



OPEN Stimulus conditions that promote habitual control

Zachary J. Pierce-Messick & Laura H. Corbit

Two experiments in rats examined how training where a stimulus signaled when to respond for reward, conditions that should favour S-R learning, might lead to habitual control of behaviour. Experiment 1 investigated how animals trained with a stimulus preceding lever insertion would impact learning relative to a group that was self-paced and could control lever insertion with a second, distinct response. Rats were then tested for sensitivity to outcome devaluation to distinguish between goal-directed and habitual control. We found that free-operant, self-paced conditions promoted goal-directed control while signaling trials with a stimulus promoted habitual control evidenced as insensitivity to outcome devaluation. Experiment 2 assessed whether the stimulus–outcome association is important for driving habitual responding when training occurs with a traditional discriminative stimulus. A comparison group was trained under free-operant conditions and experienced the same stimulus presented alongside the earned reward. Following devaluation, animals trained under discriminated-operant conditions had reduced goal-directed control, but only after extended training. The free-operant group remained goal-directed, even after extended training, and their performance was not altered by stimulus presentations, suggesting effects of a stimulus–outcome association were unlikely to account for the deficit in the discriminative stimulus group. These results extend understanding of how stimuli present during instrumental training promote the development of habitual control.

Keywords Instrumental learning, Habit, Outcome devaluation, Rat

Instrumental learning paradigms, used to study the acquisition and execution of reward-seeking behaviours, have identified two learning processes that contribute to control of instrumental responses¹. The first is a goal-directed system that relies on knowledge of the causal relationship between a response and its outcome as well as evaluation of the current value of that outcome. Because goal-directed behaviour is controlled by the outcome it produces, changes in the value of that outcome are reflected in performance. However, with extended practice under consistent conditions, the same behaviour can become automatic or habitual¹. Habits have long been argued to rely on stimulus–response (S-R) learning where an association between stimuli (S) present when a response (R) is executed and the response itself is gradually strengthened each time the response is rewarded. These stimuli can then ‘trigger’ that response when similar circumstances are encountered in the future^{2–4}. Importantly, the S-R association thought to control habits does not include encoding of the outcome itself. Thus, changes to the value of the outcome do not produce immediate changes in performance of habits. Based on this framework, the outcome devaluation task has become the standard way to determine whether a behaviour is under goal-directed or habitual control^{5–8}. In this task, animals are trained to make a response (e.g. press a lever) for some desirable outcome (e.g. a food pellet). Then, prior to the test session, the value of that outcome is reduced by pre-feeding the outcome to satiety or inducing a conditioned taste aversion by pairing consumption with illness. Lever-pressing is then tested under extinction conditions. A behaviour that is goal-directed and controlled by an expectation of its consequences should adapt (decrease) in line with the new reduced value of the food, while habitual performance should be relatively unaffected by devaluation as it depends on S-R learning rather than anticipation of the outcome^{1,6,7}.

While there has been substantial progress in defining the behavioural and neural substrates that control goal-directed learning⁹, much less is known about the precise conditions that produce habit learning. Behaviour is often classified as habitual when it fails to meet the criteria for goal-directed control (that is, when it fails to show sensitivity to manipulations of the response–outcome contingency or changes in outcome value) but there is surprisingly little direct evidence that stimuli trigger habits, or even, under free-operant conditions, about what exactly those stimuli might be.

Recognizing this problem, there have been several recent attempts to characterize the stimuli that comprise the S-R association thought to underlie habit learning^{4,10–15}. One procedure that appears to be effective in

Department of Psychology, The University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada.
 email: laura.corbit@utoronto.ca

establishing habitual control was introduced by Vandaele et al.¹⁵ and used a discrete-trial design. In that study, a lever was inserted and if the rats pressed the lever five times, reward was delivered and the lever retracted until the next trial. When tested, animals trained in this fashion were insensitive to outcome devaluation suggesting that behaviour was habitual. The authors suggested that the audiovisual properties of lever insertion were a stimulus and effective in producing habitual control because they were particularly salient.

