

Title: Making habits measurable beyond what they are not: a focus on associative dual-process models

Authors: Poppy Watson¹, Claire O'Callaghan², Iain Perkes¹, Laura Bradfield^{3,4}, Karly Turner¹

1. School of Psychology, UNSW Sydney, Australia.
2. Brain and Mind Centre and School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Australia.
3. School of Life Sciences, Faculty of Science, University of Technology Sydney, New South Wales 2007, Australia.
4. St. Vincent's Centre for Applied Medical Research, St. Vincent's Hospital Sydney, New South Wales 2011, Australia.

Corresponding author:

Dr Poppy Watson
School of Psychology
UNSW Sydney
Sydney
NSW 2052
Australia
poppy.watson@unsw.edu.au

Conflict of interest: The authors declare no conflict of interest.

Funding Acknowledgement: This work was supported by an Australian Research Council Discovery Early Career Researcher Award awarded to PW (DE200100591).

Abstract

Habits are the subject of intense international research. Under the associative dual-process model the outcome devaluation paradigm has been used extensively to classify behaviours as being either goal-directed (sensitive to shifts in the value of associated outcomes) or habitual (triggered by stimuli without anticipation of consequences). This has proven to be a useful framework for studying the neurobiology of habit and relevance of habits in clinical psychopathology. However, in recent years issues have been raised about this rather narrow definition of habits in comparison to habitual behaviour experienced in the real world. Specifically, defining habits as the absence of goal-directed control, the very specific set-ups required to demonstrate habit experimentally and the lack of direct evidence for habits as stimulus-response behaviours are viewed as problematic. In this review paper we address key critiques that have been raised about habit research within the framework of the associative dual-process model. We then highlight novel research approaches studying different features of habits with methods that expand beyond traditional paradigms.

Keywords: Habits, goal-directed control, associative dual-process models, outcome devaluation, stimulus-response

1. Introduction

Habit is used frequently in everyday language to refer to behaviours that seem to have an inexorable pull over us. These can have both negative, e.g., a drug use, as well as positive, e.g., healthy exercise, connotations. Despite intense international research into habits (>7000 papers published in the last twelve months with “habit” in the abstract; *PubMed*, June 2022), there is ambiguity regarding the definition and measurement of the construct.

Common to most definitions is the notion that a habit is a behaviour performed quickly, automatically, and triggered by stimuli or contexts (see for reviews: Balleine & Dezfouli, 2019; Gardner, 2015; Robbins & Costa, 2017; Wood & Runger, 2016). However, within this definition, there is no unanimous criteria for habits—with different approaches between fields. For example, researchers in Health Psychology use self-reported indices of automaticity to classify behaviour as habitual (Gardner et al., 2011). Those studying the development of skilled behavioural repertoires rely on the speed at which behaviours are carried out as a proxy for automaticity and habits (Wu et al., 2004; Banca et al., 2020; review: Du et al., 2022). However, many of these approaches do not provide a clear threshold beyond which a goal-directed action can be said to transition to a habitual response. By contrast, the associative dual-process framework has a narrow and clear definition of whether a behaviour is habitual or under goal-directed control (Dickinson, 1985).

Developed in the 1980s, the associative dual-process framework places the emphasis on the control of action through either its association with an outcome (response-outcome; R-O) or its potentiation by stimuli (stimulus-response; S-R). Goal-directed control, driven by anticipation of outcomes, is flexible but cognitively demanding. To demonstrate that a behaviour is habitual, and therefore driven by S-R associations, it is necessary to demonstrate that the behaviour is not goal-directed. That is, the behaviour will be carried out even when the outcome it produces is currently undesirable (outcome devaluation), and

persists when the contingency between response and the outcome has been degraded (de Wit & Dickinson, 2009; Dickinson, 1985, 1994).

Habits, as described under the dual-process model, have been intensively investigated - particularly in the field of behavioural neuroscience. Significant progress has been made in mapping the neurobiological substrates of both habits and goal-directed control in animals and humans (de Wit et al., 2012; Reber et al., 2017; Watson et al., 2018; reviews: Balleine & O'Doherty, 2010; Yin & Knowlton, 2006; Balleine, 2005, 2019). In the field of clinical neuropsychology overreliance on habits at the expense of goal-directed control has been related to various psychopathologies (Delorme et al., 2016; Gillan et al., 2011; Reber et al., 2017; see for review: Verhoeven & de Wit, 2018). However, in recent years limitations of the dual-process framework as currently applied in experimental paradigms and the use of outcome devaluation insensitivity as the main diagnostic test of habits have been raised (Figure 1). Specifically, the current framework – which has produced remarkable insights – may lend itself to a rather narrow definition of habits in comparison to those experienced in the real world. In this review paper we address key critiques that have been raised about habit research within the framework of the associative dual-process model. We then highlight novel research approaches studying the various features of habits using methods that expand beyond traditional paradigms.

2. Associative Dual-Process Model of Behaviour

Goal-directed control and habits are formalised in the associative dual-process model in terms of associative links (de Wit & Dickinson, 2009; Dickinson, 1985, 1994). It is proposed that during instrumental learning, both stimulus-response (S-R) and bidirectional outcome-response (R-O/O-R) links are formed (Ostlund & Balleine, 2007). Initially, responding is driven by the expectation and desire for an outcome that is associated with an action. However, with repeated experience, S-R associations can drive behaviour, such that the response will persist even if the outcome is not currently desired. To establish whether

behaviour is under goal-directed control, alterations in responding are measured either during contingency degradation or in extinction following outcome devaluation (see Figure 2). If a participant is goal-directed, they will be sensitive to the fact that the response no longer produces the outcome (contingency degradation) or that the outcome is no longer valuable (outcome devaluation), and they will reduce that response. By contrast, if the response continues under these conditions it is argued to be an S-R habit – directly triggered by environmental stimuli and divorced from the representation of the outcome it produces. Although both insensitivity to R-O contingency degradation and insensitivity to outcome devaluation are argued to be the two hallmarks of habit (de Wit & Dickinson, 2009), in reality the majority of habit research in this field has focused on outcome devaluation as the principal diagnostic test of whether behaviour is goal-directed or habitual. Habits assessed with the outcome devaluation paradigm are therefore the focus of this review.

The associative model has been extended to propose a computational model where habits are chunked into action sequences initiated by a goal-directed controller (Balleine & Dezfouli, 2019; Dezfouli & Balleine, 2012, 2013). Outcomes that fall within the boundary of the chunked action sequence are not assessed in a goal-directed manner (Balleine & Dezfouli, 2019). This framework is well suited to understanding more complex slips-of-action in daily life. For example, the goal-directed controller could initiate driving to work via the drycleaners but once in the familiar environment of the car, the habit controller takes over allowing you to put on a seatbelt, change gears and stop at red lights. However, the unintended result may be that the individual drives on ‘autopilot’ straight to work. Integrative models such as these help to bridge simplified experimental associations with the complex real-world experience of habits.

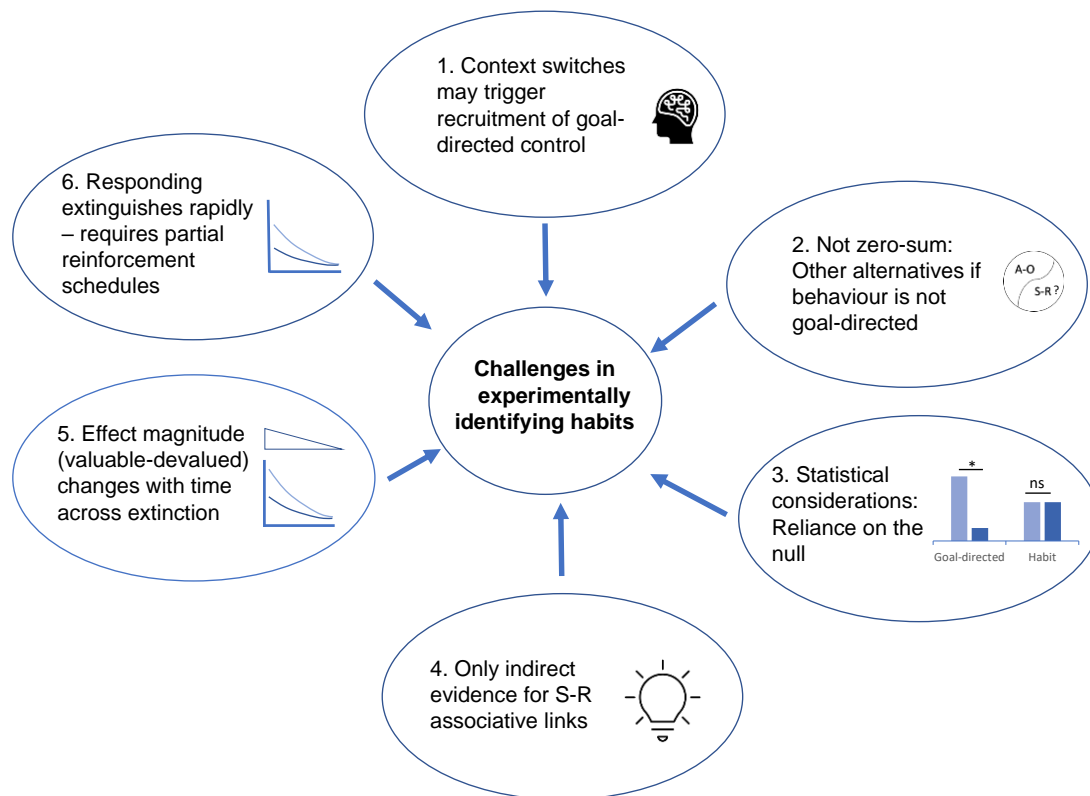


Figure 1. Challenges in interpreting the absence of experienced value sensitive behaviour as habit. Clockwise from top: 1. When changes in context between training and test are noticed by the subject (human or animal), goal-directed control may be quickly recruited, suppressing habits. 2. There may be reasons other than habit (e.g., apathy, confusion) as to why subjects do not show goal-directed actions at test. 3. Statistically a null effect (assumed indicative of habit) could occur due to noise, insufficient power or small effect size. 4. Direct (rather than indirect) evidence for S-R associative links is difficult to demonstrate experimentally. 5. Testing under extinction also leads to a time-dependent reduction in responding and change in contingency that can alter the effect of devaluation over time, requiring careful consideration for researchers in terms of how data is analysed. 6. Partial reinforcement schedules are required to ensure that a sufficient number of responses can be measured during the extinction test, limiting the assessment of habits in continuously reinforced behaviours.

3. Evidence for Dual Systems

As a general concept, dual-process models of behaviour have been criticized on the grounds that it is inappropriate to dichotomise behaviour without rigorous assessment of whether behaviour might in fact be on a continuum (Keren & Schul, 2009; Melnikoff & Bargh, 2018). However, discrete neurobiological networks underlie the control of goal-directed actions and habits (see Balleine & O’Doherty, 2010 for review). One clear distinction is in the medial and lateral subregions of the dorsal striatum, which are important for goal-directed actions and habits respectively. Manipulations in rodents have demonstrated that control of actions can

be bidirectionally shifted, such that loss of function in the dorsomedial striatum (DMS) results in habitual responding (Yin et al., 2005) and loss of function in the dorsolateral striatum (DLS) revert habitual behaviour to goal-directed control (Yin et al., 2004). Dissociable roles have also been described in subregions of the prefrontal cortex (prelimbic vs. infralimbic cortices: Killcross & Coutureau, 2003) and amygdala (basolateral vs. central nuclei: Lingawi & Balleine, 2012; Ostlund & Balleine, 2008).

