ORIGINAL ARTICLE

Associative theories of goal-directed behaviour: a case for animal–human translational models

Sanne de Wit · Anthony Dickinson

Abstract Associative accounts of goal-directed action, developed in the fields of human ideomotor action and that of animal learning, can capture cognitive belief-desire psychology of human decision-making. Whereas outcome-response accounts can account for the fact that the thought of a goal can call to mind the action that has previously procured this goal, response-outcome accounts capture decision-making processes that start out with the consideration of possible response alternatives followed only in the second instance by evaluation of their consequences. We argue that while the outcome-response mechanism plays a crucial role in response priming effects, the response-outcome mechanism is particularly important for action selection on the basis of current needs and desires. We therefore develop an integrative account that encapsulates these two routes of action selection within the framework of the associative-cybernetic model. This model has the additional benefit of providing mechanisms for the incentive modulation of goal-directed action and for the development of behavioural autonomy, and therefore provides a promising account of the multi-faceted process of animal as well as human instrumental decision-making.

S. de Wit (✉)
Department of Psychology,
Amsterdam Center for the Study of Adaptive Control in Brain
and Behavior (Acacia), University of Amsterdam,
Amsterdam, The Netherlands
e-mail: s.dewit@uva.nl

A. Dickinson
Department of Experimental Psychology,
University of Cambridge, Cambridge, United Kingdom

Introduction

According to folk psychology, intentional, *goal-directed actions* are performed when an agent desires the goal and believes that the behaviour in question will achieve the goal. Consider the example of going to the local market to buy flowers. This goal-directed action is performed because of the desire to get flowers and the belief that going to the market will achieve this outcome. This intuitive notion of goal-directed action has been captured by cognitive '*belief-desire*' *psychology*, according to which the intention to act is produced by an interaction of a belief about the causal relation between action and outcome with the value of, or current desire for, the outcome. Such an account is to be found in one form or another in action theory in human motivation, expectancy-value theory and the theory of rational action or planned behaviour in social psychology, and belief-desire psychology in the philosophy of mind (Greve, 2001), and indeed Heyes and Dickinson (1990) have argued for a belief-desire cognitive account of animal instrumental action.

Although goal-directed and planned action has been the focus of much high-level computation theorising in human cognitive psychology (Anderson et al., 2004; Daw, Niv, & Dayan, 2005; Newell, 1990) the predominant cognitive paradigms have traditionally employed instructed stimulus-response tasks (e.g. Stroop task, Simon task, flanker task) to study direct behavioural control. While these tasks have furthered our understanding of decision-making, they certainly do not capture the fact that most of our daily decisions involve choosing actions on the basis of their consequences rather than performing according to instructed stimulus-response mappings. The two research fields that have always recognised the importance of instrumental relationships between our actions and their consequences in behavioural control are those of (1) human

ideomotor action and (2) animal learning. In this review, we will discuss the strengths and weaknesses of accounts of goal-directed action derived from these research fields before developing an integrated associative model of goal-directed and habitual behaviour.

Goal-directed behaviour

A prerequisite for an action to be considered goal-directed is that it is *instrumental*. Instrumental actions are characterised by two properties. First of all, instrumental actions are *learned*. Second, the action has to be controlled by the causal relationship between the action or response (R) and its consequences or outcome (O), rather than being controlled by predictive relationships between environmental stimuli (S) and the outcome. Predictive S-O learning in animals is usually studied in Pavlovian conditioning procedures (Pavlov, 1927) in which a new response is acquired to a stimulus that predicts the occurrence of a biologically significant outcome. There are two important points to make about the distinction between instrumental and Pavlovian learning. First, although Pavlovian responses are usually biological adaptive within the ecological environment in which they evolved, such behaviour can be maladaptive when the animal is faced with new contingencies. Second, it is not always transparent which of these two types of learning underlies the acquisition of a new response.

These two points are succinctly illustrated by an experiment in which Hershberger (1986) fed chicks from a distinctive food bowl. When they were subsequently displaced from the food bowl, the chicks rapidly learned to approach it. Although the obvious account of this behaviour is that the birds learned that the approach response yielded food, Hershberger demonstrated that this was not the case by placing another group of chicks in a "looking-glass world". In this world the food bowl receded from the chicks twice as fast as they approached it, so that they could never feed by approaching the food bowl. If chicks' locomotion in this situation were instrumental, they should have had no problem in extinguishing any approach tendency and learning the withdrawal response. However, this they never did, suggesting that their locomotion was not under the control of the instrumental response-outcome (R-O) relationship. Rather their approach behaviour appears to have been controlled by the predictive stimulus-outcome (S-O) relationship so that the sight of the bowl elicited approach as a Pavlovian conditioned response. This is not to say that animals are incapable of true instrumental action that is under the control of the causal R-O relationship. There is good evidence, for example, that the action of pressing a lever by rats is directly controlled by its consequences (Dickinson, Campos, Varga, & Balleine, 1996). In contrast, the instrumental status of

human responses is rarely examined experimentally as they are usually arbitrary and instructed (Pithers, 1985).

Once we have established that behaviour is instrumental, it still needs to meet two additional criteria to be considered intentional in the sense of goal-directed, as opposed to automatic or habitual: the belief criterion and the desire criterion (Heyes & Dickinson, 1990).

Belief criterion

The purpose of the belief criterion is to capture the idea that if a behaviour is goal-directed it must be based upon knowledge of the relationship between the action and its consequences, because in the absence of such knowledge it is not clear how the action could be 'directed' to its consequences as a 'goal'. In other words, goal-directed actions are controlled by a belief that the action will cause the goal. In associative terms, behaviour is mediated by a representation of the R-O relationship. The belief criterion excludes another class of instrumental responses that are not mediated by anticipation of the goal, but instead are directly primed by contextual stimuli through S→R associations. It is usually assumed that these *habitual* responses are acquired according to Thorndike's *Law of Effect* (1911), which states that the experience of reward following a response leads to the strengthening of an association between the context and the response, so that on future occasions, the context can directly prime the response through the S→R association. To be qualified as a goal-directed agent, it is therefore not sufficient that one is sensitive to the effect of one's actions on the world, but additionally a representation of the causal R-O relationship needs to mediate performance. Finally, we argue that in addition it is crucial that a motivational criterion is met.

Desire criterion

According to the desire criterion, goal-directed actions are controlled by the affective or motivational value of the outcome at the time that the action is performed, or in other words, the action is only executed when the expected consequences are currently desired and therefore have the status of a goal. In order to establish whether actions meet the belief criterion as well as the desire criterion, researchers in the field of animal learning have developed the *outcome revaluation* paradigm. The basic idea behind this paradigm is that a change in the incentive value of an outcome should only directly influence the propensity to perform a previously acquired action if (1) the agent possesses knowledge of the instrumental contingency between the action and the outcome (belief criterion) and (2) the agent can evaluate the outcome in light of its current needs and desires and use this evaluation to decide whether to perform the action or not (desire criterion).

In a typical outcome revaluation procedure, rats are first trained to perform an action to gain a certain food outcome. Following this instrumental training, the value of this outcome is reduced in the absence of the opportunity to perform the instrumental action, either by establishing a specific food aversion by pairing the consumption of food with the induction of nausea (Garcia, Kimeldorf, & Koelling, 1955) or by unlimited feeding the animal to induce specific satiety for this food (Balleine & Dickinson, 1998a, 1998b). Subsequently animals are given the opportunity to perform the instrumental action again. If the behaviour is directly controlled by the current goal value of the outcome, the animals should immediately respond less for the now devalued outcome. Note that in order to ensure that performance reflects employment of R-O knowledge acquired during the instrumental training, it is paramount to conduct this test in extinction and therefore in the absence of outcome deliveries. If the devalued outcomes were to be delivered contingent upon responding in the test phase, then any decline in responding could well be due to a gradual weakening or inhibition of direct S→R associations between the context and the instrumental response through a punishment mechanism of the type envisaged by Thorndike in his Law of Effect (1911). While the Law of Effect can account for the gradual decline in performance of an action that is no longer followed by a desirable outcome, it would not predict immediate sensitivity to outcome devaluation in the absence of further experience with the instrumental contingency.

