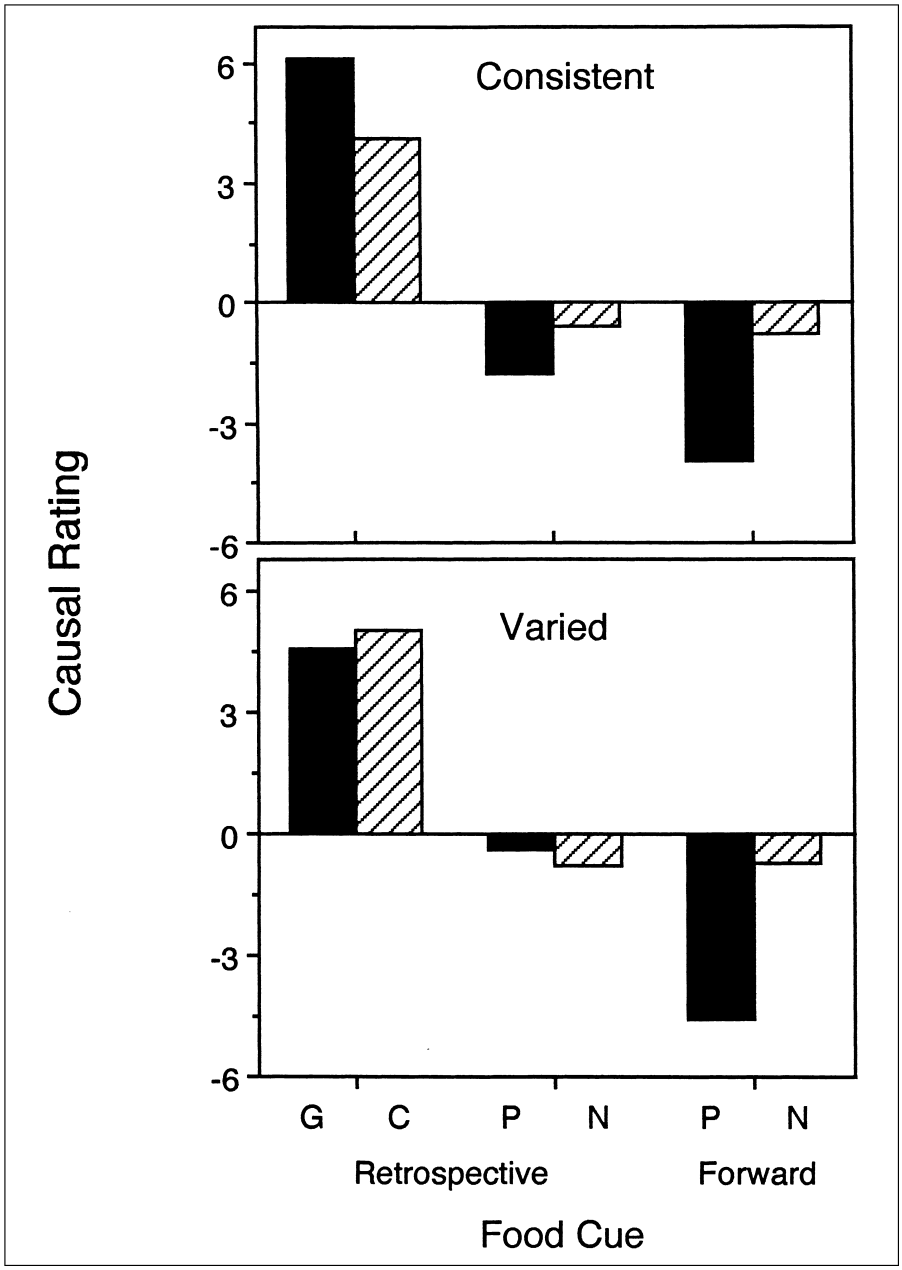

Fig. 2. Mean causal ratings for food cues G and C trained under the generative (retrospective) and control (retrospective) contingencies, respectively; for food cues P and N trained under the preventive (retrospective) and neutral (retrospective) contingencies, respectively (see Table 1); and for food cues P and N trained in the forward condition. In the forward condition, P and N were trained under the same preventive and neutral contingencies, respectively, as in the retrospective condition except for the fact that the two stages of training were reversed. The top panel graphs the ratings when the food cues were consistently paired across the compound meals, and the bottom panel graphs the ratings when these pairings were varied across compound meals. Positive and negative ratings represent generative and preventive causal relationships, respectively.

tation of food G via the within-compound association.

Furthermore, following Van Hamme and Wasserman (1994), we suggested that learning through cue representations takes the form opposite that sustained by cue presentations. Thus, according to this modified associative theory, to increase an association by the training of a cue representation, as opposed to a cue presentation, it is necessary to pair the representation with the absence of an expected outcome (such a pairing would decrease the association for a presented cue). It is exactly this pairing that was achieved during the second stage of the retrospective generative contingency. Food X activated a representation of food G, which was then paired with the omission of the expected allergic reaction. This operation should have enhanced the association for food G and hence the rating of its generative causal power. In fact, this is what happened.

The retrospective acquisition of preventive causal status also follows directly from the modified associative theory. The retrospective preventive contingency ensured that when the presentation of food X retrieved the representation of food P in the second stage, this representation was paired with the occurrence of a surprising allergic reaction. As surprising outcomes decrease associations when paired with cue representations—in contrast to their effect on presented cues—food P acquired a negative association and hence was rated as having preventive causal status.