There is, therefore, clear evidence for dissociable neurobiological systems underlying habits and goal-directed control, suggesting a dual-process model is plausible. However, these two systems likely compete and collaborate such that observable behaviour is better understood on a continuum rather than an absolute categorical basis. Indeed, it has been shown that by using concurrent ratio and interval schedules, mice can be trained to perform essentially the same action (left lever press) for the same outcome (pellet) in either a goal-directed or habitual manner, depending on the current context (Gremel & Costa, 2013). This protocol demonstrates the adaptive and flexible ability to shift between strategies within an individual. It is also possible that individual responses are binary but can rapidly oscillate throughout testing, producing a subset of habitual responses intercalated with goal-directed actions (Bouton, 2021; Bouton et al., 2020; Ceceli et al., 2020; Gremel & Costa, 2013; Halbout et al., 2019). Across multiple trials we therefore end up with a summed index that incorporates independent actions that could discretely be classified as either habitual or goal-directed. This account would speak to both the independent neural circuits controlling two unique forms of action control, as well as the potential to observe behaviour on a spectrum.

4. Habits as Null Effects

The outcome devaluation paradigm elegantly demonstrates that outcomes are encoded by animals (including humans) during instrumental training procedures. Specifically, at test, subjects selectively suppress responding for now-devalued outcomes, whilst continuing to

respond for still-valuable outcomes (thus demonstrating goal-directed behaviour). However, it has recently been highlighted that relying on a null effect when comparing responding for devalued versus still-valued outcomes in the search for habit is problematic (Balleine & Dezfouli, 2019; Berridge, 2021; Robbins & Costa, 2017). There are two main issues with interpreting null effects as habits. The first issue is the assumption that behaviour that is not goal-directed must be habitual: that this is a zero-sum game with only two options. There are likely many situations where a loss of goal-directed control is not going to equate to habit (Balleine & Dezfouli, 2019), for example if the subject is confused, has not learned the associations yet, or is exploring response options (particularly during the devaluation test phase when the subject detects that responses are no longer reinforced). The second issue is statistical — animal studies are often underpowered (Button et al., 2013) compounding the issue of relying on the null as evidence for habitual control (Berridge, 2021; Vadillo et al., 2020). Furthermore, because the test phase is conducted in extinction, subjects may respond to the changed contingency and eventually stop responding, creating a risk that the length of the analysis time window can bias the results. For example, a short window may be more likely to show goal-directed responding as this is where the greatest magnitude of divergence between valued and devalued responding is often observed, but as responding diminishes with extinction the likelihood of accepting the null hypothesis strengthens (but see Dezfouli et al., 2014 for an example where goal-directed responding can return later in the extinction test). This variability over time increases the chance that researchers may inadvertently bias their findings, simply by adjusting the parameters of the statistical analysis. Increased transparency can be achieved by presenting the curves for both valued and devalued responses over time, however this is not always reported.

If behaviour is on a continuum ranging from extremely-sensitive to extremely-insensitive to outcome devaluation, then as mentioned earlier, a binary test of habits at the group level is problematic and individual differences in response strategy need greater consideration. A recent study in humans using the outcome devaluation test with food rewards reported a

multimodal distribution in responding for devalued outcomes at test (Pool et al., 2021). One cluster (approximately one third) of participants showed sensitivity to outcome devaluation (responding significantly more for the valuable reward) whereas another cluster (two thirds of the sample) responded approximately equally often for the devalued and valued outcomes. However, at the group level the difference between responding for the devalued outcome relative to the valuable outcome was significantly different, forcing the conclusion that overall the sample were relying on goal-directed control. The authors reported that the affective components of chronic stress (e.g., worry and discontent) were a moderating factor - individuals high in affective stress showing insensitivity to outcome devaluation much faster than those who scored low in affective stress.

While these findings are largely exploratory, they reinforce the idea that by examining group-level differences in responding for devalued versus valuable outcomes at test, studies may be underestimating the prevalence of habitual responses that are insensitive to outcome devaluation. It should be noted that individual differences may be more variable in human research, with less variability in response patterns typically observed in well controlled animals studies (cf Figure 1, Bradfield et al., 2020). However, animals are also not completely homogenous in their food preferences, motivation to respond for rewards, memory abilities and susceptibility to stress (Antoniuk et al., 2019; Kuhn et al., 2019). As such, more sensitive analysis techniques may be required in both human and animal studies to characterise behaviour at an individual participant level, at different stages of the test phase. It was recently highlighted that one limitation of the outcome devaluation test in rodents is that the influence of extinction from the first to second test session (counterbalanced for devaluation of outcomes), makes the examination of individual differences difficult (van Elzelingen et al., 2022). Ultimately, this results in the first test contributing more responses than the second, which can have a compounding impact when outcomes are not equally preferred (Vandaele et al., 2017). While counterbalancing conditions across animals is critical, this extinction effect minimises the utility of individual

results. One solution has been to conduct additional reinforced training sessions in between test days, however this may not suit all experimental designs (e.g. where the number of associative pairings is critical).

An alternative to the dichotomous approach is the continuous parameters used in the Model-based/Model-free (MB/MF) computational method (Daw et al., 2005, 2011). MB/MF models of action control also propose two distinct systems underlying behaviour: the MB system makes anticipatory models of likely consequences of potential responses and uses this to drive action selection. By contrast, the MF system caches historical outcome values associated with a particular response and simply selects the most valuable response. In addition to modelling independent MB and MF parameters, the relative balance between the two systems is modelled with a weighting parameter, ω (Daw et al., 2011). This moves beyond binary, allowing for assessment of the degree of MB/MF control for each individual participant.

Despite the appeal of the MB/MF approach, there are theoretical issues with simply mapping MF behaviour to S-R habits (Dezfouli & Balleine, 2012, 2013; Miller et al., 2018, 2019). It has been argued that studies showing largely overlapping neural architecture in MB and MF decision making demonstrate that both types of computations contribute to behaviour that should be considered value-driven and largely goal-directed (Miller et al., 2018, 2019) and that neither capture a habit controller (i.e., a mechanism where behaviour is elicited by the environment without evaluation of the consequences; see also Dezfouli & Balleine, 2012, 2013). Miller proposed a realignment of the existing computational models such that both MF and MB computations interact to drive goal-directed control and that S-R habits arise simply due to frequency of action. Under this model, S-R habits emerge from motoric repetition alone and are always dissociated from outcome value (Miller et al., 2018, 2019). The model appeals to an arbitrator (shifting the balance from goal-directed to habitual control) which accounts for experimental data that at first glance seem inconsistent with this

account (e.g., findings that extended training of behavioural responses does not lead to S-R habits: de Wit et al., 2018; Pool et al., 2021). Despite a lack of direct experimental evidence in support of this model, it does highlight the important role of repetition in habits.

Although there may be issues with mapping MB and MF behaviour directly to goal-directed and S-R habits, they could inspire a more nuanced dynamical model of outcome devaluation sensitivity/insensitivity across the extinction test. Human studies of habit using variants of the outcome devaluation task often use a devaluation sensitivity index (DSI) which is simply the number of responses made for devalued outcomes subtracted from the number of responses made for valuable outcomes, during the entire test phase. This continuous index has then been related to performance on other tasks or clinical indices (cf. Dietrich et al., 2016; Sjoerds et al., 2016). This approach could be used more widely, and calculated at different stages of the extinction test, although normalisation should be adopted to ensure that difference scores are not correlated to overall propensity to lever press.

5. Specific Experimental Parameters Required for Habits

Research groups that wish to study habits in animals tend to use very specific experimental parameters (see Figure 2A, Table 1 and Table 2). For example, interval schedules of reinforcement during instrumental training are known to more rapidly lead to behaviour that is insensitive to outcome devaluation, relative to ratio schedules of reinforcement (Dickinson et al., 1983). Specific reinforcers (such as sucrose solution relative to food pellets), differ in their sensitivity to devaluation and engender different response rates during the extinction test following outcome devaluation (Colwill & Rescorla, 1985b; Vandaele et al., 2017). Different outcome devaluation methods (e.g., satiation versus conditioned taste aversion) can influence whether habits are detected or not (Trask et al., 2020), with conditioned taste aversion generally being used for habit research. The altered motivational state of satiety or feeding of outcomes prior to testing may promote the shift from habits back to goal-directed control in and of itself, possibly comprising a kind of

context shift (Abiero & Bradfield, 2021; Parkes et al., 2016). Even within a specific devaluation method, different protocols can exert a large effect on the behaviour observed at test (Amaya et al., 2020).

In addition, habits do not emerge when animals are trained on more than a single instrumental R-O contingency (Colwill & Rescorla, 1985a, 1986; Kosaki & Dickinson, 2010). This challenges the ecological validity of translational habit research, given that humans readily develop habits in the real world, where the choice context appears to be vast (Vandaele & Ahmed, 2021). The question remains as to why behaviour remains sensitive to outcome devaluation when animals are trained on two contingencies, but habits can develop when only one R-O contingency is trained (Figure 2B, Table 1). It has been proposed that choice between two possible responses (and their associated outcomes) keeps the correlation between responses and their associated outcomes high and at the forefront of attention such that R-O associations dominate the control of behaviour (Kosaki & Dickinson, 2010; see Lingawi et al., 2016 for detailed discussion).

Table 1: Experimental considerations when studying habit formation in rodents

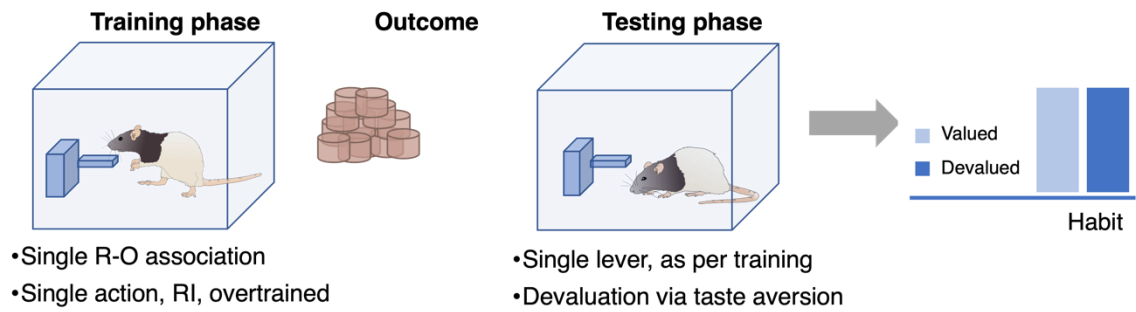
Key Factors	Observations	Key references
Schedules of reinforcement	Devaluation insensitive behaviour (habit) emerges more quickly with interval than ratio schedules of reinforcement. Leaner schedules of reinforcement might lead to habit more quickly.	Killcross & Coutureau, 2003; Gremel & Costa, 2013; Dickinson et al., 1983; Garr et al., 2020; Thrailkill et al., 2018.
Number of instrumental responses	Rats trained on more than one R-O contingency tend to remain goal-directed. One-lever training more reliably results in behaviour insensitive to outcome devaluation.	Kosaki & Dickinson, 2010; Colwill & Rescorla 1985a,
Type of reinforcer	Behaviour reinforced with sucrose more quickly becomes devaluation insensitive relative to food pellets.	Colwill & Rescorla 1985a; Vandaele et al., 2017
Type of instrumental response	Chain pulling and lever pressing may transition to habit at different rates.	Faure et al., 2005
Training Duration	Generally accepted that habits emerge from behavioural repetition but the exact amount of training required for habit is not clear.	Adams, 1982; Garr et al., 2020, Exp 4b; Garr et al., 2021; van Elzelingen et al., 2022;

Outcome devaluation protocol	LiCi more reliably produces habit than satiation. Devaluation taking place in the operant chamber more reliable than devaluation outside the operant chamber.	Parkes et al., 2016, Exp 3; Kosaki & Dickinson, 2010
------------------------------	---	--