To give an example of an outcome revaluation study, Adams and Dickinson (1981) trained rats to lever press for one type of food pellet while delivering a second type independently of responding. Following devaluation of either the response-contingent or non-contingent outcome through aversion conditioning, lever pressing was tested in extinction. In this test, the animals pressed less following devaluation of the response-contingent rather than the non-contingent outcome, thereby demonstrating that the rats had acquired knowledge of the R-O contingency during instrumental training, and that their instrumental behaviour was sensitive to the current desirability of the outcome. This instrumental outcome devaluation effect has been extensively replicated using both aversion conditioning and specific satiety to devalue the outcome (e.g. Colwill & Rescorla, 1985).

Cognitive neuroscientists have so far made only limited use of the outcome revaluation paradigm to assess whether human actions meet the desire criterion for goal-directed action, but it seems clear that in everyday life we often act in a goal-directed manner when the incentive value of outcomes of our actions change. For instance, while an ice cream from the ice cream van may be appealing most of the time, it is certainly less attractive if one has just enjoyed a bountiful ice cream desert. In cases like this, we may directly decide against pursuing the additional icecream, rather than having to rely on the gradual weakening of S→R associations. Although the ability to behave in a goal-directed manner appears to be crucial for human decision-making, to the best of our knowledge, only four studies so far have employed the outcome revaluation paradigm to investigate this ability directly in humans. Valentin, Dickinson, and O'Doherty (2007) showed that human volunteers reduced responding that had been trained with a liquid reward (tomato juice, chocolate milk or orange juice) when they had just consumed this food to satiety. In another study, the goal-directed status of adult instrumental performance was demonstrated through "instructed devaluation" (de Wit, Niry, Wariyar, Aitken, & Dickinson, 2007). In the latter study, during the preceding instrumental training, participants pressed keys to gain fruit icons and points on a computer screen. For example, right key presses were rewarded with a picture of an apple and points, whereas left key presses were rewarded with a picture of a pear and points. During the subsequent extinction test phase subjects were asked to gain certain fruit pictures but not others. For example, participants were informed that the apple would no longer earn them any points while the pear was still valuable. Participants were able to select the action trained with the apple outcome more than that trained with the pear outcome, thereby demonstrating that they had acquired knowledge about the R-O contingencies during training and could modulate their performance in accordance with the instructed values of the outcomes (de Wit et al., 2007).

Of course, the ability of adults to select an action on the basis of the current incentive value of its expected outcome hardly needs experimental demonstration (even if the process by which they do so remains undetermined). What is less clear, however, is the developmental trajectory of the capacity for goal-directed action. Although it is well known that infants are capable of instrumental learning (Rovee-Collier, 1983), what is less clear is whether the resulting actions are goal-directed rather than simply habitual. To address this issue, Klossek, Russell, and Dickinson (2008) taught children to perform two different actions in order to view videoclips from different cartoons series (Pingu vs. teletubbies), before assessing whether their performance was goal-directed by devaluing one of the cartoons through repetitive presentation. When children older than 2 years were once again allowed to choose between the two actions in an extinction test, they consistently performed the action trained with the still-valued cartoon more than that trained with the now-devalued cartoon, a result that was replicated in another scenario by Kenward, Folke, Holmberg, Johansson, and Gredeback (2009). Therefore, the instrumental behaviour of humans can from an early age meet the belief criterion as well as the desire criterion of goal-directed action.

In summary, the conception of goal-directed action presented here includes three main ideas: goal-directed actions are instrumental, they are controlled by a representation of the R-O relationship, and they are sensitive to the current incentive value of an outcome. We thereby exclude many of the behaviours that are typically referred to as goal-directed in the field of cognitive psychology. First of all, this definition excludes behaviours that are completely controlled by Pavlovian contingencies. Students of goal-directed action under our definition are therefore well-advised to avoid *target-directed* measures as the dependent variable such as food approach, grasping of targets or eye-movement towards reward, as Pavlovian influences are likely to be a major determinant of these behaviours.[1] Secondly, goal-directed actions need to meet the belief criterion and the desire criterion. This definition therefore excludes behaviours that are under direct control by external stimuli, and S-R tasks are therefore not suitable for the study of goal-directed action. For example, the response of saying 'green' when confronted by the word "green" printed in blue in the Stroop Task, or pressing a button in response to an arrow on the screen in a flanker task, appear to be predominantly controlled by instructed S→R associations.

Outcome-response (O-R) theory

Folk psychology suggests that the thought of a goal precedes the decision to perform the action that leads to it. For example, the thought of an ice cream may inspire a trip to the ice cream van. The idea that action planning starts with imagining the consequences of the action has been captured in the *ideomotor account* of human intentional action, which was developed mainly in the nineteenth century (Carpenter, 1852; Herbart, 1825; James, 1890; Lotze, 1852; Stock & Stock, 2004) and is still endorsed by many researchers in the field of human cognition (Hommel, 2003; Hommel, Musseler, Aschersleben, & Prinz, 2001). According to James (1890), experiencing a contingency between an action and its outcome (or effect) produces an association between the motor programme for the action and a representation of the outcome, so that on future occasions activating the outcome representation directly excites the action programme. Although ideomotorists have mainly studied associations between movements and their immediate *sensory* consequences, similar associative accounts have been developed in the field of animal learning to account for goal-directed action (Asratyan, 1974; Gormezano & Tait, 1976; Pavlov, 1932). We will refer to the ideomotor

account as well as related associative accounts as *outcome-response* (*O-R*) *theory*, because the basic premise is that goal-directed actions are mediated by O→R associations. In the following section, we will focus on the experimental paradigms that have provided support for the prediction of O-R theory that actions can be triggered by experiencing the consequences of these actions directly or even by mere anticipation of the consequences.

O-R theory and the belief criterion

If instrumental learning establishes O→R associations, which can then be used to guide action selection, presenting the outcome should immediately prime the associated action. Addictive drugs appear to be particularly potent (O→R) primers of drug craving and drug seeking. For example, Chutuape, Mitchell, and de Wit (1994) found that consumption of an alcoholic drink not only enhanced the craving for alcohol but also the propensity to seek the drug, and response priming by such outcomes has received extensive empirical investigation in animals (see for a review, Stewart & de Wit, 1987). However, purely sensory outcomes are also capable of priming their associated responses. In the R-O pre-training stage of Meck's experiment (1985), rats were presented with two levers. While a common food reward could be obtained after pressing both levers, each lever was also paired with a specific sensory post-reinforcement cue, either a noise or a light signal. After substantial training on this schedule, the rats were transferred to a procedure in which the noise and the light preceded the opportunity to lever press. For half of the rats (congruent group), each stimulus signalled that the response that had previously earned this reward during the R-O pre-training would be rewarded. For instance, when a tone was presented, the animals had to press the lever previously associated with the tone outcome, whereas when a light was presented, they had to press the lever associated with the light outcome. These contingencies were reversed for the remaining animals (incongruent group) in such a way that the animals had to learn to make the opposite response. Here, a tone signalled that the lever that was previously associated with the light outcome would lead to a reward and vice versa. If R-O training had established O→R connections between the outcome cues (i.e. the light and the tone) and their respective responses, the congruent group should acquire the discrimination more rapidly than the incongruent group. This prediction of O-R theory was confirmed.