While Vandaele et al.¹⁵ argued that lever insertion as a stimulus was particularly effective in producing habits, Thraill et al.¹² found that lever insertion was no better at producing habits than a tone trained in a similar fashion. However, Thraill et al.¹² did report that animals trained with lever insertion acquired the instrumental response more quickly than those trained with a tone and so it is possible that lever insertion has unique properties for supporting habits. One limitation of the discrete-trial design with lever insertion serving as the stimulus is that rats can only respond when the lever is present which restricts the comparisons that can be made; the lever as a signal to respond is confounded with the ability to respond. Without the ability to measure lever-pressing in the absence of the stimulus, evidence that the stimulus is controlling responding is limited. To avoid this issue, by adding a light in training, Experiment 1 was designed so that the lever was available throughout the duration of the test. This allowed for comparison of responding when the light was present and when it was absent. Another potential issue with previous designs is that the animal cannot control when trials occur which differs from free-operant conditions where an animal can elect to respond at any time in pursuit of a desired outcome. To address this, we included a group that could initiate trials and lever insertion at any time by pressing a separate lever. Since in training, for the light group, light presentation dictated when the animals could respond and be rewarded, we hypothesized that these conditions would favor the development of an S-R association and that animals in this group would demonstrate habitual control (insensitivity to outcome devaluation). In contrast, because animals that could initiate trials by pressing a separate lever could control when they sought reward and presumably did so with the goal of obtaining that outcome, we predicted that, while controlling for lever insertion/retraction, these conditions would favor goal-directed learning and this group would reduce responding following outcome devaluation.

Results

A summary of the behavioural paradigm is presented in Fig. 1a. Rats (8 males/7 females) were trained under one of two conditions. For animals in the Light group, a circular light to the right of the magazine would illuminate for 6 s followed by the insertion of the reward lever to the left of the magazine, which the animals could then press to earn reward. The Lever group was self-paced and able to control insertion of the reward lever by responding on a second lever (positioned to the right of the magazine). For both groups, responding on the reward lever an average of 10 times earned a reward pellet at which point the lever would retract. Following 8 days of training, all animals were tested for sensitivity to outcome devaluation induced by sensory-specific satiety^{6,8,16}. Briefly, animals were placed in individual feeding cages and allowed unlimited access to their previously earned outcome (devalued) or an alternative food (non-devalued) for one hour. This serves to reduce the current value of the preferred food. Animals were then placed in the training chambers for an extinction test where the reward lever was present for the duration of the 5-min test for both groups. Additionally, every 30 s, either the light (for the Light group) or the second lever (for the Lever group) was presented for 6 s a total of 8 times. This allowed examination of responding in the presence and absence of the stimulus.

Signaling lever availability promotes habitual control

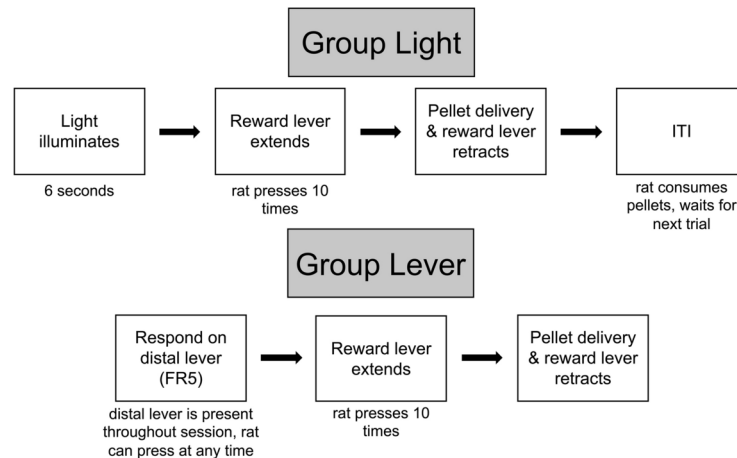
Except for one rat that was excluded, all animals acquired the instrumental response, and demonstrated stable responding throughout training with no group differences observed (largest F -values, day \times group $F [7, 77] = 1.108$, $p = 0.367$, day \times sex $F [7, 77] = 1.721$, $p = 0.116$; all other F 's < 1). Preliminary analyses detected no main effects or interactions with sex (largest F -values including sex were for an interval \times group \times sex interaction ($F [1, 11] = 2.019$, $p = 0.183$, $\eta^2 = 0.010$), and a group \times sex interaction ($F [1, 11] = 1.475$, $p = 0.250$, $\eta^2 = 0.065$). All other F 's < 1 , therefore, the following analyses collapsed across sex.

Consumption was equivalent for the two groups (Supplementary Information Fig. S2). Analyses of the test data examined lever pressing during the 6 s prior to the stimulus and the 6 s when the stimulus (light or distal lever according to group) was present using a $2 \times 2 \times 2$ ANOVA (group \times devaluation \times stimulus). Devaluation reduced responding overall ($F [1, 13] = 11.031$, $p = 0.006$, $\eta^2 = 0.101$) but as seen in Fig. 1, this appeared to differ by group (group \times devaluation interaction: $F [1, 13] = 4.174$, $p = 0.062$, $\eta^2 = 0.038$). Simple effects analyses based on our predictions confirmed that rats in the Lever group reduced responding following devaluation (Fig. 1a; $F [1, 6] = 41.496$, $p < 0.001$, $\eta^2 = 0.472$) while those in the Light group did not (Fig. 1b; $F [1, 7] = 0.554$, $p = 0.481$, $\eta^2 = 0.51$). Neither group significantly altered responding upon stimulus presentation (Lever group: Fig. 1a; $F [1, 6] = 3.508$, $p = 0.110$, $\eta^2 = 0.125$; Light group: Fig. 1b; $F [1, 7] = 4.220$, $p = 0.079$, $\eta^2 = 0.034$). Magazine data are included in Supplementary Fig. 3.

