COMMENT AND REPLY

What are association formation models?

ANTHONY DICKINSON

University of Cambridge, Cambridge, England

In his presentation of the propositional account of associative learning, De Houwer (2009) argues that association formation models (AFMs) assume excitatory link representations and automatic learning processes. However, the application of AFMs to human causal and contingency learning has assumed propositional forms of representation, although excitatory link representations are also required to explain certain nonrational consequences of associative learning. Moreover, at least two of the AFMs that have been applied to human associative learning invoke processing with nonautomatic characteristics. In conclusion, the distinction between the propositional account and AFMs of associative learning lies not in the form of representations but in the specific details of the learning processes generating the associative representations.

De Houwer (2009) has done the field of learning a service in articulating the propositional approach to associative learning, an approach that has been developed through a series of empirical articles in the last few years to a point at which it is timely for an integrative exposition. My purpose in this response is not to evaluate the proposition approach, but rather to respond to the second theme of De Houwer's article, which contrasts the propositional approach with what he calls "association formation models" (AFMs). My contention is that his characterization of AFMs misrepresents these models in a number of respects. Within the context of human causal and contingency learning, AFMs have been developed and applied by a number of authors, and clearly I cannot speak to more than my own conception of AFMs. Indeed, my only authority for doing so is that—along with David Shanks and John Evenden—I was the first to apply AFMs to human causal learning (Dickinson, Shanks, & Evenden, 1984). I shall first discuss De Houwer's characterization of the associative representations invoked by AFMs before addressing the question of whether their learning mechanisms are "automatic."

Associative Representations

There is much I agree with in Part I of De Houwer (2009), in which he argues that the term "associative learning" should be applied to learning about relationships between events with no theoretical commitment to the nature of that learning. I myself argued as much in my monograph on

animal learning (Dickinson, 1980), in which I distinguished associations purely empirically, in terms of the relationship between events and the nature of the events involved in these relationships (pp. 12–22). The nature of the information acquired about associations was addressed explicitly in the third chapter under the title of "Associative representations," and it is here that I first take issue with De Houwer.

Referring to Dickinson (1980, p. 85), he argues that AFMs assume that associative representations are "simply unqualified links between representations through which activation can spread" (De Houwer, 2009, p. 3). However, I deployed this excitatory link form of representation specifically to explain the so-called "stimulus substitution" that occurs in some forms of Pavlovian conditioning, in which the animal responds to the conditioned stimulus (CS) as it does to the unconditioned stimulus—for example, why a pigeon appears to attempt to eat a localized visual stimulus signaling food but to drink one signaling water (Jenkins & Moore, 1973). What De Houwer does not refer to is the extensive discussion of the associative representations established by instrumental conditioning (Dickinson, 1980, pp. 109–120), which concludes that the excitatory link representations cannot explain goal-directed instrumental responding. I explicitly argued in this discussion that goal-directed instrumental action, even in animals, is mediated by proposition knowledge, such as "leverpressing causes food," which is then deployed by a practical inference process to control behavior (Dickinson, 1980, p. 115). So this type of associative representation, even when acquired by an AFM, is compatible with evidence that animals can show some forms of causal reasoning (e.g., Blaisdell, Sawa, Leising, & Waldmann, 2006; Clayton & Dickinson, 2006).

So, from the very outset, I committed AFMs to propositional representations of associations. Indeed, it was this assumption that motivated the Dickinson et al. (1984) study, which was intended to investigate whether the acquisition of propositional knowledge in the form of causal judgments conformed to the predictions of AFMs within a blocking paradigm, which at the time was a touchstone for assessing the contribution of AFMs. Once it is acknowledged that AFMs can support propositional representations, the products of these models can be expected to interact with propositional information, such as that supplied by instructions and prior knowledge. Indeed, the very rationale for applying AFMs to human causal learning assumes that associative representations have a form that allows them to interact with the experimental instruction in order to generate causal judgments.

A. Dickinson, ad15@cam.ac.uk

In the 1980 monograph, I concluded that AFMs generate two forms of representation: an excitatory (or inhibitory) link to explain phenomena, such as Pavlovian stimulus substitution and instrumental habits, and a propositional representation to mediate instrumental goal-directed action. By contrast, De Houwer's (2009) propositional approach assumes a unitary form of associative representation that leaves an explanatory gap for what may be called nonrational responses to associative experiences, such as the acquisition of persistent responding on an omission schedule (see Dickinson, 1980, pp. 116–117). The explanatory gap arises from the fact that the form of a representation cannot be determined independently of the process that operates in deploying its content. In the case of an excitatory link representation, the processes are excitation and inhibition, whereas the deployment of propositional representation must surely be through some form of reasoning process. Indeed, De Houwer appears to support this conclusion by claiming "Propositional models postulate that associative learning effects will depend on the truth evaluation of the propositions considered to be relevant to the behavior being measured" (p. 3), and surely this evaluation must be through reasoning.