More recently, similar results were obtained in humans. For example, Elsner and Hommel (2001) demonstrated outcome-mediated response priming in humans. In the R-O pre-training stage, participants were instructed to press either a right or a left key as quickly as possible on appearance

---

[1] For models of target-directed movements see, for example, Wolpert and Ghahramani (2000).

of a white rectangle. Each key press triggered either a high or a low tone, for example, right key presses (R_R) were followed by a high tone (T_h) and left key presses (R_L) by a low tone (T_l). According to O-R theory, this pre-training should have established $T_h{\rightarrow}R_R$ and $T_l{\rightarrow}R_L$ associations. In the test phase, trials started with the presentation of one of the tones and the participants were instructed to press either the left or the right key as quickly and as spontaneously as possible. Right key presses following the high tone and left key presses following the low tone were counted as congruent choices, whereas the opposite choices were counted as incongruent. The prediction of O-R theory that the participants should make more congruent than incongruent choices was again born out by the results.

The response priming effect has not only been shown with tones as outcomes (Elsner & Hommel, 2001, 2004) but also with a variety of other sensory outcomes (Ansorge, 2002; Hommel, 1993, 1996; Ziessler, 1998; Ziessler & Nattkemper, 2001, 2002). Moreover, Beckers, De Houwer, and Eelen (2002) showed that the emotional valence of outcomes can prime responses through O→R associations. In their study, responses associated with shock outcomes were executed faster following negative target words than neutral responses, and relatively slowly following positive target words. Finally, Elsner and Hommel (2004) demonstrated that the two major parameters controlling animal instrumental learning, R-O contiguity and R-O contingency, also influence the priming effect as would be predicted by associative theory.

In summary, there is a wealth of evidence for outcome-mediated response priming in adult humans. Moreover, O-R learning has been demonstrated early in infancy. If babies are given the opportunity to learn that moving their legs activates a mobile, then later presentations of the moving mobile will cause them to move their legs (Fagen & Rovee, 1976; Fagen, Rovee, & Kaplan, 1976), suggesting that the ability to form O→R associations develops early in life.

However, although response priming by outcomes provides evidence for the formation of O→R associations, this effect by itself does not support the contention of O-R theory that the mere thought of a goal can prime actions through these associations. In daily life most goal-directed actions do of course not result from direct perception of the goal. For example, the sound of the bell of the ice cream van on a sunny afternoon makes us think of an ice cream which in turn makes us think of the action that has yielded ice cream in the past, namely to approach the van and buy an ice cream. In associative terms, the bell activates the ice cream representation through a (Pavlovian) S→O association, which in turn activates the response representation of buying the ice cream through an (instrumental) O→R association.

Of course, in daily life, S-O and O-R learning as well as S-R learning occur concurrently in complex, dynamic interactions so that the contribution of each learning process to performance cannot be isolated. However, many years ago animal learning researchers have developed the so-called *Pavlovian-to-Instrumental transfer* paradigm (*PIT*) to demonstrate the separate contributions of S-O and O-R learning to the control of instrumental performance (Estes, 1943, 1948). With this task it has been shown that a purely Pavlovian stimulus can prime performance of an instrumental response that was separately paired with the same outcome as the Pavlovian stimulus (Baxter & Zamble, 1982; Blundell, Hall, & Killcross, 2001; Colwill & Motzkin, 1994; Colwill & Rescorla, 1988; Corbit, Janak, & Balleine, 2007; Holland, 2004; Kruse, Overmier, Konz, & Rokke, 1983; Rescorla, 1994). As the Pavlovian stimulus and instrumental response were never trained together, this effect has to be mediated by an intervening representation of the common outcome rather than being mediated by direct S→R associations. The PIT procedure is illustrated by an experiment of Colwill and Motzkin (1994) in which rats first received Pavlovian S-O training during which, for example, a light signalled the delivery of food pellets and a noise predicted access to sucrose solution. In a separate instrumental training phase, the same rats were trained to lever press for food pellets and chain pull for sucrose solution in the absence of the light and the noise. Finally, in an extinction test the rats were given access to both response manipulanda. This test yielded outcome-specific PIT in that the stimuli caused the rats to make the response associated with the signalled outcome. During the noise the rats chain-pulled more than they pressed the lever and vice versa. These results suggest that presentation of the noise (or the light) triggered the representation of the food pellets (or the sucrose solution) that in turn led the rats to press the corresponding lever, thereby demonstrating the role of an S→O→R associative chain in the control of responding.

It seems intuitively appealing that humans, like rats, select actions via S→O→R associative chains, and in fact PIT effects have been argued to play a crucial role in drug seeking behaviour (Ludwig, Wikler, Stark, & Lexington, 1974). Moreover, the logic behind marketing is that advertisements remind consumers of the product in question, which in turn should prime the action of purchasing it. Over the last few years, the PIT and associated paradigms have been increasingly used to demonstrate the effect of this route to action selection in humans (Bray, Rangel, Shimojo, Balleine, & O'Doherty, 2008; Hogarth, Dickinson, Wright, Kouvaraki, & Duka, 2007; Talmi, Seymour, Dayan, & Dolan, 2008). For example, Hogarth et al. (2007) showed that in the presence of a stimulus associated with cigarettes addicted smokers were more likely to perform actions trained with cigarettes than those trained with money, with

the reversed pattern of performance evident when the stimulus associated with money was presented. Similarly, Bray et al. (2008) reported that Pavlovian cues for natural rewards (chocolate milk, a soft drink and orange juice) would preferentially enhance performance of responses that previously yielded these rewards during a separate instrumental training phase.

In summary, a prerequisite for goal-directed action is that the agent possesses knowledge of the instrumental R-O contingency. Evidence that animal and human behaviour is sensitive to both direct response priming by outcomes and indirect priming by cues for outcomes suggests that O-R theory provides a viable account of instrumental learning. In the next section, we will address the question whether this account can also capture the desire criterion of goal-directed action, namely, that actions are sensitive to the current motivational value of their outcomes.

O-R theory and the desire criterion

Ideomotor theory and associative O-R theory provide a viable account of the acquisition and deployment of instrumental knowledge in a way that fulfils the belief criterion of goal-directed action. The question remains, however, whether O-R accounts can explain behaviour that is sensitive to the current motivational value of an outcome—in accordance with the desire criterion—and can thus be considered a sufficient account of goal-directed actions in the strict sense. In fact, as noted earlier, most ideomotor research has focused on response priming by sensory events that did not serve as goals for the participants. Indeed, some authors have even interpreted ideomotor theory as strictly referring to associations between motor movements of the body and the perceivable motor effects (Kunde, Koch, & Hoffmann, 2004; but see Hommel, 2003). Accordingly, to date O-R theory does not provide a mechanism through which the motivational value of the outcome influences action planning.

The O-R approach could, however, easily be modified to account for such a sensitivity to affective value by incorporating the additional assumption that the extent to which the outcome mediates response activation through the O→R association depends on the current value of the outcome. Or in other words, 'a fiat that the action's consequences become actual intervenes' when actions are selected via the ideomotor mechanism (James, 1890). The sensitivity of the outcome representation to input excitation, and hence its ability to activate the associated response, could be modulated by the relevance of the outcome to the agent's current motivational state. For example, the ability of choices on a menu to excite strong or vivid thoughts of the actual meal, may well depend upon how hungry one is, and this may determine to what extent the response of ordering an option

is activated through the S→O→R associative chain. From a purely theoretical perspective, therefore, it is possible for O-R theory to provide an account of the goal-directed nature of instrumental actions. However, there is an empirical evidence from animal research that speaks against such a possibility.