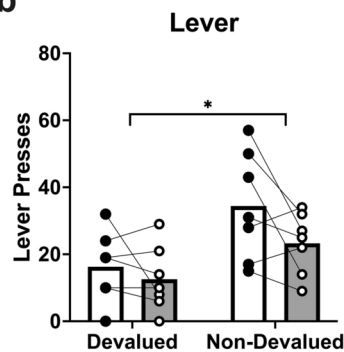
In summary, while some have suggested that lever insertion is a particularly salient stimulus that promotes habitual control, when we trained two groups that experienced lever insertion throughout training but differed in their control of when insertion happened, only rats that were trained where a light preceded lever insertion were insensitive to devaluation. In contrast, self-paced animals (the Lever group) reduced responding when the outcome was devalued, indicative of goal-directed control. These results indicate that training where a stimulus signals the availability of a response impacts the nature of learning in a manner that favors habitual control, but that lever insertion as a stimulus alone was not sufficient to produce habits when the rat controlled that insertion. Instead, behavioural autonomy appeared to be the more significant factor for preserving goal-directed control.

A surprising result from Experiment 1 was that during testing, the same light did not significantly invigorate responding. One possibility is that the audiovisual qualities of lever insertion were salient enough to overshadow the light and so it is possible that the light was not strongly associated with performance of the response. To avoid

a



b



c

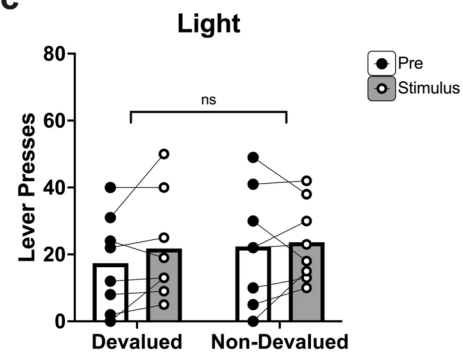


Fig. 1. Training where a light signals lever insertion results in behaviour that is insensitive to outcome devaluation. **(a)** Schematic of the training conditions in the Light and Lever groups. **(b)** The Lever group significantly reduced responding when the previously earned outcome was devalued by outcome-specific satiety. **(c)** The Light group failed to adjust instrumental responding when the earned outcome was devalued. There was no effect of interval (comparison of the 6 s interval containing the stimulus and the 6 s interval that preceded stimulus onset) in either group. “*” indicates a significant effect ($p < 0.05$).

this possibility, Experiment 2 examined behavioural control following training with more typical discriminated operant conditions where the lever is constantly present but responding is only reinforced when a discriminative stimulus (DS) is presented. This allowed investigation of whether habitual control would develop when a stimulus dictated when a response would be successful (and thus when it should be performed) beyond the specific example of lever insertion. While the DS procedure is of interest because it provides an opportunity for a stimulus to become associated with the response, thus allowing investigation of S-R learning, because rats earn and consume the outcome during the stimulus, it is possible that an association between stimulus and outcome (S-O) also forms. If so, stimulus presentations may affect responding during testing (e.g. potentially through a process akin to Pavlovian instrumental transfer^{17,18}). To control for any impact of presenting a stimulus during testing, as well as to address the apparent importance of the animals' ability to control lever access, Experiment 2 included a free-operant group that had similar exposure to the stimulus and thus the opportunity to form a S-O association, but where the stimulus did not dictate when responding would be reinforced.

Habitual control following discriminated operant training is not the result of stimulus-outcome learning

A summary of the behavioural design is presented in Fig. 2A. Rats (15 males) were divided into two groups, a discriminative stimulus (DS) group, and a free-operant (FO) group. For the DS group, the lever was always present but responding was only reinforced (after the completion of 5 presses) during the illumination of a light. This allowed for the elimination of the lever inserting and retracting throughout the training sessions. To be consistent with Experiment 1 and previous related literature, the duration of the DS was 6 s¹².