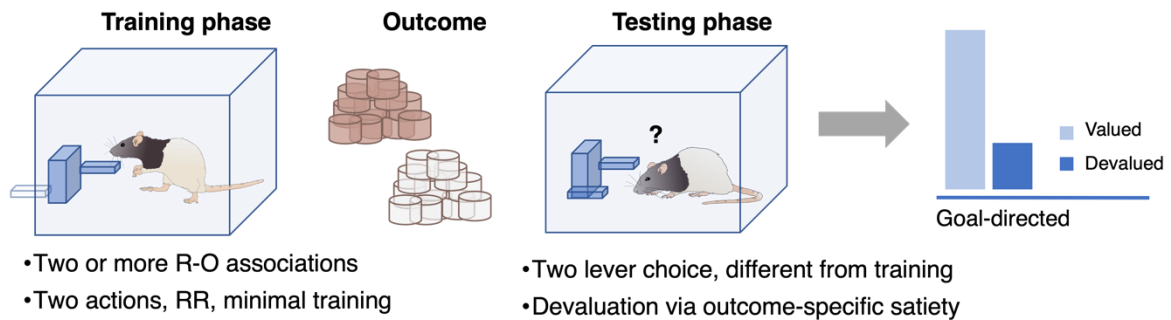
Another (not mutually exclusive) possibility is that with two-lever training only relatively weak S-R associations can be acquired because there is often no discriminative stimulus presented during instrumental learning protocols (cf. Kosaki & Dickinson, 2010). Response manipulanda are available (which can function as discriminative stimuli themselves), but there is often no discrete stimulus such as a light or a tone signalling which response is likely to be reinforced (Thrailkill et al., 2018; Vandaele et al., 2017). On the other hand, because there are discrete outcomes associated with each response, discrete O-R links are likely to be concurrently formed. Under these conditions, it is possible that the operant context acts as a general stimulus that becomes associated with both responses. The only difference may be the presence of either lever, minimising the discriminability of the two S-R associations and promoting the more easily differentiated R-O associations (see Figure 2B, training phase). Unlike discriminative training where discrete S-R associations are learned and can compete with discrete R-O associations for behaviour, this rather weak context-response association may not be sufficient to control responding at test. In line with this idea, Colwill (1994) showed that responses trained in the presence of discriminative stimuli interfered with carrying out other responses that were trained with the same outcome, but in a generic context (in the absence of discriminative stimuli). In the absence of a discriminative stimulus, the two separate and distinct R-O associations dominate action control and then animals do not readily develop habits under these conditions. Indirect evidence for this comes from animal studies that *did* report the development of behaviour that was insensitive to devaluation in a two-response choice scenario, possibly because they used discriminative stimuli during instrumental training (Faure et al., 2005, 2010). Vandaele et al (2017) examined this hypothesis and provided some evidence (albeit with only one lever) that with the same amount of training animals trained with discriminative stimuli acquired habits whereas those trained without such stimuli remained goal-directed.

Human studies tend to use more complex discriminations and, in contrast to the studies outlined above, always include (up to six) discriminative stimuli signalling which response is required to earn outcomes during training (cf. Dietrich et al., 2016; Sjoerds et al., 2016). Prior to the test phase, participants are told that responses directed towards devalued outcomes will incur a loss of points (i.e. instructed outcome devaluation). Particularly under speeded test conditions where decisions are made quickly, habitual 'slips of action', upon presentation of stimuli signalling devalued outcomes have been reported in numerous human studies (Dietrich et al., 2016; Sjoerds et al., 2016; Watson et al., 2018; see for review: Verhoeven & de Wit, 2018).

A) Putative Habits



B) Goal-directed Actions



C) Slips of Action

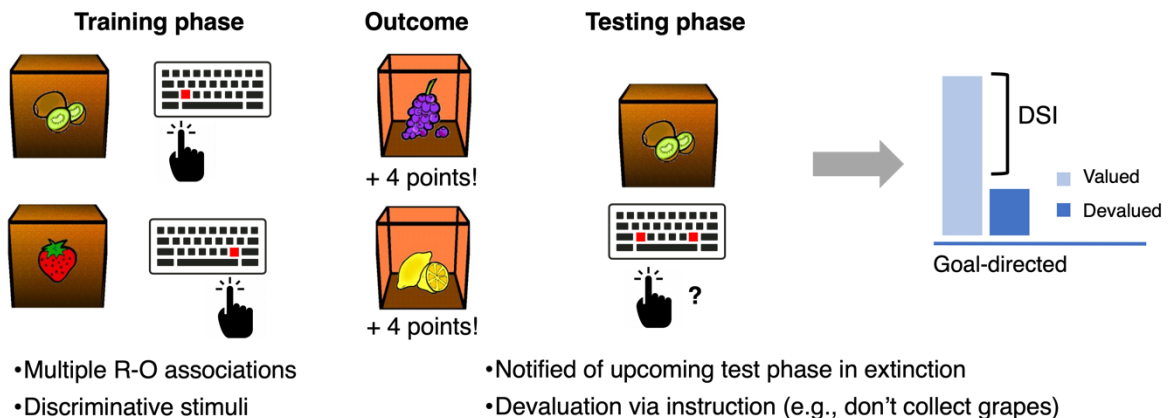


Figure 2. **A.** Experimental conditions commonly used for the development and measurement of habits. Insensitivity to goal-directed actions, including habits, in healthy animals is often generated by using a single action task, a random interval schedule, and overtraining. Devaluation is often performed via conditioned taste aversion. The test environment is generally the same as training (except under extinction conditions). A lack of significant difference between the valued and devalued responses indicates outcome devaluation insensitivity, which is often inferred as evidence for S-R habit. **B)** When aiming to measure goal-directed actions, tasks often incorporate more than one action-outcome association, train animals on a random ratio schedule, and provide sufficient training to learn the associations but not overtrained. Each action is often trained separately within a session such that only one lever is extended at any given time to help form distinct R-O associations. Devaluation is often performed via outcome-specific satiety as this is reversible and conducted within-subject. The devaluation test can be the first time the animal has been presented with two actions available, possibly generating detection of a change in context, which would hamper habit expression. Goal-directed control is confirmed when there is significantly more valued compared to

devalued responses. C) The slips-of-action task in humans uses up to six discriminative stimuli (fruits) that signal whether a right or left key presses will earn fruit outcomes (and points). Devaluation is via instruction – participants are told to stop earning some fruit outcomes in order to avoid losing points. Participants receive instructions that the test phase is in extinction signalling a change in context from training to test. The ability of participants to prevent erroneous responses directed towards stimuli signalling devalued outcomes is captured by the devaluation sensitivity index (DSI). This task reliably elicits slips-of-action towards devalued responses at test, however overtraining does not lead to increased slips-of-action as would be expected.

It is vexing that S-R habits only emerge in specific experimental conditions. However, as outlined above, there may be good theoretical reasons for this. Further research is needed to determine exactly why two action-outcome associations fail to come under habitual control in animals, even when using interval schedules or overtraining. It is likely that attention and cognition play a role such that having choice between two or more viable response options asserts goal-directed control at the choice point (Bouton, 2021; Dickinson, 1985; Kosaki & Dickinson, 2010; Thraillkill et al., 2018). Furthermore, where the goal is to measure habits, the use of discriminative stimuli during training could be incorporated into standard procedures, to support the learning of S-R associations. Considering how prolific habits are in everyday life, it is important that we devise and understand the laboratory conditions that reliably produce them.

6. Evidence for S-R associative links

In the previous sections we have assumed that both S-R and R-O associative links are formed during instrumental training, in line with associative dual-process models. In the case of R-O associations, there is considerable evidence that animals and humans learn to associate responses with the outcomes that they produce and use this information to select the appropriate course of action (Elsner & Hommel, 2004; Ostlund & Balleine, 2007; Pavlov, 1932; Rescorla, 1992). But the evidence for S-R associative links is more indirect (e.g., residual responding for devalued outcomes that occurs after outcome devaluation: Dickinson et al., 2002). This issue received considerable attention from associative learning theorists in previous decades (Colwill, 1994; Colwill & Rescorla, 1986). Recently, some authors have gone so far as to suggest that researchers should refrain from discussing S-R associative

links, and that it is hindering progress on habit research (De Houwer, 2019). One alternative to S-R associations is simply that R-O associations dominate action, and that behaviour that is characterised as 'habit' emerges because outcome devaluation does not affect behaviour in the way that researchers expect. That is, despite outcomes being devalued, it is impossible to know whether responses are carried out for some other means to an end. De Houwer (2019) gives the example of those who regularly eat popcorn in the cinema being more likely to eat stale popcorn when offered it in a cinema setting (a classic demonstration of habit from the field of Health Psychology; Neal et al., 2011). De Houwer suggests that it may not be that the stale (devalued) popcorn is eaten out of habit (triggered by the cinema context) - instead regular cinema goers may believe that popcorn (whether stale or not) rounds out the cinema experience and so are eating it in a purely goal-directed manner. According to this view, all choice behaviour is goal-directed (see e.g., Moors et al., 2017). A similar argument was advanced by Colwill and Rescorla (1986) who proposed that extensive instrumental training may render reinforcer devaluation ineffective. They also argued that different measurement techniques may be required to tap into goal-directed control, following extended instrumental training. Anecdotally however, we all experience habitual actions that are clearly not in line with our goals. For example, driving on autopilot to work in the weekend when you were intending to drive and visit a friend. Given that this unintended action is not driven by the current outcome (R-O associations), there must be an alternative system producing the behaviour.

Another alternative to S-R associations controlling behaviour when inappropriate responses are made is that confusion simply led to the incorrect R-O association being selected or that the animal is purposefully exploring other response options (Addicott et al., 2017). However, the mechanism driving alternative response selection (either in error or strategically) would surely need to involve context or stimuli in the environment. Hierarchical S: R-O theories do provide an account for this such that R-O associations are learned and can be triggered by contextual stimuli (Colwill & Rescorla, 1986; Rescorla, 1990, 1991). Recent evidence suggests

that both binary S-R associations and hierarchical S:R-O associations are learnt during instrumental training and potentially compete to control behaviour (Bradfield & Balleine, 2013; Trask & Bouton, 2014). However, evidence of S-R associations was inferred indirectly in these studies, through for example, residual responding for devalued reinforcers that occurred after outcome devaluation.

In 1994, Colwill used an elegant paradigm to obtain more direct evidence for S-R associative links. As briefly mentioned earlier, she demonstrated transfer of S-R associations that impacted on other instrumental responses. In the first phase, rats were trained with two discriminative stimuli (light and tone), two responses (nose poke and handle pulling), and one outcome (i.e., S1- R1-O1 and S2- R2-O1). Then two new R-O associations were trained without discriminative stim (lever press and chain pull) with sucrose (R3-O2) and polycose (R4-O3). Finally, R1 and R2 were each trained with O2 and O3 (in the absence of discriminative stimuli). At test the lever (R3) and chain pull (R4) were made available during presentation of the discriminative stimuli. Colwill reported that S1 selectively affected responding on R3 and S2 selectively affected responding on R4 even though R3 and R4 had not previously been trained in the presence of discriminative stimuli. The pattern of results could not be explained without appeal to S-R associative links that must have been established in the initial training phase of the experiment, such that the representation of S1-R1 affected responding of R3 (via the shared outcome representation of O2). A similar pattern of results was observed in humans, leading to suggestions that S-R associative links are also learned in human instrumental learning paradigms (Gámez & Rosas, 2007). However, beyond these studies, direct evidence for S-R links remains largely elusive.

Overtraining has long been assumed as a clear way to induce habits and show evidence for S-R links. Adams (1982) demonstrated that in a group of rats moderately-trained on a ratio schedule of reinforcement, R-O links were still dominating choice because the behaviour was sensitive to the current value of the outcome. However, after extensive training, rats just as

frequently carried out the response that earned the now devalued outcome. This finding has been interpreted to suggest that after extended training, S-R links dominated, and that behaviour was carried out regardless of the current value of the outcome (de Wit & Dickinson, 2009; Dickinson, 1985). However, this overtraining effect (with ratio schedules of reinforcement) has been difficult to replicate (Garr et al., 2021; see Table 2). In humans, whilst robust evidence for habitual ‘slips of actions’ has been observed, many studies have also shown that extended training does not lead to *more* slips of action (de Wit et al., 2018; Pool et al., 2021; see Table 3). This suggests that we are yet to clearly identify the experimental parameters that promote habitual behaviour. Dickinson (1985) suggested that habits are not simply due to overtraining, but that repetition encourages invariance of actions, a feature now considered cardinal of habits (Balleine & Dezfouli, 2019). Therefore any factor that reduces behavioural variation (and consequently the experienced behaviour-goal correlation) would also be expected to derive habits. In situations where overtraining does not lead to enhanced automaticity, then goal-directed control may remain dominant irrespective of how much training is given (see also: Camerer & Li, 2021).