A potential problem for O-R accounts of goal-directed action is that outcome-specific PIT appears to be insensitive to the current value of the outcome. Earlier we argued that PIT provides evidence for an O-R account because the separate establishment of S→O and O→R associations enables the Pavlovian stimulus to activate the response associated with the common outcome even though the stimulus and response have never been trained together. Therefore, instrumental responses can be triggered through an S→O→R associative chain. The finding that speaks against an O-R theory as a complete account of goal-directed action is that, at least in rats, the magnitude of PIT is unaffected by outcome devaluation. For example, following separate S-O and R-O training phases, Rescorla (1994) conditioned an aversion to the food outcome before testing the ability of the stimulus to enhance the corresponding response. Importantly, he found that the ability of a stimulus to facilitate responding was unaffected by devaluation of the associated food (see also Holland, 2004). In other words, PIT occurs even when the anticipated outcome is not currently a goal for the animal, and therefore, although O→R associations allow outcomes to prime associated responses, a different mechanism may mediate the pursuit of goals.

Further evidence that O-R theory may not be a sufficient account of goal-directed behaviour comes from neural dissociations between PIT and outcome devaluation effects in rats. Lesions of the nucleus accumbens core abolish the immediate sensitivity of instrumental performance to devaluation of the outcome, but do not affect outcome-specific PIT (Corbit, Muir, & Balleine, 2001). Conversely, lesions of the nucleus accumbens shell abolish this form of PIT while not influencing performance in an outcome devaluation test (Corbit et al., 2001; de Borchgrave, Rawlins, Dickinson, & Balleine, 2002). Furthermore, related research has shown that extensive amounts of training abolish the goal-directed nature of instrumental performance but leave outcome-specific PIT intact (Holland, 2004). In summary, several rodent PIT studies provide evidence that response priming via O→R associations and outcome devaluation are mediated by different mechanisms, suggesting that O-R theory may not provide a complete account of goal-directed behaviour. Note, however, that these studies do not demonstrate that O→R associations *cannot* mediate goal-directed behaviour. It certainly remains possible that the lack of sensitivity of response priming via O→R associations to the current incentive value of an outcome and to

extensive training merely arise as artefacts of the PIT procedure. In any case, it appears wise to be cautious in accepting an O-R account of goal-directed behaviour. In the following section we will, therefore, discuss an alternative candidate theory of goal-directed action.

Response-outcome (R-O) theory

Although O-R theorists focus on the role of outcome expectancies as a cue for action control through O→R associations, they commonly assume that the association between outcome and response is bidirectional (Elsner & Hommel, 2004; Pavlov, 1932; Asratyan, 1974). An alternative account of goal-directed action emphasises the importance of the (forward) R→O association, rather than the O→R association in which the direction of the association is 'backward' with respect to the instrumental relationship between response and outcome. Like the O-R theory, this alternative R-O theory also has an intuitive appeal. R-O theory assumes that the decision process starts with the thought of the alternative courses of action that are available, one at a time. The thought of each response then retrieves the thought of its outcome through the R→O association. If the outcome is then evaluated as a goal, this evaluation is fed back to the response mechanism to enhance the likelihood that this response will be performed. As an example, assume that you are visiting Amsterdam and that you just had lunch on the Dam Square. As you come out of the restaurant, you consider turning to the right, which you know will bring you back to the central train station, or turning to the left, which you know will bring you to the shops in Kalverstraat. According to the R-O account, the thought of turning right activates a representation of the station, which because it is not currently a goal, supplies no evaluative feedback to the response of turning right. By contrast the thought of turning left activates a representation of shopping in Kalverstraat, a thought that is evaluated as a current goal with the consequence that this evaluation is fed back to augment the propensity to turn left.

To the best of our knowledge, such an R-O theory was first proposed by Thorndike (1931). He characterised it as a "representative or ideational theory" and on those grounds dismissed it as an account of instrumental behaviour, at least for animals. Subsequently, a number of theorists espoused what Estes (1969) called "contiguity-cybernetic" accounts and indeed Sutton and Barto (1981) simulated such a model within the context of the left versus right turn choice, as in the Amsterdam scenario. Unhappy with an implied commitment to a purely contiguity-based process for R-O learning, Dickinson and Balleine (1993, 1994) rechristened this class of theories as *associative-cybernetic* (*AC*) *accounts*.

One of the problems faced by AC theory is how the thoughts of candidate responses are brought to mind in the first place. In the case of the Amsterdam shopping trip, at least in its idealised form, there are only two response alternatives and so one can assume that random sampling of the two alternatives would rapidly lead one to consider turning left for shopping. In this particular scenario, therefore, the R-O mechanism for goal-directed action allows for efficient decision-making. In the case of the ice cream vendor, however, it seems implausible that when we hear the jingle from our living room, we need to consider all response alternatives that are available in that moment until the thought of going to the van reminds us of the currently desired outcome of ice cream. To explain the apparent ease of decision-making in such a context R-O theory needs to take into consideration the effect of the prior experience of the S: R-O relationship. To solve the problem of response sampling, Dickinson and Balleine suggested that the R-O process could be interfaced with an S-R mechanism (Dickinson, 1994; Dickinson & Balleine, 1993).

As we have noted earlier, the "Law of Effect" of instrumental behaviour (1911) assumes that delivering a reward contingent upon a response strengthens or reinforces an S→R association, in our example, between the jingle and the response of going in the direction of the van. The critical feature of this stimulus-response/reinforcement process is that it allows stimuli to directly bring to mind the thought of a response that in the past has been rewarded in the presence of that particular stimulus. There is a good evidence for the reality of this habit process even at the lowest levels in the central nervous system. For example, Wolpaw and colleagues (1997) discovered that the gain of the stretch reflex in both rats and monkeys can be altered by instrumental reinforcement. As this reflex is mediated by a single synapse between the afferent and efferent limb of the reflex arc in the spinal chord, this adaptation represents a prototypical form of S-R learning.

Because the nature of the outcome is not encoded in the S→R associative structure, this process by itself cannot mediate goal-directed action, but rather is a model for the acquisition of habits. S-R habit formation can, however, be integrated with R-O learning to address the problem of response sampling. In our ice cream van scenario, the initial trips to the van to purchase an ice cream form a jingle→van approach (S→R) association, albeit weak initially, reinforced by the reward of consuming the ice cream. Although this weak jingle→approach association may not be strong enough to support reliable behaviour on its own, it at least enables future presentations of the jingle to activate the thought of going to the van more strongly than those of alternative responses. This thought then calls to mind the ice cream via the R→O association, which in turn supplies a positive feedback upon the thought of approaching the
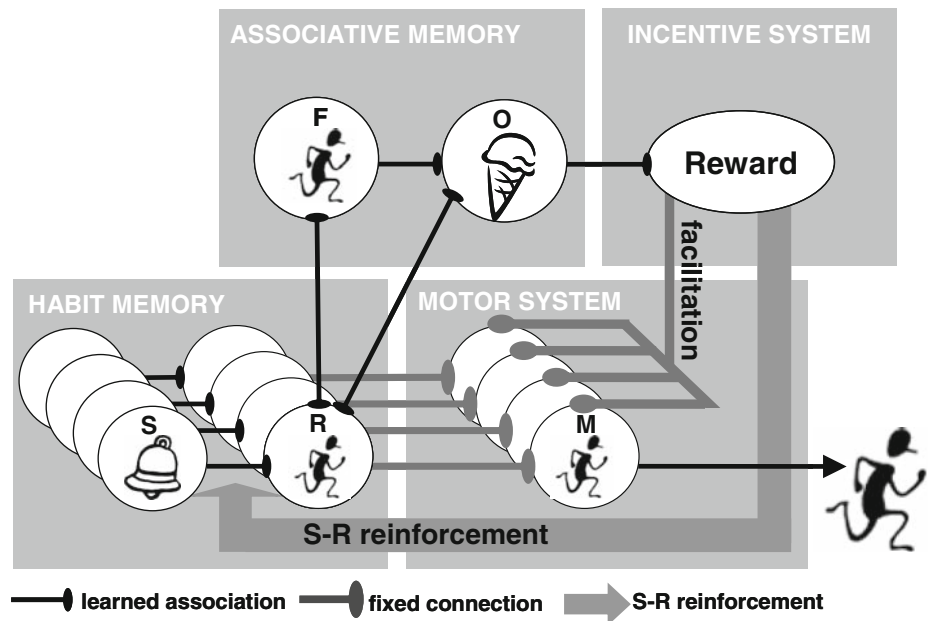
van so that it becomes activated strong enough to trigger the actual behaviour. Therefore, the AC theory assumes that goal-directed actions are selected through an S→R→O associative chain in contrast to the S→O→R chain of O-R theories.