The FO animals could respond at their own pace and completion of 5 lever-presses led to illumination of the light (6 s) and delivery of a food pellet which the animal could then collect and consume. This allowed for the stimulus to form an association with the earned outcome (S-O) and would serve to control for any impact of stimulus presentations during testing. If any S-O association that forms during DS training is important for

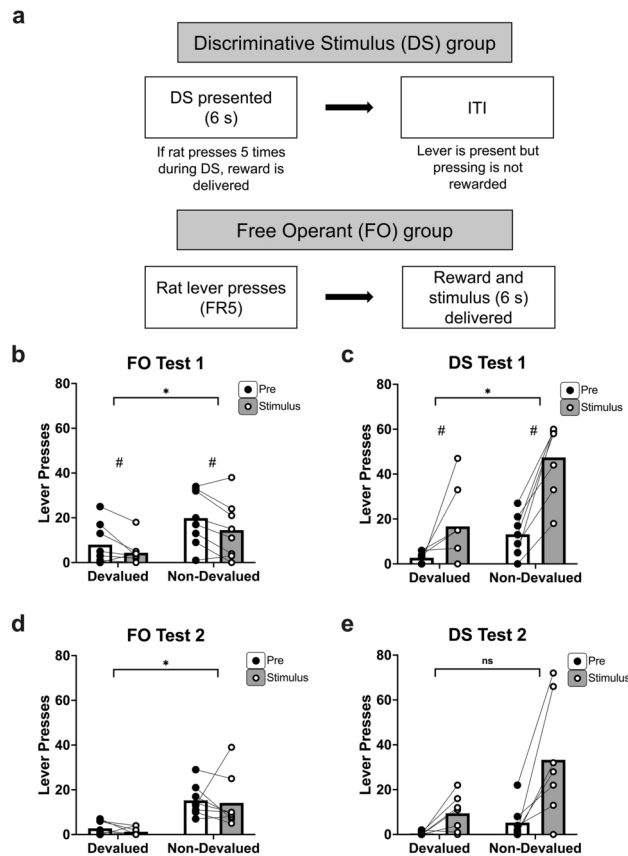


Fig. 2. Training where a discriminative stimulus indicates when instrumental responding will be rewarded decreases sensitivity to outcome devaluation following extended training. **(a)** Schematic of the training conditions in the free operant (FO) and discriminative stimulus (DS) groups. During Test 1, both the FO group **(b)** and DS group **(c)** reduced responding following outcome devaluation. The FO group reduced responding during stimulus presentations relative pre-stimulus intervals, while the DS group increased responding during stimulus presentations. During Test 2, the FO group **(d)** continued to reduce responding following outcome devaluation. In contrast, for the DS group **(e)** the devaluation effect was no longer significant but responding increased significantly during stimulus presentations. “*” indicates a significant effect of devaluation ($p < 0.05$); # indicates a significant effect of interval (comparison of the 6 s interval containing the stimulus and the 6 s interval that preceded stimulus onset).

disrupting sensitivity to devaluation at test, then the impact of stimulus presentation should be seen in both the DS and FO groups. However, if the DS’s relationship with the response (S-R) is the primary association that governs stimulus driven, habitual responding, then only the DS group should show impaired devaluation. We also predicted that the DS group would show more responding during the stimulus than in its absence. In contrast, we predicted that the FO group would reduce responding following outcome devaluation and that stimulus presentations would have little effect on behaviour.

Both groups acquired the instrumental response earning the maximum available outcomes throughout training, thus, no group differences existed. Next, animals underwent devaluation by outcome-specific satiety as described in Experiment 1. Consumption was equivalent between groups and data are reported in Supplementary Fig. S2. The test that followed was 5 min in duration and included eight, 6 s presentations of the light stimulus separated by 30 s interstimulus intervals to assess behavioural control in each group. Due to the nature of the training for the two groups, each had different response rates during the 6 s stimulus presentations of interest which breached the assumption of equality of variance. As such, we examined the effects of devaluation and stimulus presentations separately for each group. The FO group reduced responding when the outcome was devalued (Fig. 2b; $F [1, 7] = 9.611, p = 0.017, \eta^2 = 0.470$), and responded less when the stimulus was present, compared to when it was absent ($F [1, 7] = 9.145, p = 0.019, \eta^2 = 0.079$). There was no devaluation \times interval interaction ($F [1, 7] = 0.447, p = 0.525, \eta^2 = 0.003$). These results indicate that animals in the FO group were goal-directed during both pre-stimulus and stimulus intervals but also responded less during the stimulus presentations.

The DS group also showed a significant devaluation effect (Fig. 2c; $F [1, 6] = 32.659, p = 0.001, \eta^2 = 0.845$), and an effect of interval ($F [1, 6] = 24.521, p = 0.003, \eta^2 = 0.803$) characterized by significantly more responding in the 6 s stimulus interval relative to the 6 s baseline interval that preceded it. There was also a significant

devaluation x interval interaction ($F [1, 6] = 8.101, p = 0.029, \eta^2 = 0.575$), characterized by a larger effect of stimulus in the non-devalued condition.

The failure to find habitual control in the DS group might suggest that training with a stimulus that precedes availability of a response (as in Experiment 1) is particularly effective at promoting S-R learning and habitual control. However, others have found habitual control emerges with more extensive training^{12,19}, thus both groups were trained for an additional 4 sessions and then tested again for sensitivity to outcome devaluation.