This associative account of retrospective revaluation depends crucially on the formation of within-compound associations during training with the compound meals. My colleagues and I investigated the role of within-compound associations by assessing whether retrospective revaluation is reduced if the formation of these associations is minimized (Larkin et al., 1998). Although I have referred to the various cues by a single designation letter, in fact a number of different foods served the role of cues G and C, P and N, and X and Y for each participant. The conditions I have already described were the

consistent conditions, in which the same two foods were always presented together in compound meals. Such consistency favors the formation of the within-compound associations necessary for retrospective revaluation. However, we also included varied conditions, in which within-compound associations were minimized because different foods were paired in the first stage of the retrospective contingencies (e.g., each G food and each P food was paired with various X foods, a different X food on every compound meal in the first stage).[2] The lower panel of Figure 2 shows that retrospective revaluation was abolished in both the generative and the preventive contingencies when within-compound associations were reduced in the varied condition.

ASSOCIATION VERSUS COMPUTATION

The development of modified versions of associative theory means that retrospective revaluation is no longer the empirical touchstone for distinguishing between associative and computational theories. Moreover, although the results for the varied and consistent conditions are consistent with the associative account of retrospective revaluation, they are also consistent with the computational account. According to the latter approach, the varied condition, compared with the consistent condition, required the computational system to retrieve information about many more compound meals at the time of judgment, so it is not surprising that retrospective revaluation was attenuated. However, this manipulation can be used to drive an empirical wedge between the two theories by contrasting the effect of varying the pairings of the foods in compound meals during retrospective preventive training and during equivalent forward training. The forward preventive and neutral contingencies were produced by reversing the two stages of training so the participants were trained with food X and food Y alone in the first stage and with the compound meals of foods P and X and foods N and Y in the second stage. Therefore, the forward preventive contingency was identical to the standard preventive contingency displayed in Table 1.

This retrospective-forward distinction is important because the computational theories argue that exactly the same processes mediate causal judgments in the two conditions. Therefore, if the varied condition imposes a larger memory or processing load than the consistent one in the retrospective condition, it should also do so under forward training. By contrast, in the modified associative theory, within-compound associations play a role only in retrospective training. In forward training, all that matters is whether a food cue is paired with an unexpected outcome to produce generative learning or with the surprising omission of an outcome to produce preventive learning. Thus, the computational theories predict that the varied training should interfere with causal learning in both the retrospective and the forward conditions, whereas modified associative accounts predict that the interference is restricted to the retrospective condition. To test these divergent predictions, we compared the forward and retrospective preventive contingencies (Table 1). In contrast to the results for the retrospective contingencies, in which the varied condition failed to establish food P as a preventive cause, the results for forward training showed that food P was rated an equally strong preventive cause in the consistent and varied conditions[3] (Fig. 2; Larkin et al., 1998).

FUTURE ISSUES

Although our analysis of causal learning favors the modified associative theory over the computation account, this conclusion is far from the end of the matter. There are different theories of associative learning. Certain accounts claim that the ability of an outcome to enter into association with a cue is directly determined by whether or not the outcome is surprising, whereas according to other theories, the predictability of the outcome affects learning by controlling attention to the cues. Analyzing the role of these processes in causal learning is a goal of future research.

An equally important issue concerns the nature of the information represented by an association. I have assumed that a cue-outcome association represents a causal relationship, whereas within-compound associations represent simply the conjoint occurrence of the elements of the compound, such as the foods in a meal. However, scenarios in which the relationship between the elements of a compound is itself causal should produce a very different pattern of cue interactions. For example, by analogy to higher-order conditioning, training that pairs cue X with the outcome should enhance rather than block the causal status of cue B (see Table 1) if the relationship between cues B and X is itself causal rather than simply contiguous. Therefore, a developed theory must specify the conditions for the acquisition of associations with differing representational content and how this content is marked.

Whatever these unresolved issues, contemporary research on causal learning has endorsed the

conclusion of the great empiricist philosopher of causation, David Hume (1739/1888), who noted that "'tis sufficient to observe, that there is no relation which produces a stronger connexion in the fancy, and which makes one idea more readily recall another, than the relation of cause and effect betwixt their objects" (p. 11).

Recommended Reading

Cheng, P.W. (1997). (See References)
Dickinson, A. (2001). (See References)
Shanks, D.R., Holyoak, K.J., & Medin, D.L. (Eds.). (1996). *The psychology of learning and motivation: Vol. 34. Causal learning*. San Diego: Academic Press.

Acknowledgments—This research was supported by a grant from the Biotechnology and Biological Sciences Research Council.

Notes

1. Address correspondence to Anthony Dickinson, Department of Experimental Psychology, University of Cambridge, Downing St., Cambridge CB2 3EB, United Kingdom.

2. Subsequent tests verified that the trained compounds were more difficult to recognize following varied than following consistent compound training, indicating that the varied training reduced within-compound learning.

3. We found a comparable dissociation of the effects of varied and consistent compound meals in retrospective and forward generative contingencies (Dickinson & Burke, 1996).

References

Aitken, M.R.F., Larkin, M.J.W., & Dickinson, A. (2000). Super-learning of causal judgments. *Quarterly Journal of Experimental Psychology, 53B*, 59–81.

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

Dickinson, A. (2001). Causal learning: An associative analysis. *Quarterly Journal of Experimental Psychology, 54B*, 3–25.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology, 49B*, 60–80.

Hume, D. (1888). *A treatise on human nature* (L.A. Selby-Brigge, Ed.). Oxford, England: Clarendon Press. (Original work published 1739)

Larkin, M.J.W., Aitken, M.R.F., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1331–1352.

Van Hamme, L.J., & Wasserman, E.A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation, 25*, 127–151.