The associative-cybernetic model

In order to illustrate the operation of the AC theory, Dickinson and Balleine developed a model (Dickinson, 1994, Dickinson & Balleine, 1993), which is illustrated in Fig. 1 by applying it to the scenario of purchasing an ice cream upon hearing the vendor's jingle. Behaviour originates in the *habit memory*, which contains (S) sensory units responsive to environmental stimuli that are potentially connected to response (R) units that encode a relatively high-level specification of the corresponding behaviour. The output of the habit memory takes place via the *motor system*, which contain units (M) corresponding to the motor programmes that actually drive the behaviour. The connections between the stimulus and response units in the habit memory are plastic and subject to strengthening through the reinforcement process envisaged by Thorndike (1911). This process is implemented in the model by a reward unit in the *incentive system*, which delivers a reinforcement signal to the habit memory when a rewarding outcome is received so that the S→R connection that is active contiguously with that signal is strengthened. Consequently, in the ice cream scenario, the first ice cream we enjoy will strengthen a tendency to approach the van when we next hear its jingle.

Within the AC model, the reinforcement learning in this direct S-R *feed-forward* pathway through the habit memory

and motor system is a slow process so that early in acquisition the relevant S→R connection is not sufficiently strong for the output of the habit memory to drive the corresponding motor unit reliably. And, of course, the feed-forward pathway on its own cannot give rise to goal-directed behaviour. To address this issue, the model contains a fourth component, an *associative memory* that contains units that encode sensory/perceptual events that are critical for goal-directed action. The first is a unit (F) representing the sensory feedback that arises from performing the target response, in this case van approach, whereas the second unit (O) represents the outcome of the response, the ice cream. Activation of F and O units corresponds to the thoughts of van approach and of the ice cream, respectively, in the folk psychology account.

The *feedback loop* is then completed by the acquisition of three associations. The first is a bidirectional R⇔F association between the response unit in the habit memory and the feedback unit in the associative memory brought about by the contingency between the firing of the response unit and activation of the feedback unit by immediate sensory consequences of performing the target response, a contingency that is experienced every time the response is performed. Within our scenario, activation of the response unit for van approach in the habit memory will activate the thought of the response in the associative memory. The second is an F→O association formed in the associative memory by experience with the instrumental contingency between the action and outcome. This F→O association represents instrumental knowledge as stipulated in the belief criterion for goal-directed action and in our scenario means that the thought of van approach will activate the thought of ice cream.

Recall, however, that in order for behaviour to be goal-directed in the strict sense, it has to also meet the desire criterion. In other words, a goal-directed action is not only guided by instrumental knowledge but it is also sensitive to the current incentive value of its outcome. Within the AC model the value of an outcome is learnt by experiencing directly the hedonic reactions to it. There is now extensive evidence for this process of incentive learning in rats (Dickinson & Balleine, 1994, 2008). Incentive learning is captured in the AC model by assuming that experiencing the delight of ice cream leads to the formation of a connection between its sensory (O) representation in the associative memory and the reward unit in the incentive system. Of course, devaluation of the outcome, for example by eating a lot of ice cream (satiety) or feeling sick after eating ice cream (aversion conditioning), would render this association ineffective so that the thought of the outcome would no longer excite the reward unit. The feedback loop is completed by assuming that activation of the reward unit, as well as supplying the reinforcing signal to the habit memory, also exerts a general facilitatory influence on the motor system. The AC model therefore argues that a habit process, mediated by the feed-forward pathway, and the goal-directed process through the feedback loop work in concert to control instrumental behaviour.

We will now return to the ice cream example to illustrate the integrated operation of the AC model. After some limited experience with the ice vendor, the hearing of the jingle will activate an approach tendency through the S→R association in the habit memory, which in turn will tend to excite the M unit for this response in the motor system. Concurrently, thoughts of the approach response and of ice cream will be activated in the associative memory through the R→F and F→O associations, respectively, which in turn will activate the reward unit in the incentive system, bringing about a general and indiscriminate influence on all units in the motor system. The important feature of the positive feedback is that it is only effective when it coincides with an excitatory input of the habit system to the approach motor units. Consequently, the excitatory influences mediated by the feed-forward and feedback pathways summate on the motor unit for approach thereby enhancing the likelihood of this behaviour.

Having sketched out the AC model, we shall consider a number of features of goal-directed actions and habits and of their interaction that find a ready explanation within the model.

Behavioural autonomy

The '*slips of action*' that punctuate daily life illustrate that our behaviour is not always goal-directed in nature. Folk wisdom suggests that such slips of action occur when well-practised responses intrude to compromise our goal-directed behaviour. For example, it is a common-place experience to find oneself arriving at the door of the old office although one's original intention was to get to the new one. Adams (1982) was the first to show that in rats extensive training of the instrumental response of lever pressing rendered it impervious to devaluation of the food outcome, a finding that has now been replicated in a number of rodent studies (Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995; Holland, 2004). The development of *behavioural autonomy* of the current incentive value of the goal follows directly from the architecture of the AC model. Continued reinforcement should strengthen the S→R association in the habit memory to such a degree that the presentation of the stimulus can reliably trigger the motor unit for the response before the longer feedback pathway through the associative memory can evaluate whether the outcome is currently a goal for the animal. Whether or not extensive training establishes comparable behavioural autonomy in humans awaits investigation and, to date, the only experimental demonstration of behavioural autonomy in adult humans is that engendered by response conflict due to *stimulus-outcome associative interference* (de Wit et al., 2007).

De Wit et al. (2007) trained adults on instrumental discriminations, in which each trial started with the presentation of a fruit icon on a computer screen signalling whether a left or right key press was correct. Choice of the correct response yielded another fruit icon and points, whereas the incorrect response did not yield a fruit outcome and earned zero points. Conflict was induced in an *incongruent discrimination* by having the same fruit function as the discriminative stimulus for one response in one component of the discrimination, and as an outcome for the opposite response in the other component. For example, the correct left key press to a coconut stimulus yielded a cherry outcome, whereas the coconut was the outcome for the correct right key press to the cherry stimulus. Whereas limited training on a standard instrumental discrimination should give rise to goal-directed performance, the incongruent discrimination should not be solvable in this manner because each fruit icon should become associated with opposite responses in its roles of stimulus and outcome. However, although de Wit et al. (2007) found that the acquisition of the incongruent discrimination was retarded relative to a standard discrimination participants did learn to choose the correct response. Furthermore, de Wit et al. showed that under these circumstances participants do not encode the R-O relationship but rather solve the task by acquiring simple S→R associations. Unlike in the standard discrimination (instructed) devaluation of one of the outcomes had no impact on the response choice in the incongruent discrimination. Therefore, stimulus-outcome associative interference

can cause a shift in learning strategy from goal-directed action to habits in humans (in Experiment 2; and in rats in Experiment 1). Whereas extensive training is proposed to engender behavioural autonomy mainly through enhancing the S-R feed-forward pathway in the AC model, the incongruent discrimination renders the feedback pathway ineffective by bringing about conflict in the associative memory leaving the residual control with the feed-forward pathway.