The second devaluation test yielded results consistent with the expected impact of extended training for the DS group only.

As shown in Fig. 2e, the DS group showed a significant effect of interval ($F [1, 6] = 16.224, p = 0.007, \eta^2 = 0.730$) characterized by an increase in responding in the stimulus relative to the baseline interval preceding the stimulus. However, extended training resulted in loss of the devaluation effect ($F [1, 6] = 4.662, p = 0.074, \eta^2 = 0.437$), consistent with the known effects of extended training on behavioural control. However, there remained a numeric decrease in responding following devaluation suggesting that goal-directed control was dampened but perhaps not eliminated with extended training in this group. There was no devaluation x interval interaction ($F [1, 6] = 4.063, p = 0.090, \eta^2 = 0.404$).

As shown in Fig. 2d, the FO group still showed a significant devaluation effect ($F [1, 7] = 23.477, p = 0.002, \eta^2 = 0.582$) suggesting that the loss of sensitivity in the DS group was not due to repeated testing. There was no effect of interval ($F [1, 7] = 0.397, p = 0.549, \eta^2 = 0.006$), and no devaluation x interval interaction ($F [1, 7] = 0.007, p = 0.936, \eta^2 < 0.01$).

Discussion

The current experiments extend the results from Thraillkill et al.¹² and Vandaele et al.¹⁵ by demonstrating that the presence of discrete stimuli that signal when to respond can promote habitual control consistent with the long-held view that habits are the result of S-R associations. Experiment 1 found that animals trained where a light stimulus predicted lever availability (Light group) demonstrated habitual control. In contrast, animals that underwent the same amount of training where the reward lever was also inserted and retracted but under the animal's control (Lever group) were goal-directed. This indicates that an animal's level of behavioural autonomy is important for determining whether habits develop. Specifically, animals that control when to respond seem more resistant to the development of habits, while those without such control are more reliant on environmental stimuli to guide behaviour.

Experiment 1 was designed such that light preceded lever insertion, but necessarily, lever insertion was also an antecedent stimulus to lever availability. Thus, it is possible that lever-insertion rather than the light was the more salient stimulus and promoted habit formation¹⁵. Thraillkill et al.¹² demonstrated that lever insertion did not support habitual control, measured as insensitivity to devaluation, more readily than a tone stimulus. However, they did find that animals acquired responding more quickly when trained with an inserting and retracting lever and that performance in this group (but not the tone-stimulus group) was resistant to change when the probability of reinforcement was changed, suggesting some unique properties of lever insertion. In line with this finding, in Experiment 2 where the lever was present throughout the training session but responding was only reinforced during the 6 s light, evidence of habits was only found after extended training suggesting that the light alone did not promote S-R learning as readily as lever-insertion. However, other procedural differences between the two experiments could also account for this effect (e.g. the light preceded the lever and response in Experiment 1 whereas animals responded during the stimulus in Experiment 2).

Another possible difference between the Light and Lever groups is that the light may have promoted sign-tracking in the corresponding group which may have competed with lever pressing. Although this possibility could not be directly assessed as sessions were not videorecorded, this seems an unlikely explanation for the results since were the light to promote sign-tracking behaviour in some animals, any sensitivity to devaluation could still be observed in the stimulus-free intervals which was not the case. Additionally, phenomena such as the Simon effect suggest that the congruence between stimulus and response location is important for task performance²⁰ and here the light was positioned on the opposite side of the chamber relative to the lever. Because the light was physically separated from the lever, and the audiovisual qualities of lever insertion are necessarily inextricable from the lever itself, this further suggests that the light had disadvantages for competing with lever insertion for association with the response.

Whatever the case, the salience of lever insertion does not necessarily mean that it has unique ability to become engaged in S-R habitual control as, importantly, the same lever insertion occurred for the Lever group, and this group demonstrated robust goal-directed control. Thus, whatever the salience of lever insertion, it alone cannot explain the group differences in behavioural control. An important difference between these groups was that the Lever group was self-paced. Since they could choose when to initiate responding, and presumably did so in pursuit of the reward lever and ultimately, reward, this may have served to maintain attention to the consequences of responding and thus, goal-directed control. In contrast, the uncontrollable trial structure and brief access to the lever in group Light may have been optimal for ‘triggering’ responding, and promoting habitual control as it was predicted to do.