These failures to find overtraining effects across all reinforcement schedules have been interpreted as an issue for the associative dual-process model – regardless of the schedule of reinforcement, strong S-R links should build up during training which should then lead to stronger habit tendencies at test. However, the failure to find overtraining effects might reflect the fact that while stronger S-R links are being built up with extended training, concurrent R-O links are also being strengthened at the same time. It is clear that repetition of behaviour alone is insufficient to induce habits and that factors such as predictability of the response-reinforcer relationship (and thereby attention towards these various components of behaviour) play a critical role in determining whether behaviour that is insensitive to reinforcer devaluation will emerge (Dickinson, 1985; Thrailkill et al., 2018). As mentioned earlier, many animal studies do not use discriminative training which may impair acquisition of distinct S-R associations. While human paradigms tend to always use

discriminative stimuli and thus should be optimal conditions for building S-R associations, the fact remains that on every training trial the S-R and R-O associations are both being strengthened. To date there is very limited experimental evidence to suggest that stronger associative links should form between S-R relative to R-O when both associations are being trained within the same trial.

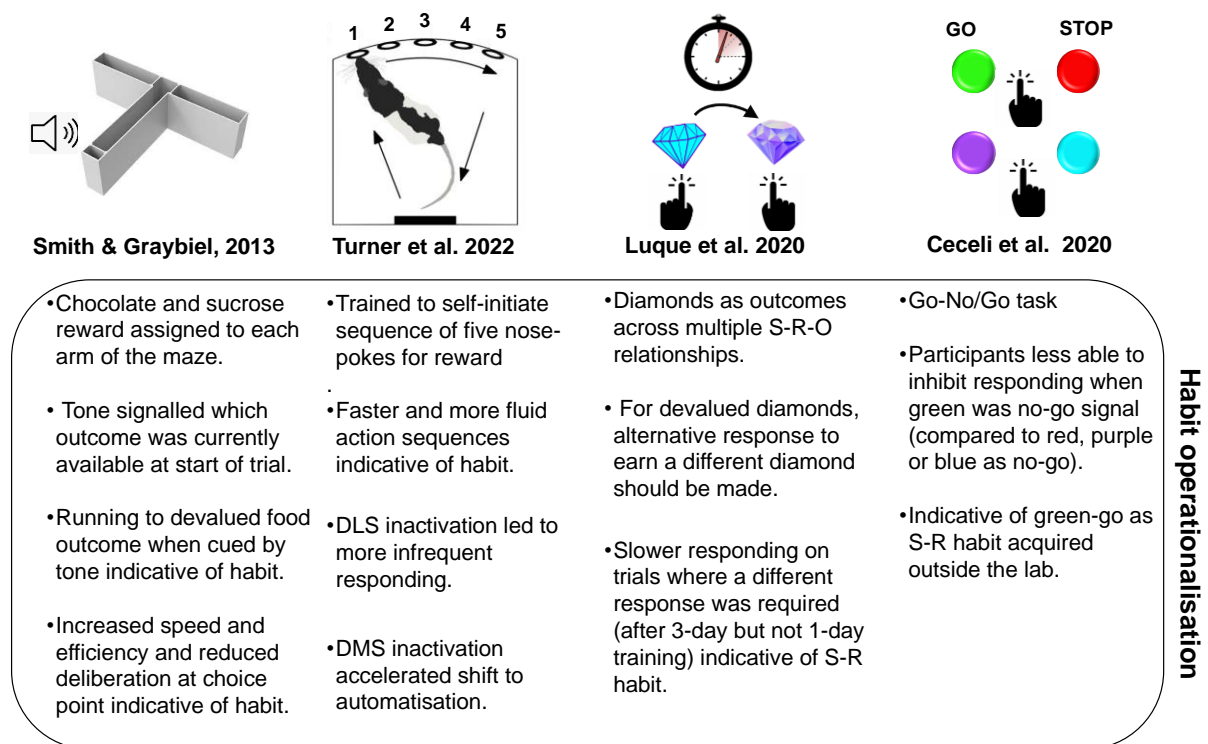


Figure 3. Some of the novel paradigms being used to measure other features of habit. DLS = dorsolateral striatum ; DMS = dorsomedial striatum.

One recent study demonstrated the ‘unmasking’ of stronger S-R associative links after overtraining, using a time-pressure manipulation with contingency reversal (Hardwick et al., 2019). Human participants learned to make a unique keyboard response for each of four discriminative stimuli, for which they received ‘correct/incorrect’ feedback. They had either limited training (performance to criterion, minimum of 20 trials) or extensive training (4000 trials over four days). During the test phase, the contingency between responses and ‘correct’ feedback was changed — two of the responses were swapped such that Stimulus 1 should now be followed by Response 3 rather than Response 1 (and vice versa for stimulus

3). The response preparation time was then manipulated in the test phase, with the authors showing that when participants were restricted to 300-600ms to prepare a response they were more likely to make the previously learned response for the remapped stimuli. At longer preparation times these slips-of-action no longer occurred, and participants chose the correct (remapped) response. Importantly these action slips under time pressure were only observed in the extended 4-day training condition as compared to the minimal training condition.

The results of Hardwick et al. suggest that stimuli can very quickly trigger a prepotent response which can be inhibited and replaced by the correct response, in less than one second (see also: Dezfouli & Balleine, 2012). It seems logical that habits can be rapidly overridden when cognitive resources are available to make better choices. This highlights that manipulating conditions during training and/or at test that allow S-R associations to dominate during learning/expression remains a critical area of research (Watson & de Wit, 2018).

Context changes from training to test arguably lead to recruitment of goal-directed control (in both humans and animals), that may mask any S-R learning (review: Bouton, 2021). In animal designs, interleaved single-lever training phases followed by a two-lever choice test after outcome devaluation could be a significant context change for the animal (see Table 2). This in turn may lead to a switch from habit to goal-directed control (although we note that many studies do not use the double-lever at test design yet still report goal-directed control after extended training; cf. Garr et al., 2020). Human studies tend to use more complex discriminations and often include (up to six) discriminative stimuli signalling which response is required to earn points during training (see figure 2C: Dietrich et al., 2016; Sjoerds et al., 2016). However, participants are aware that they are in a psychological experiment about learning and memory and are alerted to the fact that they are entering a new phase of the task (i.e. the test phase), which may function as a context switch,

reasserting goal-directed control processes. Stress inducement has already been suggested as a key manipulation to reduce goal-directed control (Pool et al., 2021; Schwabe & Wolf, 2009) as has time pressure (Hardwick et al., 2019). Another obvious manipulation would be to tax cognitive control through working memory manipulations (either during learning or at test). Under conditions where R-O associative processing cannot dominate, habitual S-R associations should govern behavioural control.

In summary, in the event that goal-directed control cannot dominate responding, S-R associations are a parsimonious way to account for observed behaviour. However, evidence that overtraining leads to stronger S-R links and eventual habits, regardless of training schedule, is still limited. Although there are reasonable theoretical explanations for this lack of evidence, it is critical that experiments are carried out using sophisticated experimental manipulations that identify the factors controlling the shift from goal-directed control to S-R habits.

7. Beyond the Standard Tests of Habit

Tests of instrumental choice responding after outcome devaluation remain a critical tool for assessing whether behaviour can meet the strict criteria for goal-directed action, as defined in the field of associative learning. However, it is often impossible to devalue outcomes that are associated with the many daily behaviours that we may recognise as habitual, such as going directly to the gym after work or brushing teeth before bed. As mentioned in the introduction, most researchers would agree that habits also have other defining features — such as the repetition of context (repeatedly happening at the same time and/or in the same place), the rapid speed at which they are initiated and executed, and the refinement and invariance of movement sequences or action chunks (Balleine & Dezfouli, 2019; Gardner, 2015; Wood & Runger, 2016). In recent years researchers have used novel approaches to studying habits that expand beyond comparing the number of responses made for devalued versus valuable outcomes in the outcome devaluation test, and focus instead on some of the positive features of habits outlined above (Ceceli et al., 2020; Luque et al., 2020; Turner et

al., 2022). While not a comprehensive list of every novel test posited to measure habits, here we highlight some of the diverse approaches being taken in the field (Figure 3)

One recent study in humans used reaction times on remapped versus consistent trials at test as a measure of habit (Luque et al., 2020; see Figure 3). Participants could earn diamonds of different value in the presence of discriminative stimuli, by carrying out one of two responses. On test blocks the value of the diamonds could change, meaning that the optimal response to make also changed. After three days (but not one day) of instrumental training, RTs were slower when participants had to carry out a different response to that which was learned during training (taken as evidence of interference from the previously learned response). It is notable that participants in the study of Luque et al. made very few ‘slips of action’ (i.e., overtly making the previously learned response) even after extended training. This data thus fits with that of Hardwick et al., (2019) demonstrating that the stimulus very quickly triggers a prepotent response which can then be inhibited if necessary (leading presumably to a slowing in response time rather than significant errors in the study of Luque et al.). This reinforces the idea that participants in a lab context remain in a state of goal-directed control and that manipulations designed to inhibit this control (e.g., response time pressure, stress induction, taxing working memory) may be required for reliable measurement of S-R habits. If the evolutionary purpose of habits is to reduce cognitive load by operating in the background, then it may well be that additional cognitive load (as is commonly experienced in real world examples) is required to relegate functions to habitual control.

The approach used by Luque et al. (2020) and the aforementioned study of Hardwick et al. (2019) was subtly different to the standard instrumental training procedure because there was no unique outcome associated with each S-R pairing. For example, in the commonly used slips-of-action test, a left key press might be associated with winning apples (in the

presence of a discriminative stimulus) and a right key press associated with winning bananas. During the test phase, those fruit outcomes become devalued via instruction such that participants should stop attempting to acquire them (cf. Sjoerds et al., 2016). By contrast, there were no unique outcomes in the study of Hardwick and colleagues- only correct/incorrect feedback was provided across multiple responses. Luque et al. (2020) used a more traditional design with instrumental S-R learning reinforced with outcomes (coloured diamonds), some of which were then devalued at test. However, during training, a particular diamond was common to multiple S-R mappings. For example, participants could earn a blue diamond during training by pressing R1 in the presence of S1 or R2 in the presence of S2. This 'common outcome' feature of the experimental designs of Luque and Hardwick may explain why these studies found evidence indicative of S-R habits following overtraining whereas studies using more traditional 'differential outcome' instrumental learning phases have not. When the same outcome is used across multiple R-O relationships, contingency learning is slower (for both humans and other animals) relative to when unique outcomes are used (Mok & Overmier, 2007; Trapold, 1970; see for review: Urquioli, 2005). The designs used by Hardwick and Luque, therefore, have the optimal conditions for R-O associations to develop more slowly than S-R associations, potentially allowing for the latter to dominate action control following extended training.

Ceceli et al., (2020) used a go/no-go task and capitalised on participant's experience with red being associated with an instrumental stopping response and green being associated with going (e.g., when driving; see Figure 3). This study reported that participants were better at inhibiting responding when red circles were the no-go signal (as opposed to when green, purple or blue functioned as no-go signals). Poorer performance on red-go trials relative to the other conditions was also found (Experiment 2). This study demonstrated the benefit and detriment conferred by overtrained habits where training was acquired over a lifetime, rather than in the lab. How much training and exposure participants had to the 'red-stop' and 'green-go' instrumental contingencies was therefore unable to be measured (and the

degree to which Pavlovian activation/inhibition could influence instrumental go and no-go responses is also not clear). Demonstrating the flexibility of behaviour to shift back and forth between habit and goal-directed control, Ceceli et al reported that performance impairments on green-stop trials (relative to e.g., red-stop trials) could be eradicated when participants received salient performance-contingent feedback and reward for correct responses.