The importance of this point is worth stressing. Consider the simple case of why a rat leverpresses for food, having acquired the belief that this action causes food. As pointed out above, according to the account given in Dickinson (1980, p. 115; see also Heyes & Dickinson, 1990), the process that operates on this causal representation is one of practical inference. Specifically, the causal belief interacts with a desire for food to generate an intention to leverpress. The important point about this account is that not only is the practical inference causal in producing the action of leverpressing; this process also respects the intentionality of the representations, such as the truth value of the causal belief. If the intention to leverpress is executed and the causal belief is true, then *of necessity* the desire must be fulfilled. However, it may well be that learning about the relationship between leverpressing and food causes the rat's heart rate to rise when the rat is confronted with the opportunity to leverpress. However, this response cannot be explained by the process of practical inference, because there is no necessary relationship between a change in heart rate and the fulfillment of the rat's desire for food.

De Houwer (2009, p. 10) discusses this issue by reference to human electrodermal conditioning and the question of how propositional knowledge can cause a change in skin conductance. Again, unlike an instrumental escape or avoidance response, sweating is not a rational consequence of knowing that the CS predicts an aversive event (except, of course, in the sense that elicited responses may be fitness enhancing and subject to natural selection). De Houwer (p. 10) simply asks us to accept "the idea that propositional knowledge can have automatic effects on cognition, emotion, and behavior" by fiat, without providing any process that could bridge the explanatory gap. The concept of an excitatory link representation, in conjunction with the process of excitation, provides just such an account, and I would argue that De Houwer's assumption

that associative representations can have nonrational, elicited effects is tantamount to an implicit endorsement of excitatory link representations.

Furthermore, it should be noted that the claim that associative learning forms two types of psychological representations does not necessarily require us to assume dual neural representations. Because the form and content of a representation cannot be specified independently of the processes that deploy it, it is possible that excitatory link and propositional representation are encoded by a common neural engram, and that the psychological distinction arises from the nature of the processes that use this engram in the control of behavior and thought. Indeed, this perspective would view the evolution of generic causal cognition in terms not of the evolution of new forms of neural representations of associations, but rather of the phylogenetic emergence and selection of novel processes for deploying these representations.

Finally, I should note that there are ways in which the associative representations supported by AFMs and those envisaged by De Houwer's (2009) approach might well be differentiated. For example, AFMs impose some constraints on the nature of the representation of a causal relationship, in that they assume that the content confounds a number of features of the learning experience in a single variable, associative strength. This path-independence assumption can be illustrated by considering the acquisition functions for causal judgments, such as those reported by Shanks (1987) under different instrumental contingencies. In this study he found that 10-sec experience with a strong contingency yielded the same causal judgment as 60-sec exposure to a weaker contingency. In the absence of discriminative stimuli signaling the different contingencies, AFMs assume that the propositional representations of the causal relationship generated by these two training regimes do not discriminate between their acquisition histories. These histories all coalesce in the single representation of associative strength. It is unclear to me whether De Houwer is proposing that these two training regimes generated discriminable causal representations independently of incidental information, such as that encoded in episodic memories of the training episode.

Mechanisms of Learning

De Houwer's (2009) second core assumption is that associative learning depends upon nonautomatic processes that require cognitive resources and processing time. He argues that none of the predictions that arise from this assumption "originated from traditional association formation models" (p. 12). Although De Houwer cites authority to support this claim (p. 12), it is far from clear that this assumption is generally endorsed by proponents of AFMs. Of the five traditional AFMs (Mackintosh, 1975; Miller & Matzel, 1988; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1981), the two that specify processing mechanisms both have characteristics of nonautomatic processing.

When we originally argued that AFMs could mediate the acquisition of human causal beliefs (Dickinson et al., 1984), we supported this contention with simulations of the Pearce–Hall model (Kaye & Pearce, 1984). Within the

context of causal learning, this model assumes that associative learning is controlled by the attention to, and hence by the associability of, the putative cause; in the original application of the theory to Pavlovian conditioning, Pearce and Hall (1980) explicitly argued that “the loss of associability of a CS is regarded as a transition from controlled to automatic processing” (p. 549). Although one might argue the merits of this claim, the important point is that Pearce and Hall originally assumed that the attentional processing that supports associative learning is nonautomatic; indeed, there is evidence that the attentional processing invoked by this theory, and the acquisition of propositional knowledge, are closely linked. One of the most counterintuitive predictions of the Pearce–Hall theory is that unreliable predictors receive more attentional processing than do reliable predictors (Kaye & Pearce, 1984). Recently, Hogarth, Dickinson, Austin, Brown, and Duka (2008) have validated this prediction by measuring the amount of time a visual signal was fixated in a human predictive learning paradigm. In agreement with the prediction of the Pearce–Hall model, they found that an unreliable predictor attracted more visual orientation than did a reliable one, but—importantly—only in participants who were aware of the predictive relationship.

The other classic AFM that makes processing assumptions is Wagner’s (1981) SOP model, which has been explicitly applied to human causal learning in a modified form (Aitken & Dickinson, 2005). Although Wagner entitled SOP a “model of automatic memory processing,” it is notable that his model fulfils two of De Houwer’s (2009) criteria for nonautomatic processing: (1) the dependence on “cognitive resources,” and (2) “time.” Without going into details, this model assumes that associative learning depends on how *long* event representations reside concurrently within a limited capacity *resource*; indeed, this limitation has provided the rationale for using a concurrent task to interfere with associative learning (Aitken, Larkin, & Dickinson, 2001).