Experiment 2 explored behavioural control following training where a DS indicated when a lever response would be reinforced and compared this with a free-operant group where the same stimulus was paired with reward delivery to control for any influence of S-O associations during testing. The results support the idea that discriminated-operant conditions are more likely than free-operant conditions to promote habit development. While this effect was subtle and emerged only after extended training, which has been reported elsewhere to promote habitual control in free operant conditions as well^{1,5}, it is important to note that the free operant group included here remained goal directed after the same amount of training, suggesting that the presence of the

stimulus that dictated when to respond allowed habits to form more quickly. Future experiments using yet further training may produce a more complete effect.

While Thrailkill et al.¹² attempted to rule out conditioned reinforcement as an explanation by preventing stimulus offset from coinciding with outcome delivery, they did not control for the possibility that reinforcement within a DS interval could also allow a S–O association to form that may influence behaviour during testing. There is evidence that a DS is able to form an association with the earned outcome and that the outcome then mediates future responding. Colwill and Rescorla^{21,22} found that a DS was capable of selectively enhancing a separately trained instrumental response provided that it earned the same outcome as the original response trained under that DS. Although Colwill and Rescorla trained multiple responses that earned distinct rewards and outcome-specific associations may be more likely to form when animals are trained with multiple outcomes^{23,24}, these data suggest a DS can become associated with the outcome. We attempted to control for this possibility in Experiment 2. Despite animals in the DS and FO groups having equal exposure to the stimulus, and each had the opportunity to form an S–O association, no effect of stimulus presentation was observed in the FO group, indicating that the S–O association had little influence on behaviour, at least under our training parameters. It is possible that the timing of reward delivery relative to stimulus onset produced weaker S–O learning in the FO group. In the DS group, following stimulus onset, rats had to press the lever 5 times to earn reward which would impose a delay between stimulus onset and reward receipt although reward still occurred during the stimulus. In contrast, in the FO group, stimulus onset and activation of the pellet dispenser were programmed to occur simultaneously. However, the animal still had to collect and consume the pellet and so in practice there was also a delay between stimulus onset and reward receipt though it was likely shorter than in the DS group. Since delay conditioning is generally expected to produce stronger learning than simultaneous conditioning, these differences in timing may have affected the strength of conditioning in the two groups. On the other hand, the intervening event of lever-pressing that occurred between stimulus onset and reward receipt in the DS group might also be expected to undermine formation of any S–O association in favour of the S–R association.

There are a number of limitations and directions for future research from these two experiments. Future research should further examine habitual control in paradigms with distinct R–O associations under discriminated control (e.g. S1:R1–O1, S2:R2–O2). Finding S–R habits in animals trained with distinct R–O associations would enhance validity of animal models of habits as in real-world situations, habits form even though animals (including humans) perform many different behaviours to achieve various outcomes. Another limitation was that while Experiment 1 included half male and female animals, Experiment 2 included only males. While no sex differences were observed in Experiment 1, it should nevertheless be considered in future studies given the increasing recognition of the importance of including sex as a biological variable in biomedical research^{25,26}. Finally, it would be of interest to better understand whether habits trained with these conditions, once established, can be disrupted to restore goal-directed control. Thrailkill et al.^{12,13} found that responding during a DS was resistant to habit formation when the rats were only reinforced on half of the trials suggesting that the predictability of reinforcement is critical for habit formation and that partial reinforcement maintains attention to the consequences of behaviour, thus promoting goal-directed control. Future experiments could extend the results of Experiment 1 and assess whether training where a stimulus dictates when to respond but where reinforcement is variable still leads to habitual control. Pierce-Messick and Corbit¹⁰ found that for animals trained under free operant conditions, where the context itself is likely to serve as the stimulus that forms an S–R association with extended training, exposing animals to the training context without the lever available restored goal-directed control. These results suggest established habits may be broken by altering the predictive relationship between a stimulus (the context or a DS) and a response. Thus, future experiments should test whether exposure to the Light or DS used here when the lever is absent restore goal-directed control.

Overall, our results provide further insight into how the presence of discrete stimuli during instrumental learning alters the nature of behavioural control that forms and provide some direct evidence that S–R associations underlie such behavior. This is important because while often suggested, evidence for such associations is rarely direct. Adding a stimulus that immediately precedes the opportunity to respond or dictates when to respond allows investigations of how stimuli direct responding. The observation that such stimuli promote habitual control in animals that do not have the ability to control response opportunity is consistent with the view that habitual responses are triggered by antecedent stimuli. Of note, when animals were trained with the same stimuli present, but where they did not dictate when to respond, animals remained goal-directed after the same amount of training. This suggests that the mere presence of stimuli does not ensure S–R learning or habitual control. Rather, when stimuli inform when to respond, attention is likely focused on detecting the next stimulus (and not contemplation of the outcome) which then impacts the relationships that are encoded.