A novel paradigm designed to explore the benefit rather than detriment of habitual responding was recently used by Turner and colleagues (Turner et al., 2022). Rats learned to make a sequence of five nose pokes for reward. The initiation of the action sequence was not cued, but self-initiated. Sessions were time limited meaning that faster and more fluid action sequences resulted in animals earning more rewards. The researchers found that pharmacological pre-training lesions of either the DMS or DLS had opposing effects on behaviour. As expected, given previous findings that habits rely on an intact DLS (Yin et al., 2004) rats with DLS lesions initiated less action sequences and completed them more slowly relative to DMS-lesioned rats. By contrast DMS lesions accelerated the shift towards automatization, suggesting that disengagement of the DMS allows the DLS to more readily take control of behaviour. This study highlights the potential benefits in not relying on outcome devaluation to allow habits to emerge but focusing on chunked action sequences that may be initiated by goal-directed processes and become fluid through repeated practice, to the benefit of the organism.

This follows extensive work from the laboratory of Graybiel untangling the role of the dorsal striatum in habit and skill learning, particularly in bracketing chunked action sequences (Graybiel, 1998; Graybiel & Grafton, 2015; Smith & Graybiel, 2013). By applying outcome devaluation and an extinction test on the rodent T-maze task, Smith and Graybiel (2013) demonstrated that animals persisted in navigating the maze to an outcome that had been devalued through taste aversion – but only after extended training. In the presence of a discriminative stimulus signalling its availability, rats trained to criterion reduced navigation

towards the devalued outcome – instead they turned in the opposite direction at the critical juncture (indicative of goal-directed control). Although the T-maze is not frequently used to investigate habits (because of the strong Pavlovian component inherent to the design), the authors' interpretation of habitual responding was also supported by other measures such as increased speed and efficiency (decreased deliberative head turning at the choice point). In a similar vein, Desrochers et al (2010) demonstrated that non-human primates performing a non-instructed eye tracking task for reward over many months developed highly repetitive eye gaze patterns. These optimal but repetitive patterns developed spontaneously and became 'habit-like', leading the authors to suggest an inbuilt drive for repetitiveness.

Defining habits as fluid and quickly deployed movement sequences overlaps with definitions of skills and procedural learning, which are seen by associative learning theorists as different behaviours and fields of research, distinct from habits. Indeed, in the absence of a demonstration that a behaviour is insensitive to outcome value (or decrement in R-O contingency), many would argue that behaviour cannot be termed a 'habit'. However, the brain regions implicated as being critical for habit-like behaviour in the aforementioned tasks, align with those identified as critical for habitual behaviour that is insensitive to outcome devaluation (Balleine & O'Doherty, 2010; Turner et al., 2022; see for further discussion: Graybiel & Grafton, 2015). Anecdotally, examples of human slips-of-action are often related to procedural behaviours – for example, accidentally typing an old password on a familiar webpage or accidentally taking a familiar route that leads somewhere other than where you were intending to go. As mentioned in the introduction, some theorists argue that habits should emerge from repetition alone and always be dissociated from outcome value (Miller et al., 2018, 2019). Direct experimental evidence for this notion (and the arbitrator that determines whether S-R habits control responding) is still required. Nonetheless, this raises the question of whether human habits are in fact simply well learned sequences of actions and blurs the line between habit and skill (see Du et al., 2022; Haith & Krakauer, 2018 for further discussion of this topic).

Finally, a recent study adapted the chained seeking-taking lever design (cf. Zapata et al., 2010) to examine whether animals would develop habitual behaviour after extended training (van Elzelingen et al., 2022). Rats were first trained to press the seeking lever, which was then retracted for three seconds before a taking lever was extended. Pressing the taking lever earned a food pellet. Training continued for ten weeks. Standard outcome devaluation testing demonstrated that seeking-lever presses were insensitive to changes in outcome value. The unique adaptation in this study was to include a weekly brief extinction test where both the seeking and taking levers were presented and the lever choice was assessed. Early in training the animals were more likely to select the taking lever (being the most optimal lever to gain access to reward), but then across the ten weeks of training, choice for the seeking lever steadily increased, which the authors attribute to a positive test for habit. Although this is an interesting paradigm, it does not address whether animals could also be selecting the seeking lever in a very strategic and goal-directed manner (based on their extensive seek-take lever training), which may be under quite different control in this choice situation compared to the outcome insensitivity demonstrated for seek lever presses when that was the only option available. The fact that animals made more head entries to the food magazine following taking rather than seeking lever presses aligns with the idea that the animals retained the seeking→taking→reward structure of the task at test and could have been acting in a goal-directed manner in their choice of the seeking lever when offered both levers at test.

However, the study of van Elzelingen et al. raises an interesting point about magazine entries following lever pressing in extinction. Although simple magazine entries during instrumental and Pavlovian tasks are often reduced following outcome devaluation (Holland, 1998), a recent study reported that when a magazine entry followed a lever press, pressing was insensitive to outcome devaluation whereas presses without a magazine check remained sensitive to devaluation and were reduced compared to responses on a valued lever (Halbout et al., 2019). That is, lever presses without a magazine check were reduced on the devalued

compared to valued lever, whereas press-magazine sequences were not. This suggests that lever presses without a magazine check remain under goal-directed control, whereas press-magazine sequences are insensitive to changes in outcome value. Therefore, despite overall press rates indicating rats were under goal-directed control, there was evidence of divergence of control within the behavioural sub-components. It is tempting to suggest that sub-components under S-R control, such as these press-magazine sequences, help to explain the residual levels of responding on the devalued option despite overall performance suggesting it is under goal-directed control (Dickinson et al., 2002). There may also be potential to exploit this lever-press-to-magazine-check sequence as an additional measure of habit in operant paradigms.

8. Conclusions and Outstanding Questions

Associative dual-process models of behaviour have provided a strong framework for the experimental analysis of the factors that can lead to the emergence of habit, defined by some as the absence of goal-directed behaviour, that is insensitive to R-O contingency degradation and/or outcome devaluation. Despite the strong neurobiological evidence for S-R habits and goal-directed control, there are several aspects of the dual-process framework that are underexplored. We have attempted to provide theoretical justification for why habits tend not to emerge in two-lever choice scenarios and why evidence for stronger S-R links following overtraining remains elusive. We argue that researchers need to design experiments that better support S-R learning at the expense of R-O by utilising discriminative stimuli during training in addition to the use of outcomes that are common to multiple responses. Manipulating conditions during training and/or at test to shift the balance from goal-directed control and allowing S-R associations to dominate learning/expression remains a critical area of research. Key to this is minimising changes in context from training to test and inhibiting goal-directed control during training and/or test through stress inducement, time pressure or working memory manipulations.

On balance it seems acceptable to assume that S-R associations of some form drive behaviour when goal-directed processes are otherwise engaged. However, more research is required to develop and assess hierarchical S: R-O theories (particularly in human research). The fact remains however, that classic tests of sensitivity to outcome devaluation or contingency degradation can only be used to confirm whether behaviour is goal-directed or not and further criteria are needed to positively identify habit (Balleine & Dezfouli, 2019).

Using a null result to infer evidence for habits is also statistically questionable. There are some ways around this issue, such as deciding as a field the threshold at which we declare some evidence for habit (see e.g., Alderson, 2004). Bayesian techniques that quantify the strength of evidence for the null could also be used more widely (cf. Quintana & Williams, 2018). In general, however, binary tests of habit at a group level neglect the diverse and dynamic nature of decision-making. We should aim to measure the relative balance between habits and goal-directed control in a more nuanced fashion, including showing the adaptive changes in responding over time in the extinction test. If we assume that responding during test reflects the relative balance of goal-directed and habitual control for a particular individual, then behaviour can be dimensionally graded on how goal-directed it is and this can be related to individual differences (as is common in the human slips-of-action literature; cf. Watson et al., 2018). It is notable in clinical research that the outcome devaluation test is sometimes applied as a test of intact goal-directed control with minimal claims about habit (cf. R. W. Morris et al., 2015; Perkes et al., 2022);

We need to recognise that traditional paradigms measuring outcome devaluation and contingency degradation capture specific, narrow features of habits – namely the tendency to carry out a behaviour that is no longer appropriate. These tasks work best with very simple instrumental responses and require high rates of responding in the absence of immediate reward to ensure that sufficient responses can be measured during the extinction test to perform meaningful statistical analyses on response rates. This leaves many behaviours

unable to be examined by this framework as the detection of extinction conditions leads to rapid deterioration of responding, and likely restores goal-directed control. Moreover, there are many other features of habits that are not captured by the key metrics from outcome devaluation/contingency degradation paradigms such as the speed at which actions are performed and the refined and invariant characteristics of movement sequences (Balleine & Dezfouli, 2019). Attempts are being made to develop novel behavioural tasks that tap into these positive features of habits, and this should be encouraged as should be tasks that are not designed to assess whether that behaviour is sensitive to outcome devaluation, but capture the definition of habits through other measures.

Habits probably overlap with skills and procedural learning in that they both involve chunking of action sequences (Banca et al., 2020; Du et al., 2022; Graybiel, 2008). Many of the habit examples from our daily life involve automated sequences of well-learned behaviours that are triggered by contexts (being in the car, being on the computer), despite the fact that the original goal (to go via the dry cleaners, to enter the correct login password) has never changed. Of course, the risk of calling any fluid behavioural repertoire 'habit' is that clarity on what exactly is under investigation and what associative structure underpins the behaviour may be lost. As mentioned in the introduction, there are no clear cut-offs for what defines a habit when measurements are by self-report or reaction time. However, on the flip side, we may discover a richer representation of what habits are, if we are willing to find commonalities (in terms of neurobiology and associative structures) across different types of habit research. Clearer terminology as to the type of habit under investigation may be required by researchers to ensure that others can assess at a glance what exactly is under investigation (e.g., devaluation-insensitive habits vs. procedural habits).

To date, most studies of habit have focused on the transition from goal-directed to habitual control (following e.g., overtraining or pharmacological manipulations). However, recent evidence from various labs suggests that behaviour can transition rapidly between these two

modes (Bouton et al., 2020; Ceceli et al., 2020; Gremel & Costa, 2013; Hardwick et al., 2019). The transition from habit back to goal-directed control is an important yet relatively unexplored piece of the puzzle. It seems reasonable to assume that habit may be the dominant mode of action control as we move through the familiar environment – for example, waking at the same time each day, performing the same set of actions and heading out the door to work. But when changes or errors are detected, we can rapidly deploy goal-directed resources as and where they are needed. Perhaps under simplified and novel testing conditions this happens more often than in complex and well-rehearsed real-world scenarios, hence reducing the reliance on habits or increasing sensitivity to task goals and structure during experiments. Difficulties in switching from habits to goal-directed control may be relevant for understanding some features of psychopathology for example, the inability to override maladaptive habits in OCD (see for review: Robbins et al., 2019). It is likely that the switch back to goal-directed control from habit requires definition of an additional error detection system, which makes theoretical models more complex – but we need to consider that the transition from habits *back* to goal-directed control may be just as important as the well-researched transition from goal-directed control to habit.

In summary, this review highlights the remarkable insights that have stemmed from research conducted under the associative learning framework and is not intended to dishearten associative learning theorists in their pursuit of habits. However, we aimed to point out some of the challenges that emerge when defining habits experimentally. We encourage researchers to broaden the tools they use in their search for habits – in the hope that we might capture in the laboratory more of the rich diversity associated with habits in the real world.