Behavioural autonomy is also observed in other conditions that should render the feedback pathway in effective, at least in rats. One such condition is when there is only a weak instrumental contingency between the action and the outcome. The strength of this contingency can be varied by contrasting training on different reinforcement schedules, which basically fall into two classes. The first are *ratio* schedules in which on average there is a fixed probability that an action will produce an outcome. A ratio schedule models cases in which there is an inexhaustible source of resources or outcomes so that the more the agent works the more outcomes it receives. Ratio schedules can be contrasted with another class of schedule in which the resources deplete when taken and then regenerate with time. An example would be the task faced by a nectar foraging hummingbird. Once it has depleted a flower of its nectar, however frequently it re-visits the flower it will not receive further nectar until the flower has regenerated the nectar resource. Such depleting-regenerating sources are modelled by *interval* schedules under which there is some average interval between available resources and which have an upper limit on the rate at which outcomes can be produced. Therefore, with interval schedules there is no constant ratio between the rate of responding and the number of outcomes gained. As ratio schedules therefore arrange a much stronger R-O relationship than interval schedules, they should generate a stronger representation of the instrumental contingency in the associative memory, or in other words a stronger F→O association. Ratio performance should therefore engage the feedback loop of the AC model more than interval performance and so should be more susceptible to outcome devaluation, a prediction that has received support from the rodent literature (Dickinson, Nicholas, & Adams, 1983; Hilario, Clouse, Yin, & Costa, 2007).

The instrumental behaviour of very young infants also appears to be behaviourally autonomous. When discussing the desire criterion, we noted that infants over the age of approximately 2 years are able to acquire goal-directed actions. In contrast, instrumental performance on the same task by younger infants is insensitive to outcome devaluation (Klossek et al., 2008). Within the AC model, the developmental transition from autonomous habitual behaviour to goal-directed action is linked to the engagement of the feedback loop through the associative memory and incen-

tive system, which Klossek et al. (2008) speculated may depend upon the maturation of the cortical-striatal-thalamic loops.

Balleine and Ostlund proposed that the capacity for goal-directed action depends upon the functioning of such loops on the basis of the fact that lesions in components of these loops render instrumental responding behaviourally autonomous (Balleine, 2005; Balleine & Ostlund, 2007). Although rats with lesions to prelimbic prefrontal cortex (Balleine & Dickinson, 1998; Corbit & Balleine, 2003; Killcross & Coutureau, 2003), the dorsomedial striatum (Yin, Ostlund, Knowlton, & Balleine, 2005), the core of the nucleus accumbens (Corbit, Muir, & Balleine, 2001), and the dorsomedial thalamus (Corbit, Muir, & Balleine, 2003) can all acquire instrumental responding, their performance is not directly sensitive to subsequent outcome devaluation. The functioning of the feedback loop of the AC model may therefore critically depend upon the integrity of the relevant cortical-striatal-thalamic loops.

In summary, the AC model provides a framework for integrating the diverse circumstances under which instrumental behaviour becomes autonomous of the current value of the goal: extensive training, conflict, development and neurobiological dysfunction.

Integrating O-R and R-O learning

As discussed above, neurobiological dissociations which demonstrate a loss of the capacity for goal-directed action while leaving basic S-R habit learning intact, accord with the structure of the AC model. In contrast, demonstrations that other brain dysfunctions leave rats with the capacity for goal-directed action while abolishing the development of behavioural autonomy appear to be problematic for the AC model. Dysfunction in the rodent infralimbic prefrontal cortex (Killcross & Coutureau, 2003) and dorsolateral striatum (Yin, Knowlton, & Balleine, 2004) prevents the development of behavioural autonomy so that instrumental responding remains sensitive to outcome devaluation even after extensive training. The reason why this result is potentially problematic for the AC model is that the architecture of the model assumes that the capacity for goal-directed action emerges from the modulation of the S-R feed-forward pathway by the feedback loop through the associative memory. In other words, the associative memory is 'parasitic' on the habit memory. Therefore, dysfunction in the habit memory mediating the S-R pathway should also impact on goal-directed action.

One possible way to resolve this dilemma is by incorporating O-R learning into the model. We have already noted that O-R learning plays an important role in the control of instrumental behaviour as manifest in direct and indirect PIT priming of responses by outcomes. Although PIT

priming effects are insensitive to the current value of the outcome and O-R learning therefore appears to fail to meet the desire criterion, these priming effects do require that O-R learning is incorporated into any complete associative account of instrumental action.

The connection between the response (R) unit in the habit memory and the response feedback (F) unit the associative memory is depicted as being bidirectional in Fig. 1, a bidirectionality that allows for ideomotor influences within the AC model. However, not only is the sensory feedback from performing an action highly response contingent and contiguous, so are the sensory properties of the outcome itself, albeit to a lesser degree. Consequently, not only should the F unit in the associative memory develop a bidirectional association with the R unit, but so should the outcome (O) unit (see Fig. 1). The incorporation of O-R learning into the AC model then allows for two routes in the control of performance by outcome value.

The first route is the one that has already been developed, namely the priming of a response by the S→R association in the habit memory and the concurrent incentive evaluation of the outcome through the R⇔F→O associative chain. This represents the case in which the thought of a response activates the thought of the outcome. The alternative route is that via the bidirectional R⇔O association so that the activation of the outcome representation in the associative memory not only produces evaluation of the outcome through the reward unit in the incentive system but also concurrent priming of the response in the habit memory, thereby resulting in summation of activation on the motor unit for this response if the outcome is currently a goal. This is the case in which the thought of a goal directly primes the associated response.[2]

On the basis of neurobiological double dissociations, some researchers (e.g. Daw et al., 2005) have argued that independent processes mediate habitual and goal-directed behaviour. However, the integration of R-O and O-R learning within the AC model also allows for the observed dissociations. The loss of the capacity for goal-directed action results from dysfunction in the positive feedback loop, whereas the failure to develop behavioural autonomy reflects failure of S-R learning in the habit memory, leaving the response primed through the O→R association.

Summary and conclusions

Actions are truly goal-directed only if the decisions to undertake them are based on the knowledge of the causal relationship between the actions and their consequences (belief criterion), and if those consequences are currently a goal (desire criterion). In the current paper, two classes of associative theories of goal-directed action were reviewed, namely, O-R (or ideomotor) theory and R-O theory. In fact, it seems clear that in most situations we form both types of associations. While sometimes the thought of an action reminds us of the consequences (R→O), on other occasions the thought of a goal reminds us of the appropriate course of action (O→R). The question of interest therefore is what differential roles these associations may play in decision-making.

We argue that while O-R theory offers the most parsimonious account of response priming effects by outcomes as well as by cues for outcomes (PIT), an R-O mechanism may be responsible for the sensitivity of goal-directed behaviour to the current value of its consequences (outcome revaluation). With the aim of capturing the full complexity of goal-directed action we have integrated these two accounts in the associative-cybernetic (AC) model. This model postulates associative as well as incentive processes to account for response selection via R→O and O→R associations.