These findings are important for advancing our understanding of behavioural control, both from the perspectives of effectively modeling habits in animal models, as well as understanding how habits are expressed in humans and relate to relevant psychopathologies such as substance use disorders. It is important to refine the way that we model habits by focusing on positive evidence for habits and S–R learning, rather than relying heavily on a failure to observe the devaluation effect⁴. These findings also have implications for how humans may be able to adjust particularly when long-practiced behaviours have become entrenched, yet an individual has the desire to break unwanted habits. Treatments that target learned S–R associations may be more effective in producing behaviour change than those that focus on goals if goals are not what drive the behaviour in question.

Methods

All procedures were approved by the Animal Care Committee at the University of Toronto and performed in accordance with ethical standards and guidelines established by the Canadian Council on Animal Care and are reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

Subjects

Sixteen adult Long Evans rats (8 male, 8 female; Charles River, St. Constant, QC, Canada) were used in Experiment 1 and one was excluded from final analyses because they did not acquire the instrumental response. Sixteen male Long Evans were used in Experiment 2. Rats were housed in pairs in individually ventilated cages in a room maintained on a 12:12 light/dark cycle with lights on at 7am. Animals were trained and tested during the light phase. Rats had free access to water and environmental enrichment (e.g. cages with two levels, red tubes, woodblocks, nyla bones) while in the home cage.

Apparatus

Training took place in operant chambers (Med Associates, East Fairfield, Vermont) housed within sound- and light-attenuating shells. The chambers were equipped with pellet dispensers that, when activated, delivered a 45 mg pellet (grain-based formula, BioServ FO165; chocolate sucrose formula; TestDiet 5TUT) into a food magazine. The chambers contained two retractable levers that could be inserted to the left and right sides of the magazine and two key lights (Med Associates, ENV-221 M-LED, white color 2.5 cm diameter), one above each lever (3 cm from lever to light base). A 3 W, 24 V house light mounted on the wall opposite the lever and magazine illuminated the chamber. Computers equipped with MED-PC software (Med Associates) controlled the equipment and recorded responses.

Statistics

Devaluation test data were analyzed using mixed-model analysis of variance (ANOVA). Lever pressing data from the devaluation tests were assessed for normality using the Shapiro–Wilk test. Experiment 2 yielded a significant Shapiro–Wilk p -value in the devalued condition, indicative of a breach in normality. This was not solved by performing a log transformation on the data. Inspecting the data, however, yielded a statistical outlier (using the interquartile range $\times 1.5$ method); removal of this rat corrected the breach of the assumption of normality.

Magazine training

Three days prior to training, access to home cage chow was restricted to maintain rats at 90% of their free feeding weight to motivate them to respond for food pellets. One day prior to the start of training, rats were exposed to the grain and chocolate pellets used in the experiment in the home cage. This was done by putting 1 g of each pellet type in the rats' home cages in such a way that an experimenter could monitor the rats and verify that each cage-mate (as they were dual housed) consumed both outcomes. Rats were randomly assigned to two groups that would earn grain or chocolate pellets throughout training. During magazine training, each animals' assigned outcome was delivered on a random time (RT) 60 s schedule for a 10 min long session. The house light was illuminated but levers were not available during this session.

Lever training

Experiment 1

Sessions terminated once 20 outcomes were earned or 60 min elapsed, whichever was reached first. All rats were first trained to press the reward lever. In the first session each response was rewarded. Thereafter, responding was reinforced according to random ratio (RR) schedules (two days of RR5 and four days of RR10). During this training the reward lever was constantly present (i.e. it did not extend and retract). After this initial phase of training, responding was relatively stable and rats were divided into two groups (Light and Lever) based on their total responses for the last day of RR10 training. For this initial phase of training, there were no differences between groups, ($F [1, 12] = 0.008, p = 0.932, \eta^2 < 0.01$), no effect of sex, ($F [1, 12] = 0.264, p = 0.617, \eta^2 = 0.022$), and no sex \times group interaction, ($F [1, 12] = 1.241, p = 0.287, \eta^2 = 0.094$).

For the second phase of training (8 days), animals in both groups now experienced the reward lever extending and retracting.

For group Light, the light located to the right of the magazine would illuminate for 6 s, and at its termination, the reward lever would extend. The 6 s stimulus duration was selected based on Thrailkill et al.¹² and satisfied some of our aims, such as having a stimulus that was brief enough to be presented in serial (stimulus \rightarrow lever extension) and therefore served as an antecedent stimulus, but the presentations were also long enough to evaluate changes in responding during presentations of this stimulus at test. Light trials were separated by a variable intertrial interval (ITI) that averaged 33 s (range 15–60 s). To minimize the difference in session length between groups, the average ITI for the Lever group was calculated each day and used as the ITI for the Light group on the subsequent day and as a result, the ITI reduced daily for both groups as the Lever group completed the session more quickly. On day 1, the average ITI was 60 s, and by the final day, the average ITI was only 15 s. Because of this lagged procedure, the experienced ITI was slightly longer in the Light group (see Supplementary Fig. 1). Responding on the reward lever yielded a pellet based on a RR10 schedule and would then retract.