Table 2: Rodent studies of habit referred to in this review

Reference	Demonstrating differences between goal-directed actions and habit	Instrumental training	Instrumental outcomes	Schedule of reinforcement	Outcome devaluation	Extinction test following outcome devaluation	Sensitivity to outcome devaluation (indicative of goal-directed control) vs. insensitivity (indicative of habits).
Yin et al., 2005	Neurobiological dissociations	Left/right lever. Trained in separate sessions	Punch and food pellets	VR	Satiety	Two levers	Sham animals showed sensitivity. Rats with lesions to posterior DMS did not.
Yin et al., 2004	Neurobiological dissociations	Single lever	Sucrose	VI	LiCi paired/unpaired	Single lever	Sham animals showed insensitivity. Rats with lesions to DLS remained sensitive.
Lingawi & Balleine, 2012	Neurobiological dissociations	Single lever	Sucrose	VI	LiCi paired/unpaired	Single lever	Sham animals and those with lesions to posterior CeA showed insensitivity. Rats with lesions to anterior CeA remained sensitive.
Ostlund & Balleine, 2008	Neurobiological dissociations	Left/right lever. Trained in separate sessions	Sucrose and food pellets	VR	Satiety	Two levers	Sham animals and those with lesions to thalamus showed sensitivity. Rats with lesions to BLA did not.
Killcross & Coutureau, 2003	Schedules of reinforcement (Within)/ Neurobiological dissociations (between)	Left/right lever. Trained in separate sessions	Sucrose and food pellets	One lever trained extensively VI, the other trained less on VR.	Satiety (repeated for each outcome)	Single lever (repeated for each lever).	Sham rats showed sensitivity on the low trained lever and insensitivity on the high trained. ILC lesions led to sensitivity on both. PLC lesions led to insensitivity on both.
Gremel & Costa, 2013 [*mice]	Schedules of reinforcement (Within)/ Neurobiological dissociations (between)	Left/right lever. Trained in separate contexts.	One outcome per animal - either pellets OR sucrose.	VI in one context, VR in the other	Satiety	Single lever (separate contexts)	Sham animals showed sensitivity but only on the response trained under the VR schedule. DMS lesions led to insensitivity for both. DLS lesions led to sensitivity for both.

Dickinson et al., 1983	Schedules of reinforcement	Single lever	Sucrose pellet	Half rats had VI. Half had VR	LiCi paired/unpaired	Single lever	Insensitivity only in the rats trained on VI.
Thraill et al., 2018, Exp 4	Schedules of reinforcement	Single lever, reinforced during discriminative stimulus (tone)	Food pellets	VI30 – for one group it was always the first lever press in any 30 second bin that was reinforced (CRF)	LiCi paired/unpaired	Single lever, discriminative stimulus	Group CRF showed insensitivity and group partial did not.
Garr et al., 2020, Exp. 3	Schedules of reinforcement	Single lever	Food pellet	either relatively-dense RI 10-s or a relatively lean RI 45-s schedule.	Satiety	Single lever	Rats trained on dense schedule showed sensitivity, rats trained on lean schedule did not.
Bouton et al., 2020, Exp. 1	Shift in context during training	Single lever	Food pellets	Group 1: 13 sessions VI Group 2: 12 sessions VI, 1 session trained with different reinforcer (VI)	LiCi paired/unpaired	Single lever	Group 1 show insensitivity. Group 2 show sensitivity.
Trask et al., 2020, Exp. 1.	Shift in context during training	Lever OR chain pull in context A. On last 4 days rats exposed to context B, half the rats received additional instrumental training of other response (R2) in context B.	Sucrose pellets	VI	LiCi paired/unpaired	Context A response (lever or chain pull) tested in Context A.	Insensitivity in the group that had not been trained on R2, Sensitivity in group trained in R2.

Trask et al., 2020, Exp. 4	Shift in context during training	Lever OR chain pull in context A. On last 4 days half the rats received additional instrumental R2 training of other response in context B. Other half received non-contingent reinforcer in context B.	Sucrose pellets	VI	LiCi paired/unpaired	Context A response (lever or chain pull) tested in context A.	Both contingent and non-contingent groups showed sensitivity.
Bouton et al., 2020	Shift in context during test	Single lever	Grain OR Sucrose pellets	VI	LiCi paired/unpaired	Single lever. Half group given pre-feeding on other outcome prior to test	Group given pre-feeding on other outcome show sensitivity, others show insensitivity.
Colwill & Rescorla 1985a	Differences in reinforcer type/ Two lever training effects	Lever pressing and chain pulling Separate sessions.	Pellets and sucrose	VI	LiCi paired with one outcome	Both responses available	Sensitivity to outcome devaluation. Sucrose and pellets produced different rates of responding during test.
Vandaele et al., 2017	Differences in reinforcer type/ possible role of discriminative stimuli	Single lever. Discrete trial procedure (start of trial indicated by lever retraction).	Grain or sucrose pellets	FR	Satiety	Single lever. Discrete trial procedure.	After 1 session sensitivity, after 6 sessions insensitivity. Interaction with food reward – response trained with pellets remained sensitive after 43 days of training while sucrose-response developed insensitivity by day 17.
Faure et al., 2005	Effect of response type/ Possible role of discriminative stimuli	Discriminative stimuli (light or tone) signalled whether chain pull or lever press would	Sucrose and food pellets	VI	Satiety	Both levers available at test, both stimuli presented intermixed.	For sham rats, lever press responding was insensitive but the chain pull remained sensitive.

		be reinforced. Trained separately.					
Faure et al., 2010	Possible role of discriminative stimuli	Left/right lever press only reinforced in light or tone. Trained separately.	Sucrose and food pellets	VI	Satiety	Both levers available at test, both stimuli presented intermixed	Sham group showed insensitivity.
Colwill & Rescorla 1985a	Overtraining effects/Two lever training effects	Lever pressing and chain pulling Separate sessions. One response trained more than other.	Pellets and sucrose	VI.	LiCi paired with one outcome.	Two extinction tests, each with one response available.	Remained sensitive, regardless of training length.
Adams, 1982 Exp 1	Overtraining effects	Single lever 100 lever presses vs. 500 lever presses	Food pellet	CRF	LiCi paired/unpaired	Single lever	Animals trained on 100 trials showed sensitivity, insensitivity in the group trained on 500 trials.
Garr et al., 2021	Overtraining effects	One lever. 2 or 10 training sessions	Food pellet	CRF	Satiety and LiCi	Single lever	Remained sensitive, regardless of training length.
Garr et al., 2020, Exp 4b	Overtraining effects/GD emerges following habit expression	One lever. For either 2 or 20 sessions	Food pellet	RI	Satiety	Single lever	After 2 days insensitive, after 20 days sensitive
Dezfouli et al., 2014, Exp 2	GD emerges following habit expression	Single lever	Sucrose	VI	LiCi paired/unpaired	Single Lever	For first 5 mins of extinction training responding was insensitive. But then sensitivity to outcome value emerged for last 5 minutes
Parkes et al., 2016, Exp 3	Context of outcome devaluation	Left and right levers trained separately within same chamber	Grain and sugar pellets	VR	Satiation either in operant chamber or in different chamber	5 hrs after test	Rats sated inside operant chamber showed sensitivity to OD whereas the rats who were sated outside the chamber showed insensitivity.
Kosaki & Dickinson, 2010, Exp 1	Context of outcome devaluation/	One group concurrently trained on two	Pellets and sucrose	VI	LiCi outside chamber (first test) followed by LiCi	Two lever test or one lever test	First test: All showed sensitivity. Second test: single lever training resulted in

	Single vs. two lever training	levers. Other group received single lever training with non-contingent presentation of other outcome			inside chamber (second test)	(depending on training)	insensitivity, two-lever training resulted in sensitivity.
Parkes et al., 2016, Exp. 1	Time course of satiation effects	Left and right levers trained separately within same chamber	Grain and sugar pellets	VR	Satiation	Immediate, 2hrs or 5 hrs test post-satiation	Only rats tested 0 or 2 hours after satiation showed sensitivity. But all three groups showed preference for the valuable over devalued in a consumption test.
Halbout et al., 2019	Alternative habit measure	Two levers, separate training	Grain and sucrose pellets	VI	Satiety	Single lever	Rats were more likely to attempt to retrieve reward (under extinction conditions) following a press on the devalued relative to the valuable lever.
van Elzelingen et al., 2022	Alternative habit measure	Seeking lever reinforced by retraction of the taking lever	Food pellet	Taking lever trained on VI schedule, taking lever always FR1.	-	Both levers available	Preference for seeking over taking lever is taken as test of habit.
Turner et al., 2022	Alternative habit measure	Rats learned to make a sequence of five nose pokes for reinforcer.	Food Pellet	CRF, self initiated.	-	Nose poke holes	Invariance of motor sequences observed across blocks

Table Note: All studies involve rats other than as indicated by * (mice). DLS = dorsolateral striatum ; DMS = dorsomedial striatum; CeA = central nucleus Amygdala; BLA = basolateral amygdala; ILC = Infralimbic cortex; PLC = Prelimbic cortex; VI = variable interval, VR = variable ratio, CRF = continuous reinforcement, FR = fixed ratio, LiCl = Lithium chloride

Table 3: Human studies of habit referred to in this review

Reference	Demonstrating differences between goal-directed actions and habit	Instrumental training	Instrumental outcomes	Schedule of reinforcement	Outcome devaluation	Extinction test following outcome devaluation	Results: Sensitivity to outcome devaluation (indicative of goal-directed control) vs. insensitivity (indicative of habits).
Dietrich et al., 2014	Individual differences	Concurrent discrimination training. Six discriminative stimuli. Two response keys.	Six animal pictures (and points)	CRF	Instructed devaluation, inhibition of response required in presence of some stimuli	Discriminative stimuli presented.	Degree of sensitivity correlated with self-reported sensation seeking.
Sjoerds et al., 2016	Individual differences	Concurrent discrimination training. Six discriminative stimuli. Two response keys.	Six animal pictures (and points)	CRF	Instructed devaluation, inhibition of response required in presence of some stimuli	Discriminative stimuli presented.	Degree of sensitivity correlated with model-based parameter from model-based/model-free task.
Reber et al., 2017	Neural correlates	Two discriminative stimuli, two responses	Two snack Foods	VR	Satiation	Discriminative stimuli presented.	Individuals with vmPFC damage showed insensitivity, healthy controls showed sensitivity.
Watson et al., 2018	Neural correlates/Role of cognitive control at test	Concurrent discrimination training. Six discriminative stimuli. Two response keys.	Six fruit pictures (and points)	CRF	Instructed devaluation	Discriminative stimuli presented.	Brain regions associated with cognitive control recruited on trials where stimulus is signalling a devalued outcome (requiring response inhibition).

Schwabe & Wolf, 2009	Role of cognitive control at test	Three discriminative stimuli, two responses	Flavoured drinks	VR	Satiation	Discriminative stimuli presented. Half of participants did the socially evaluated cold pressor test) before test.	Control group showed sensitivity. Stress group showed insensitivity.
de Wit et al., 2018, Exp. 2	Overtraining	Concurrent discrimination training. Six discriminative stimuli. Two response keys. Some S-R-O relationships trained more than others (within subjects)	Six fruit pictures (and points)	CRF	Instructed devaluation	Discriminative stimuli presented	Sensitivity towards devalued outcomes, regardless of training length
de Wit et al., 2018, Exp. 3	Overtraining	Concurrent discrimination training. Two discriminative stimuli, two responses. Short vs. long training (between groups)	Two snack Foods	VI	Satiation	Discriminative stimuli presented.	Sensitivity towards devalued outcomes, regardless of training length
Pool et al., 2021	Overtraining effects	Concurrent discrimination training. Two discriminative stimuli, two responses. Short	Two snack foods	VI	Satiation	Discriminative stimuli presented.	Both short and long training groups showed sensitivity.

		vs. long training (between groups)					
Luque et al., 2020, Exp 2.	Overtraining effects/novel test of habits	Four stimuli, two responses. Short vs. long training (between groups)	Three diamonds (and points)	VR	Instructed devaluation, different response required in presence of some stimuli	Discriminative stimuli presented.	Slower RT when a different response was required in the presence of each stimulus, but only after extended training.
Hardwick et al., 2019	Overtraining effects/novel test of habits	Four discriminative stimuli, four responses. Short vs. long training (between groups)	Correct vs incorrect feedback	CRF	Two responses swapped at test.	Discriminative stimuli presented	Slips-of-action observed when response window 300-600ms, only for the group who received extensive training.
Neal et al., 2011	Habits in real world	Relied on movie theatre context being associated with popcorn eating response.	Popcorn	-	Participants given either fresh or stale popcorn	Amount of popcorn eaten in the movie theatre vs. meeting room contexts (not tested in extinction)	Participants who habitually ate popcorn at the movies were more likely to eat the stale popcorn, but this did not occur in the meeting room context.
Ceceli et al., 2020, Exp 1.	Habits in real world	Relied on previous associations of green-go, red-stop.	Correct/incorrect feedback	-	Swapped responses (go-red, stop-green)	Go/No-go task	No go accuracy was better on red, relative to green, blue or purple. Go accuracy not affected.