The additional benefit of the AC model is that it offers an explanation for the transition from goal-directed actions to S-R habits, or the development of behavioural autonomy. Habit formation renders behaviour less flexible but it has the advantage that it allows for fast selection of appropriate responses in stable environments despite distraction, as for example the changing of gears while navigating through busy traffic. So, although habit formation has to our knowledge only been demonstrated in a single study in humans (de Wit et al., 2007), it makes intuitive sense that much of our daily behaviour is driven directly by S→R associations. We would therefore like to argue that translating behavioural assessments of habitual versus goal-directed action, as developed in animal research, to the domain of cognitive

---

[2] The O→R pathway can capture response priming as envisaged by Pavlov in his bidirectional theory (Pavlov, 1932; Asratyan, 1974). According to this theory, O→R associations are formed as a consequence of experience with the consequent R-O contingency. An alternative O-R account was proposed by Trapold and Overmier (1972), who argued that O→R associations arise because environmental cues elicit an anticipatory representation "O" of the instrumental outcome that precedes performance of the instrumental response. This antecedent "O"-R relationship then allows for the representation of the anticipated outcome to enter into association with the instrumental response through an S-R reinforcement mechanism. The latter account could be integrated into the AC model by allowing "O" units in the habit memory to enter into association with the response units. However, we should note that there is good reason to favour the bidirectional theory over the "O"-R account. Whereas bidirectional theory predicts specific O→R priming following concurrent instrumental training, it is less clear that Trapold and Overmier (1972) would predict this effect in the absence of reliable predictors of the outcomes during training. However, PIT does occur following concurrent training (see e.g. Colwill & Rescorla, 1988). Moreover, Rescorla and Colwill (1989) showed that Pavlovian stimuli activate responses that previously yielded the same outcome rather than responses that during training were preceded by anticipation of this outcome (see also Rescorla, 1992).

neuroscience can provide us with insights into associative and neural mechanisms of human decision-making. Moreover, habit formation could be an important concept in neuropsychiatry. For instance, compulsive behaviours in obsessive-compulsive disorder (e.g. Evans, Lewis, & Iobst, 2004) and drug-seeking in addiction have been characterised as habitual (e.g. Everitt, Dickinson, & Robbins, 2001; Everitt & Robbins, 2005; Redish, Jensen, & Johnson, 2008). Parkinson's disease, on the other hand, has been associated with impaired habit formation (e.g. Yin & Knowlton, 2006). The AC model offers a promising framework within which these suggestions of accelerated and impaired habit formation can be meaningfully explored. Finally, application of the model to cognitive development gives rise to the intriguing suggestion that the ability to form S→R associations, as well as O→R associations, emerges earlier in child development than the ability to behave in a goal-directed manner, which depends on the formation of F→O associations in the associative memory as well as the engagement of the incentive system.

Finally, while we argue that the AC model captures much of the complexity of animal as well as human behaviour, we surely do not wish to negate the importance of abstract rule-based reasoning for human decision-making (for dual-process theories, see e.g. Evans, 2003; Sloman, 1996). Indeed, we subscribe to the notion that ultimately associative processes, incentive processes, as well as higher-order knowledge interact in the control of much of daily behaviour. However, we insist that an understanding of the operation of fundamental associative and incentive processes as captured by the AC model is an important prerequisite for gaining insight into human intentional action.

References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 34B*, 77–98.

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 33B*, 109–121.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglas, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111*, 1036–1060.

Ansorge, U. (2002). Spatial intention-response compatibility. *Acta Psychologia, 109*(3), 285–299.

Asratyan, E. A. (1974). Conditional reflex theory and motivated behavior. *Acta Neurobiologiae Experimentalis, 34*, 15–31.

Balleine, B. (2005). Neural basis of food-seeking: affect, arousal, and reward in corticostriatolimbic circuits. *Physiology and Behavior, 86*(5), 717–730.

Balleine, B. W., & Dickinson, A. (1998a). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37*, 407–419.

Balleine, B. W., & Dickinson, A. (1998b). The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. *Animal Learning & Behavior, 26*(1), 46–59.

Balleine, B. W., & Ostlund, S. B. (2007). Still at the choice-point: action selection and initiation in instrumental conditioning. *Annals of the New York Academy of Sciences, 1104*, 147–171.

Baxter, D. J., & Zamble, E. (1982). Reinforcer and response specificity in appetitive transfer of control. *Animal Learning & Behavior, 10*, 201–210.

Beckers, T., De Houwer, J., & Eelen, P. (2002). Automatic integration of non-perceptual action effect features: the case of the associative affective Simon effect. *Psychological Research, 66*(3), 166–173.

Blundell, P., Hall, G., & Killcross, S. (2001). Lesions of the basolateral amygdala disrupt selective aspects of reinforcer representation in rats. *Journal of Neuroscience, 21*(22), 9018–9026.

Bray, S., Rangel, A., Shimojo, S., Balleine, B., & O'Doherty, J. P. (2008). The neural mechanisms underlying the influence of Pavlovian cues on human decision making. *Journal of Neuroscience, 28*, 5861–5866.

Carpenter, W. B. (1852). On the influence of suggestion in modifying and directing muscular movement, independently of volition. *Proceedings of the Royal Institution*, *1*, 147–154.

Chutuape, M. A. D., Mitchell, S. H., & de Wit, H. (1994). Ethano preloads increase ethanol preference under concurrent random-ratio schedules in social drinkers. *Experimental and Clinical Psychopharmacology, 2*, 310–318.

Colwill, R. M., & Motzkin, D. K. (1994). Encoding of the Unconditioned Stimulus in Pavlovian Conditioning. *Animal Learning & Behavior, 22*(4), 384–394.

Colwill, R. M., & Rescorla, R. A. (1985). Postconditioning devaluation of a reinforcer affects instrumental responding. *Journal of Experimental Psychology: Animal Behavior Processes, 11*(1), 120–132.

Colwill, R. M., & Rescorla, R. A. (1988). Associations between the discriminative stimulus and the reinforcer in instrumental learning. *Journal of Experimental Psychology: Animal Behavior Processes, 14*(2), 155–164.

Corbit, L. H., & Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behavioural Brain Research, 146*, 145–157.

Corbit, L. H., Janak, P. H., & Balleine, B. W. (2007). General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. *The European Journal of Neuroscience, 26*(11), 3141–3149.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2001). The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. *Journal of Neuroscience, 21*(9), 3251–3260.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *European Journal of Neuroscience, 18*, 1286–1294.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704–1711.

de Borchgrave, R., Rawlins, J. N., Dickinson, A., & Balleine, B. W. (2002). Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. *Experimental Brain Research, 144*(1), 50–68.

de Wit, S., Niry, D., Wariyar, R., Aitken, M. R. F., & Dickinson, A. (2007). Stimulus-outcome interactions during conditional discrimination learning by rats and humans. *Journal of Experimental Psychology: Animal Behavior Processes, 33*(1), 1–11.

Dickinson, A. (1994). Instrumental conditioning. In N. J. Mackintosh (Ed.), *Animal learning and cognition* (pp. 45–79). San Diego, CA: Academic Press.

Dickinson, A., & Balleine, B. W. (1993). Actions and responses: the dual psychology of behaviour. In N. Eilan & R. A. McCarthy (Eds.), *Spatial representation: problems in philosophy and psychology* (pp. 277–293). Malden: Blackwell.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior, 22*(1), 1–18.

Dickinson, A., & Balleine, B. (2008). The cognitive/motivational interface. In M. L. Kringelbach & K. C. Berridge (Eds.), *Pleasures of the brain. The neural basis of taste, smell and other rewards*. Oxford: Oxford University Press.

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. (1995). Motivational control after extended instrumental training. *Animal Learning & Behavior, 23*(2), 197–206.

Dickinson, A., Campos, J., Varga, Z. I., & Balleine, B. (1996). Bidirectional instrumental conditioning. *Quarterly Journal of Experimental Psychology, 49B*(4), 289–306.

Dickinson, A., Nicholas, D. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 35B*, 35–51.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology-Human Perception and Performance, 27*(1), 229–240.