One complaint about the AFM approach concerns the plethora of models that have been brought to bear on associative learning in general. To take but one example: As noted above, my colleagues and I have offered an account of visual attention during predictive learning in terms of the Pearce–Hall model (Hogarth et al., 2008), while appealing to a modified version of SOP (Aitken & Dickinson, 2005) to explain retrospective revaluation in causal paradigms. However, it is now widely, if not universally, accepted that no single AFM successfully accounts for the gamut of associative learning effects, and that what is required is a hybrid model (Le Pelley, 2004). This conclusion is not just a product of a cavalier and opportunistic selection of models to fit the data, but rather a hard-won conclusion from four decades of careful empirical research that has developed procedures for controlling the contribution of one process in order to reveal the operation of another. Moreover, from an evolutionary perspective, such hybrid models can be expected; for example, consider a prediction error signal (Schultz & Dickinson, 2000) such as that instantiated in the phasic dopamine responses (Waelti, Dickinson, & Schultz, 2001). Initially it evolved

to exert a direct control over learning in a way captured computationally by the Rescorla–Wagner rule (Rescorla & Wagner, 1972). Given such a signal, it is then but a relatively small adaptation to use that prediction error also in the control of attentional (Pearce & Hall, 1980) and associability processes (Mackintosh, 1975).

Conclusion

In this response, I have made three main points. First, as a proponent of AFMs, I have at least endorsed a propositional form of associative representation for goal-direct action and human causal learning. Second, the acquisition of nonrational behavior under associative contingencies is problematic for a purely propositional account, in that there is a major explanatory gap between this form of representation and such elicited behavior. This gap can be closed, however, if it is assumed that associative representations not only have propositional content but can also act as excitatory links. Finally, AFMs are neither in principle nor in actuality incompatible with nonautomatic forms of processing.

My own view is that the burgeoning field of associative learning has passed beyond the stage of disputes between generic classes of theory, such as the propositional and AFM approaches, and should now concentrate on determining the specific learning and deployment processes engaged by associative experiences. This being said, the general theoretical exegesis offered by De Houwer (2009) has the virtue of forcing us to be explicit about theoretical assumptions often only implicit in both theory and empirical procedures.

AUTHOR NOTE

I thank Cecilia Heyes for her comments on a draft of this reply. Correspondence concerning this article should be addressed to A. Dickinson, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England (e-mail: ad15@cam.ac.uk).

REFERENCES

- AITKEN, M. R. F., & DICKINSON, A. (2005). Simulations of a modified SOP model applied to retrospective revaluation of human causal learning. *Learning & Behavior*, **33**, 147–159.
- AITKEN, M. R. F., LARKIN, M. J. W., & DICKINSON, A. (2001). Re-examination of the role of within-compound associations in the retrospective revaluation of causal judgements. *Quarterly Journal of Experimental Psychology*, **54B**, 27–51.
- BLAISDELL, A. P., SAWA, K., LEISING, K. J., & WALDMANN, M. R. (2006). Causal reasoning in rats. *Science*, **311**, 1020–1022.
- CLAYTON, N., & DICKINSON, A. (2006). Rational rats. *Nature Neuroscience*, **9**, 472–474.
- DE HOUWER, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, **37**, 1–20.
- DICKINSON, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.
- DICKINSON, A., SHANKS, D. R., & EVENDEN, J. L. (1984). Judgement of act–outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology*, **36A**, 29–50.
- HEYES, C., & DICKINSON, A. (1990). The intentionality of animal action. *Mind & Language*, **5**, 87–104.
- HOGARTH, L., DICKINSON, A., AUSTIN, A., BROWN, C., & DUKA, T. (2008). Attention and expectation in human predictive learning: The role of uncertainty. *Quarterly Journal of Experimental Psychology*, **61**, 1658–1668.

- JENKINS, H. M., & MOORE, B. R. (1973). The form of the auto-shaped response with food or water reinforcers. *Journal of the Experimental Analysis of Behavior*, **20**, 163-181.
- KAYE, H., & PEARCE, J. M. (1984). The strength of the orienting response during Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 90-109.
- LE PELLEY, M. E. (2004). The role of associative history in models of associative learning: A selective review and hybrid model. *Quarterly Journal of Experimental Psychology*, **57B**, 193-243.
- MACKINTOSH, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, **82**, 276-298.
- MILLER, R. R., & MATZEL, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51-92). San Diego: Academic Press.
- PEARCE, J. M., & HALL, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, **87**, 532-552.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- SCHULTZ, W., & DICKINSON, A. (2000). Neural coding of prediction errors. *Annual Review of Neuroscience*, **23**, 473-500.
- SHANKS, D. R. (1987). Acquisition functions in contingency judgment. *Learning & Motivation*, **18**, 147-166.
- WAELTI, P., DICKINSON, A., & SCHULTZ, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, **412**, 43-48.
- WAGNER, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

(Manuscript received July 28, 2008;
accepted for publication September 8, 2008.)