Table Note: VI = variable interval, VR = variable ratio, CRF = continuous reinforcement

REFERENCES

Abiero, A. R., & Bradfield, L. A. (2021). The contextual regulation of goal-directed actions.

Current Opinion in Behavioral Sciences, *41*, 57–62.

<https://doi.org/10.1016/j.cobeha.2021.03.022>

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer

devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *34*(2), 77–

98. <https://doi.org/10.1080/14640748208400878>

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A Primer

on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research.

Neuropsychopharmacology, *42*(10), 1931–1939.

<https://doi.org/10.1038/npp.2017.108>

Alderson, P. (2004). Absence of evidence is not evidence of absence. *BMJ : British Medical*

Journal, *328*(7438), 476–477.

Amaya, K. A., Stott, J. J., & Smith, K. S. (2020). Sign-tracking behavior is sensitive to outcome

devaluation in a devaluation context-dependent manner: Implications for analyzing

habitual behavior. *Learning & Memory*, *27*(4), 136–149.

<https://doi.org/10.1101/lm.051144.119>

Antoniuk, S., Bijata, M., Ponimaskin, E., & Włodarczyk, J. (2019). Chronic unpredictable mild

stress for modeling depression in rodents: Meta-analysis of model reliability.

Neuroscience & Biobehavioral Reviews, *99*, 101–116.

<https://doi.org/10.1016/j.neubiorev.2018.12.002>

- Balleine, B. W. (2005). Neural bases of food-seeking: Affect, arousal and reward in corticostriatolimbic circuits. *Physiology & Behavior*, *86*(5), 717–730.
<https://doi.org/10.1016/j.physbeh.2005.08.061>
- Balleine, B. W. (2019). The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. *Neuron*, *104*(1), 47–62. <https://doi.org/10.1016/j.neuron.2019.09.024>
- Balleine, B. W., & Dezfouli, A. (2019). Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits. *Frontiers in Psychology*, *10*.
<https://doi.org/10.3389/fpsyg.2019.02735>
- Balleine, B. W., & O’Doherty, J. P. (2010). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*, *35*(1), 48–69. <https://doi.org/10.1038/npp.2009.131>
- Banca, P., McNamee, D., Piercy, T., Luo, Q., & Robbins, T. W. (2020). A Mobile Phone App for the Generation and Characterization of Motor Habits. *Frontiers in Psychology*, *10*.
<https://www.frontiersin.org/article/10.3389/fpsyg.2019.02850>
- Berridge, K. C. (2021). Comment on Vandaele and Ahmed: Rethinking habits in addiction. *Neuropsychopharmacology*, *46*(4), 687–688. <https://doi.org/10.1038/s41386-020-00932-0>
- Bouton, M. E. (2021). Context, attention, and the switch between habit and goal-direction in behavior. *Learning & Behavior*. <https://doi.org/10.3758/s13420-021-00488-z>
- Bouton, M. E., Broomer, M. C., Rey, C. N., & Thrailkill, E. A. (2020). Unexpected food outcomes can return a habit to goal-directed action. *Neurobiology of Learning and Memory*, *169*, 107163. <https://doi.org/10.1016/j.nlm.2020.107163>
- Bradfield, L. A., & Balleine, B. W. (2013). Hierarchical and binary associations compete for behavioral control during instrumental biconditional discrimination. *Journal of*

Experimental Psychology. Animal Behavior Processes, 39(1), 2–13.

<https://doi.org/10.1037/a0030941>

Bradfield, L. A., Leung, B. K., Boldt, S., Liang, S., & Balleine, B. W. (2020). Goal-directed actions transiently depend on dorsal hippocampus. *Nature Neuroscience*, 23(10), 1194–1197. <https://doi.org/10.1038/s41593-020-0693-8>

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>

Camerer, C. F., & Li, X. (2021). Neural autopilot and context-sensitivity of habits. *Current Opinion in Behavioral Sciences*, 41, 185–190. <https://doi.org/10.1016/j.cobeha.2021.07.002>

Ceceli, A. O., Myers, C. E., & Tricomi, E. (2020). Demonstrating and disrupting well-learned habits. *PLOS ONE*, 15(6), e0234424. <https://doi.org/10.1371/journal.pone.0234424>

Colwill, R. M. (1994). Associative Representations of Instrumental Contingencies. In *Psychology of Learning and Motivation* (Vol. 31, pp. 1–72). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60408-9](https://doi.org/10.1016/S0079-7421(08)60408-9)

Colwill, R. M., & Rescorla, R. A. (1985a). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *Journal of Experimental Psychology: Animal Behavior Processes*, 11(4), 520–536. <https://doi.org/10.1037/0097-7403.11.4.520>

Colwill, R. M., & Rescorla, R. A. (1985b). 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. (1986). Associative Structures In Instrumental Learning. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 20, pp. 55–104). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60016-X](https://doi.org/10.1016/S0079-7421(08)60016-X)
- Daw, N., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. <https://doi.org/10.1038/nn1560>
- De Houwer, J. (2019). On How Definitions of Habits Can Complicate Habit Research. *Frontiers in Psychology*, *10*, 2642. <https://doi.org/10.3389/fpsyg.2019.02642>
- de Wit, S., & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal–human translational models. *Psychological Research*, *73*(4), 463–476. <https://doi.org/10.1007/s00426-009-0230-6>
- de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., & Gillan, C. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology. General*, *147*(7), 1043–1065. <https://doi.org/10.1037/xge0000402>
- de Wit, S., Standing, H. R., DeVito, E. E., Robinson, O. J., Ridderinkhof, K. R., Robbins, T. W., & Sahakian, B. J. (2012). Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology*, *219*(2), 621–631. <https://doi.org/10.1007/s00213-011-2563-2>
- Delorme, C., Salvador, A., Valabrègue, R., Roze, E., Palminteri, S., Vidailhet, M., de Wit, S., Robbins, T., Hartmann, A., & Worbe, Y. (2016). Enhanced habit formation in Gilles de

la Tourette syndrome. *Brain: A Journal of Neurology*, 139(Pt 2), 605–615.

<https://doi.org/10.1093/brain/awv307>

Desrochers, T. M., Jin, D. Z., Goodman, N. D., & Graybiel, A. M. (2010). Optimal habits can develop spontaneously through sensitivity to local cost. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47), 20512–20517.

<https://doi.org/10.1073/pnas.1013470107>

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences, and reinforcement learning. *The European Journal of Neuroscience*, 35(7), 1036–1051.

<https://doi.org/10.1111/j.1460-9568.2012.08050.x>

Dezfouli, A., & Balleine, B. W. (2013). Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized. *PLoS Computational Biology*, 9(12). <https://doi.org/10.1371/journal.pcbi.1003364>

Dezfouli, A., Lingawi, N. W., & Balleine, B. W. (2014). Habits as action sequences: Hierarchical action control and changes in outcome value. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1655).

<https://doi.org/10.1098/rstb.2013.0482>

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1135), 67–78.

Dickinson, A. (1994). Instrumental Conditioning. In N. J. Mackintosh (Ed.), *Animal Learning and Cognition*. Elsevier Science.

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

Experimental Psychology Section B, 35(1), 35–51.

<https://doi.org/10.1080/14640748308400912>

Dickinson, A., Wood, N., & Smith, J. W. (2002). Alcohol seeking by rats: Action or habit? *The Quarterly Journal of Experimental Psychology Section B*, 55(4), 331–348.

<https://doi.org/10.1080/0272499024400016>

Dietrich, A., de Wit, S., & Horstmann, A. (2016). General Habit Propensity Relates to the Sensation Seeking Subdomain of Impulsivity But Not Obesity. *Frontiers in Behavioral Neuroscience*, 10. <https://doi.org/10.3389/fnbeh.2016.00213>

Du, Y., Krakauer, J. W., & Haith, A. M. (2022). The relationship between habits and motor skills in humans. *Trends in Cognitive Sciences*, 0(0).

<https://doi.org/10.1016/j.tics.2022.02.002>

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research*, 68(2–3), 138–154. <https://doi.org/10.1007/s00426-003-0151-8>

Faure, A., Haberland, U., Condé, F., & El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *Journal of Neuroscience*, 25(11), 2771–2780. <https://doi.org/10.1523/JNEUROSCI.3894-04.2005>

Faure, A., Leblanc-Veyrac, P., & El Massioui, N. (2010). Dopamine agonists increase perseverative instrumental responses but do not restore habit formation in a rat model of Parkinsonism. *Neuroscience*, 168(2), 477–486.

<https://doi.org/10.1016/j.neuroscience.2010.03.047>

Gámez, A. M., & Rosas, J. M. (2007). Associations in human instrumental conditioning. *Learning and Motivation*, 38(3), 242–261.

<https://doi.org/10.1016/j.lmot.2006.11.001>

- Gardner, B. (2015). A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. *Health Psychology Review, 9*(3), 277–295. <https://doi.org/10.1080/17437199.2013.876238>
- Gardner, B., de Bruijn, G.-J., & Lally, P. (2011). A systematic review and meta-analysis of applications of the Self-Report Habit Index to nutrition and physical activity behaviours. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine, 42*(2), 174–187. <https://doi.org/10.1007/s12160-011-9282-0>
- Garr, E., Link to external site, this link will open in a new window, Bushra, B., Tu, N., & Delamater, A. R. (2020). Goal-directed control on interval schedules does not depend on the action–outcome correlation. *Journal of Experimental Psychology: Animal Learning and Cognition, 46*(1), 47–64. <http://dx.doi.org/10.1037/xan0000229>
- Garr, E., Padovan-Hernandez, Y., Janak, P., & Delamater, A. R. (2021). *Maintained goal-directed control with overtraining on ratio schedules*. PsyArXiv. <https://doi.org/10.31234/osf.io/hp42x>
- Gillan, C. M., Pappmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., & de Wit, S. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *The American Journal of Psychiatry, 168*(7), 718–726. <https://doi.org/10.1176/appi.ajp.2011.10071062>
- Graybiel, A. M. (1998). The Basal Ganglia and Chunking of Action Repertoires. *Neurobiology of Learning and Memory, 70*(1), 119–136. <https://doi.org/10.1006/nlme.1998.3843>
- Graybiel, A. M. (2008). Habits, Rituals, and the Evaluative Brain. *Annual Review of Neuroscience, 31*(1), 359–387. <https://doi.org/10.1146/annurev.neuro.29.051605.112851>

- Graybiel, A. M., & Grafton, S. T. (2015). The Striatum: Where Skills and Habits Meet. *Cold Spring Harbor Perspectives in Biology*, 7(8), a021691.
<https://doi.org/10.1101/cshperspect.a021691>
- Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, 4(1), 2264. <https://doi.org/10.1038/ncomms3264>
- Haith, A. M., & Krakauer, J. W. (2018). The multiple effects of practice: Skill, habit and reduced cognitive load. *Current Opinion in Behavioral Sciences*, 20, 196–201.
<https://doi.org/10.1016/j.cobeha.2018.01.015>
- Halbout, B., Marshall, A. T., Azimi, A., Liljeholm, M., Mahler, S. V., Wassum, K. M., & Ostlund, S. B. (2019). Mesolimbic dopamine projections mediate cue-motivated reward seeking but not reward retrieval in rats. *eLife*, 8, e43551.
<https://doi.org/10.7554/eLife.43551>
- Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, 3(12), 1252–1262. <https://doi.org/10.1038/s41562-019-0725-0>
- Holland, P. (1998). Amount of training affects associatively-activated event representation. *Neuropharmacology*, 37(4), 461–469. [https://doi.org/10.1016/S0028-3908\(98\)00038-0](https://doi.org/10.1016/S0028-3908(98)00038-0)
- Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science*, 4(6), 533–550.
- Killcross, S., & Coutureau, E. (2003). Coordination of Actions and Habits in the Medial Prefrontal Cortex of Rats. *Cerebral Cortex*, 13(4), 400–408.
<https://doi.org/10.1093/cercor/13.4.400>