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research, 68*(2–3), 138–154.

Estes, W. K. (1943). Discriminative conditioning: I. A discriminative property of conditioned anticipation. *Journal of Experimental Psychology, 32*, 150–155.

Estes, W. K. (1948). Discriminative conditioning: II. Effects of a Pavlovian conditioned stimulus upon a subsequently established operant response. *Journal of Experimental Psychology, 38*, 173–177.

Estes, W. K. (1969). Reinforcement in human learning. In J. T. Tapp (Ed.), *Reinforcement and Behavior* (pp. 63–94). New York: Academic Press.

Evans, J. S. B. T. (2003). In two minds: dual process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454–459.

Evans, D. W., Lewis, M. D., & Iobst, E. (2004). The role of the orbitofrontal cortex in normally developing compulsive-like behaviors and obsessive-compulsive disorder. *Brain and Cognition, 55*(1), 220–234.

Everitt, B. J., Dickinson, A., & Robbins, T. W. (2001). The neuropsychological basis of addictive behavior. *Brain Research Reviews, 36*(2–3), 129–138.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience, 8*(11), 1481–1489.

Fagen, J. W., & Rovee, C. K. (1976). Effects of quantitative shifts in a visual reinforcer on the instrumental response of infants. *Journal of Experimental Child Psychology, 21*(2), 349–360.

Fagen, J. W., Rovee, C. K., & Kaplan, M. G. (1976). Psychophysical scaling of stimulus similarity in 3-month-old infants and adults. *Journal of Experimental Child Psychology, 22*(2), 272–281.

Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science, 122*(3160), 157–158.

Gormezano, I., & Tait, R. W. (1976). The Pavlovian analysis of instrumental conditioning. *The Pavlovian Journal of Biological Science, 11*(1), 37–55.

Greve, W. (2001). Traps and gaps in action explanation: theoretical problems of a psychology of human action. *Psychological Review, 108*, 435–451.

Herbart, J. F. (1825). *Psychologie als Wissenschaft neu gegrundet auf Erfahrung, Metaphysik und Mathematik. Zweiter, analytischer Teil*. Konigsberg: Unzer.

Hershberger, W. A. (1986). An approach through the looking glass. *Animal Learning & Behavior, 14*, 443–451.

Heyes, C., & Dickinson, A. (1990). The intentionality of animal action. *Mind and Language, 5*, 87–104.

Hilario, M. R. F., Clouse, E., Yin, H. H., & Costa, R. M. (2007). Endocannabinoid signaling is critical for habit formation. *Frontiers in Integrative Neuroscience*, 1.

Hogarth, L., Dickinson, A., Wright, A., Kouvaraki, M., & Duka, T. (2007). The role of drug expectancy in the control of human drug seeking. *Journal of Experimental Psychology: Animal Behavior Processes, 33*(4), 484–496.

Holland, P. C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *Journal of Experimental Psychology: Animal Behavior Processes, 30*(2), 104–117.

Hommel, B. (1993). The role of attention for the Simon effect. *Psychological Research, 55*(3), 208–222.

Hommel, B. (1996). The cognitive representation of action: automatic integration of perceived action effects. *Psychological Research, 59*(3), 176–186.

Hommel, B. (2003). Planning and representing intentional action. *The Scientific World Journal* (3), 593–608.

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *The Behavioral and Brain Sciences, 24*(5), 849–878. discussion 878-937.

James, W. (1890). *The principles of psychology*. New York: Dover Publications.

Kenward, B., Folke, S., Holmberg, J., Johansson, A., & Gredeback, G. (2009). Goal-directedness and decision-making in infants. *Developmental Psychology* (in press).

Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex, 13*(4), 400–408.

Klossek, U., Russell, J., & Dickinson, A. (2008). The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology: General, 137*, 39–51.

Kruse, J. M., Overmier, J. B., Konz, W. A., & Rokke, E. (1983). Pavlovian conditioned stimulus effects upon instrumental choice behavior are reinforcer specific. *Learning and Motivation, 14*, 165–181.

Kunde, W., Koch, I., & Hoffmann, J. (2004). Anticipated action effects affect the selection, initiation, and execution of actions. *Quarterly Journal of Experimental Psychology, 57A*(1), 87–106.

Lotze, R. H. (1852). Medicinische Psychologie oder die Physiologie der Seele. In *Weidmann'sche Buchhandlung*. Leipzig.

Ludwig, A. M., Wikler, A., Stark, L. H., & Lexington, K. (1974). The first drink: psychobiological aspects of craving. *Archives of General Psychiatry, 30*, 539–547.

Meck, W. H. (1985). Postreinforcement signal processing. *Journal of Experimental Psychology: Animal Behavior Processes, 11*(1), 52–70.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Pavlov, I. P. (1927). *Conditional reflexes; an investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford University Press.

Pavlov, I. P. (1932). The reply of a physiologist to psychologists. *Psychological Review, 39*, 91–127.

Pithers, R. T. (1985). The roles of event contingencies and reinforcement in human autoshaping and ommission responding. *Learning and Motivation, 16*, 210–237.

Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process. *Behavioural Brain Science, 31*(4), 415–437.

Rescorla, R. A. (1992). Response outcome versus outcome response associations in instrumental learning. *Animal Learning & Behavior, 20*(3), 223–232.

Rescorla, R. A. (1994). Transfer of instrumental control mediated by a devalued outcome. *Animal Learning & Behavior, 22*(1), 27–33.

Rescorla, R. A., & Colwill, R. M. (1989). Associations with anticipated and obtained outcomes in instrumental learning. *Animal Learning & Behaviour, 17*(3), 291–303.

Rovee-Collier, C. K. (1983). Infants as problem-solvers: a psychobiological perspective. In M. D. Zeiler & P. Harzem (Eds.), *Advances in the analysis of behaviour* (Vol. 3). Oxford, UK: Wiley.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.

Stewart, J., & de Wit, H. (1987). Reinstatement of drug-taking behavior as a method of assessing incentive motivational properties of drugs. In M. A. Bozarth (Ed.), *Methods of assessing the reinforcing properties of abused drugs* (pp. 211–227). New York: Springer.

Stock, A., & Stock, C. (2004). A short history of ideo-motor action. *Psychological Research, 68*(2–3), 176–188.

Sutton, R. S., & Barto, A. G. (1981). An adaptive network that constructs and uses an internal model of its world. *Cognition and Brain Theory, 4*, 217–246.

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human pavlovian-instrumental transfer. *The Journal of Neuroscience, 28*(2), 360–368.

Thorndike, E. L. (1911). *Animal intelligence: experimental studies*. New York: Macmillan.

Thorndike, E. L. (1931). *Human learning*. New York: Century.

Trapold, M. A., & Overmier, J. B. (1972). The second learning process in instrumental learning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 427–452). New York: Appleton-Cwntury-Croft.

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *The Journal of Neuroscience, 27*(15), 4019–4026.

Wolpaw, J. R. (1997). The complex structure of a simple memory. *Trends in Neuroscience, 20*(12), 588–594.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3*, 1212–1227.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews: Neuroscience, 7*(6), 464–476.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *The European Journal of Neuroscience, 19*(1), 181–189.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience, 22*, 513–523.

Ziessler, M. (1998). Response-effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 962–978.

Ziessler, M., & Nattkemper, D. (2001). Learning of event sequences is based on response-effect learning: further evidence from a serial reaction task. *Journal of Experimental Psychology-Learning. Memoty and Cognition, 27*(3), 595–613.

Ziessler, M., & Nattkemper, D. (2002). Effect anticipation in action planning. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and performance XIX* (pp. 645–672). Oxford: Oxford University Press.