- Kosaki, Y., & Dickinson, A. (2010). Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(3), 334–342. <https://doi.org/10.1037/a0016887>
- Kuhn, B. N., Kalivas, P. W., & Bobadilla, A.-C. (2019). Understanding Addiction Using Animal Models. *Frontiers in Behavioral Neuroscience*, *13*, 262. <https://doi.org/10.3389/fnbeh.2019.00262>
- Lingawi, N. W., & Balleine, B. W. (2012). Amygdala Central Nucleus Interacts with Dorsolateral Striatum to Regulate the Acquisition of Habits. *Journal of Neuroscience*, *32*(3), 1073–1081. <https://doi.org/10.1523/JNEUROSCI.4806-11.2012>
- Lingawi, N. W., Dezfouli, A., & Balleine, B. W. (2016). The Psychological and Physiological Mechanisms of Habit Formation. In R. A. Murphy & R. C. Honey (Eds.), *The Wiley Handbook on the Cognitive Neuroscience of Learning* (pp. 409–441). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118650813.ch16>
- Luque, D., Molinero, S., Watson, P., López, F. J., & Le Pelley, M. E. (2020). Measuring habit formation through goal-directed response switching. *Journal of Experimental Psychology: General*, *149*(8), 1449–1459. <https://doi.org/10.1037/xge0000722>
- Melnikoff, D. E., & Bargh, J. A. (2018). The Mythical Number Two. *Trends in Cognitive Sciences*, *22*(4), 280–293. <https://doi.org/10.1016/j.tics.2018.02.001>
- Miller, K. J., Ludvig, E. A., Pezzulo, G., & Shenhav, A. (2018). Chapter 18—Realigning Models of Habitual and Goal-Directed Decision-Making. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-Directed Decision Making* (pp. 407–428). Academic Press. <https://doi.org/10.1016/B978-0-12-812098-9.00018-8>
- Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without Values. *Psychological Review*, *126*(2), 292–311. <https://doi.org/10.1037/rev0000120>

- Mok, L. W., & Overmier, J. B. (2007). The Differential Outcomes Effect in Normal Human Adults Using a Concurrent-Task Within-Subjects Design And Sensory Outcomes. *The Psychological Record*, *57*(2), 187–200. <https://doi.org/10.1007/BF03395571>
- Moors, A., Boddez, Y., & De Houwer, J. (2017). The Power of Goal-Directed Processes in the Causation of Emotional and Other Actions. *Emotion Review*, *9*(4), 310–318. <https://doi.org/10.1177/1754073916669595>
- Morris, R. W., Quail, S., Griffiths, K. R., Green, M. J., & Balleine, B. W. (2015). Corticostriatal Control of Goal-Directed Action Is Impaired in Schizophrenia. *Biological Psychiatry*, *77*(2), 187–195. <https://doi.org/10.1016/j.biopsych.2014.06.005>
- Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives? *Personality & Social Psychology Bulletin*, *37*(11), 1428–1437. <https://doi.org/10.1177/0146167211419863>
- Ostlund, S. B., & Balleine, B. W. (2007). Selective reinstatement of instrumental performance depends on the discriminative stimulus properties of the mediating outcome. *Learning & Behavior*, *35*(1), 43–52. <https://doi.org/10.3758/BF03196073>
- Ostlund, S. B., & Balleine, B. W. (2008). Differential Involvement of the Basolateral Amygdala and Mediodorsal Thalamus in Instrumental Action Selection. *Journal of Neuroscience*, *28*(17), 4398–4405. <https://doi.org/10.1523/JNEUROSCI.5472-07.2008>
- Parkes, S. L., Marchand, A. R., Ferreira, G., & Coutureau, E. (2016). A time course analysis of satiety-induced instrumental outcome devaluation. *Learning & Behavior*, *44*(4), 347–355. <https://doi.org/10.3758/s13420-016-0226-1>
- Pavlov, I. P. (1932). The Reply of a Physiologist to Psychologists. *Psychological Review*, *39*(2), 91–127.

- Perkes, I. E., Morris, R. W., Griffiths, K. R., Quail, S., Waters, F., O'Brien, M., Hazell, P. L., & Balleine, B. W. (2022). *The motivational control of human action, its neural bases and functional impact in OCD* (p. 2022.03.19.22272645). medRxiv. <https://doi.org/10.1101/2022.03.19.22272645>
- Pool, E., Gera, R., Fransen, A., Perez, O. D., Cremer, A., Aleksic, M., Tanwisuth, S., Quail, S., Ceceli, A. O., Manfredi, D., Nave, G., Tricomi, E., Balleine, B., Schonberg, T., Schwabe, L., & O'Doherty, J. P. (2021). *Determining the effects of training duration on the behavioral expression of habitual control in humans: A multi-laboratory investigation*. PsyArXiv. <https://doi.org/10.31234/osf.io/z756h>
- Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, *18*(1), 178. <https://doi.org/10.1186/s12888-018-1761-4>
- Reber, J., Feinstein, J. S., O'Doherty, J. P., Liljeholm, M., Adolphs, R., & Tranel, D. (2017). Selective impairment of goal-directed decision-making following lesions to the human ventromedial prefrontal cortex. *Brain: A Journal of Neurology*, *140*(6), 1743–1756. <https://doi.org/10.1093/brain/awx105>
- Rescorla, R. A. (1990). Evidence for an association between the discriminative stimulus and the response-outcome association in instrumental learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*(4), 326–334. <https://doi.org/10.1037/0097-7403.16.4.326>
- Rescorla, R. A. (1991). Associative relations in instrumental learning: The eighteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section B*, *43*(1), 1–23. <https://doi.org/10.1080/14640749108401256>

- Rescorla, R. A. (1992). Response-outcome versus outcome-response associations in instrumental learning. *Animal Learning & Behavior*, *20*(3), 223–232.
<https://doi.org/10.3758/BF03213376>
- Robbins, T. W., & Costa, R. M. (2017). Habits. *Current Biology*, *27*(22), R1200–R1206.
<https://doi.org/10.1016/j.cub.2017.09.060>
- Robbins, T. W., Vaghi, M. M., & Banca, P. (2019). Obsessive-Compulsive Disorder: Puzzles and Prospects. *Neuron*, *102*(1), 27–47.
<https://doi.org/10.1016/j.neuron.2019.01.046>
- Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, *29*(22), 7191–7198. <https://doi.org/10.1523/JNEUROSCI.0979-09.2009>
- Sjoerds, Z., Dietrich, A., Deserno, L., de Wit, S., Villringer, A., Heinze, H.-J., Schlagenhaut, F., & Horstmann, A. (2016). Slips of Action and Sequential Decisions: A Cross-Validation Study of Tasks Assessing Habitual and Goal-Directed Action Control. *Frontiers in Behavioral Neuroscience*, *10*. <https://doi.org/10.3389/fnbeh.2016.00234>
- Smith, K. S., & Graybiel, A. M. (2013). A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics. *Neuron*, *79*(2), 361–374.
<https://doi.org/10.1016/j.neuron.2013.05.038>
- Thrailkill, E. A., Trask, S., Vidal, P., Alcalá, J. A., & Bouton, M. E. (2018). Stimulus control of actions and habits: A role for reinforcer predictability and attention in the development of habitual behavior. *Journal of Experimental Psychology. Animal Learning and Cognition*, *44*(4), 370–384. <https://doi.org/10.1037/xan0000188>
- Trapold, M. A. (1970). Are expectancies based upon different positive reinforcing events discriminably different? *Learning and Motivation*, *1*(2), 129–140.
[https://doi.org/10.1016/0023-9690\(70\)90079-2](https://doi.org/10.1016/0023-9690(70)90079-2)

- Trask, S., & Bouton, M. E. (2014). Contextual control of operant behavior: Evidence for hierarchical associations in instrumental learning. *Learning & Behavior*, *42*(3), 281–288. <https://doi.org/10.3758/s13420-014-0145-y>
- Trask, S., Shipman, M. L., Green, J. T., & Bouton, M. E. (2020). Some factors that restore goal-direction to a habitual behavior. *Neurobiology of Learning and Memory*, *169*, 107161. <https://doi.org/10.1016/j.nlm.2020.107161>
- Turner, K. M., Svegborn, A., Langguth, M., McKenzie, C., & Robbins, T. W. (2022). Opposing roles of the dorsolateral and dorsomedial striatum in the acquisition of skilled action sequencing in rats. *Journal of Neuroscience*.
- Urcuioli, P. J. (2005). Behavioral and associative effects of differential outcomes in discrimination learning. *Animal Learning & Behavior*, *33*(1), 1–21. <https://doi.org/10.3758/BF03196047>
- Vadillo, M. A., Linssen, D., Orgaz, C., Parsons, S., & Shanks, D. R. (2020). Unconscious or underpowered? Probabilistic cuing of visual attention. *Journal of Experimental Psychology: General*, *149*(1), 160–181. <https://doi.org/10.1037/xge0000632>
- van Elzelingen, W., Warnaar, P., Matos, J., Bastet, W., Jonkman, R., Smulders, D., Goedhoop, J., Denys, D., Arbab, T., & Willuhn, I. (2022). Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. *Current Biology*. <https://doi.org/10.1016/j.cub.2021.12.027>
- Vandaele, Y., & Ahmed, S. H. (2021). Habit, choice, and addiction. *Neuropsychopharmacology*, *46*(4), 689–698. <https://doi.org/10.1038/s41386-020-00899-y>

- Vandaele, Y., Pribut, H. J., & Janak, P. H. (2017). Lever Insertion as a Salient Stimulus Promoting Insensitivity to Outcome Devaluation. *Frontiers in Integrative Neuroscience, 0*. <https://doi.org/10.3389/fnint.2017.00023>
- Verhoeven, A. A. C., & de Wit, S. (2018). The Role of Habits in Maladaptive Behaviour and Therapeutic Interventions. In B. Verplanken (Ed.), *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts* (pp. 285–303). Springer International Publishing. https://doi.org/10.1007/978-3-319-97529-0_16
- Watson, P., & de Wit, S. (2018). Current limits of experimental research into habits and future directions. *Current Opinion in Behavioral Sciences, 20*(Supplement C), 33–39. <https://doi.org/10.1016/j.cobeha.2017.09.012>
- Watson, P., van Wingen, G., & de Wit, S. (2018). Conflicted between Goal-Directed and Habitual Control, an fMRI Investigation. *ENeuro, 5*(4). <https://doi.org/10.1523/ENEURO.0240-18.2018>
- Wood, W., & Rünger, D. (2016). Psychology of Habit. *Annual Review of Psychology, 67*(1), 289–314. <https://doi.org/10.1146/annurev-psych-122414-033417>
- Wu, T., Kansaku, K., & Hallett, M. (2004). How Self-Initiated Memorized Movements Become Automatic: A Functional MRI Study. *Journal of Neurophysiology, 91*(4), 1690–1698. <https://doi.org/10.1152/jn.01052.2003>
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience, 7*(6), 464–476. <https://doi.org/10.1038/nrn1919>
- 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., Knowlton, B. J., & Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action–outcome learning in instrumental conditioning. *European Journal of Neuroscience*, *22*(2), 505–512.

<https://doi.org/10.1111/j.1460-9568.2005.04219.x>

Zapata, A., Minney, V. L., & Shippenberg, T. S. (2010). Shift from Goal-Directed to Habitual Cocaine Seeking after Prolonged Experience in Rats. *The Journal of Neuroscience*, *30*(46), 15457–15463. <https://doi.org/10.1523/JNEUROSCI.4072-10.2010>