For the Lever group, a second distinct lever was available throughout the entire session and responding on this lever five times resulted in the reward lever being inserted into the chamber. Once a pellet was earned by responding on the reward lever (RR10), the reward lever was retracted, and rats were required to press the distal lever to gain access to the reward lever again. Responding on the distal lever while the reward lever was present had no consequences. Responses on the reward lever were stable across the 8 days of training for both groups with all animals earning the maximum possible rewards (20) each day, and no significant effects or interactions were observed (largest F -values, day \times group $F [7, 77] = 1.108, p = 0.367, \eta^2 = 0.092$; day \times sex $F [7, 77] = 1.721, p = 0.116, \eta^2 = 0.135$; all other F 's < 1). One animal in group Lever failed to acquire the task and was excluded.

Devaluation testing was conducted by allowing animals to pre-feed on either the outcome that they earned during training (e.g. grain pellets; devalued) or a different outcome that they had not earned during training (e.g. chocolate pellets; non-devalued), counterbalanced across animals⁸. Animals were offered 15 g of reward pellets

and given 1 h access to the food. Consumption data are included as Supplementary data. Then, animals were placed into their training chambers for the devaluation test which occurred in extinction. During this test, the reward lever was extended and lever pressing was recorded. Six s presentations of either the light stimulus (for Group Light) or the distal lever (for Group Lever) occurred eight times, with a fixed 30 s inter-stimulus interval between presentations. Animals were tested two times on separate days to allow for within-subject comparisons of responding following pre-feeding on rats' earned outcomes (devalued condition) or the different outcome (non-devalued condition).

Experiment 2

Following magazine training as in Experiment 1, rats had an initial lever training session with continuous reinforcement (CRF) that terminated once 20 outcomes were earned or 60 min elapsed. Following this, animals were separated into two groups, the DS group and the FO group. After the initial acquisition session, the DS group received one session where the lever was made available only during DS (6 s light) presentations and each response was reinforced. Thereafter, the lever was always available throughout sessions and responding was only reinforced in the presence of the DS. The DS group received one day of training where responding during the stimulus was reinforced on a CRF schedule followed by two days on a FR3 schedule before being advanced to FR5. Animals in the FO group were able to respond freely, and the same stimulus (6 s light) was presented at the time of reward delivery which allowed for evaluation of how any S–O association may impact performance during testing. Animals in the FO group received the same number of CRF, FR3, and FR5 sessions as the DS group and training continued for all animals until the DS group demonstrated stable discriminative control of responding (19 days).

The ITI for DS presentations was initially programmed to average around 30 s. This short ITI was selected to minimize group differences in the total time spent in the chamber and was also used in Thraillkill et al.¹² for some groups trained under a similarly short (6 s) DS (Experiments 2 & 3). However, after 9 days of training with this ITI, animals in the DS group showed poor discriminative control. Therefore, 10 additional days of training were conducted with a ~60 s ITI to facilitate discrimination. By the final 2 days of training rats responded significantly more during the 6 s DS interval (day 18, $M=96.429$, $SE=6.121$; day 19, $M=92.143$, $SE=6.906$) relative to pre-stimulus intervals of the same length (day 18, $M=66.857$, $SE=10.951$; $F[1, 6]=12.195$, $p=0.013$, $\eta^2=0.670$; day 19, $M=65.571$, $SE=11.769$, $F[1, 6]=9.373$, $p=0.022$, $\eta^2=0.610$). Animals in the FO condition received the same number of training sessions and had the same number of response–outcome pairings during the FR5 phase of training (380). All animals in the FO group earned the maximum number of outcomes (20) available in each session which corresponded with 100 lever presses each session.

Devaluation tests were conducted as described for group Light in Experiment 1, that is, the lever was present throughout and every 30 s the light was illuminated for 6 s. Each test contained 8 such light presentations. After the first round of devaluation tests, all rats received an additional 4 training sessions (plus one session between test 1 and 2 for a total of 24 sessions and 456 reinforced responses) to see if additional training would alter devaluation performance in line with the established impact of extended training in promoting habitual control¹².

Data availability

Raw data for all results presented in this manuscript are included as supplemental files.

Received: 19 August 2024; Accepted: 26 November 2024

Published online: 05 December 2024

References

- Dickinson, A. Actions and habits: The development of behavioural autonomy. *Philos. Trans. R. Soc. B Biol. Sci.* **308**(1135), 67–78. <https://doi.org/10.1098/rstb.1985.0010> (1985).
- Hardwick, R. M., Forrence, A. D., Krakauer, J. W. & Haith, A. M. Time-dependent competition between goal-directed and habitual response preparation. *Nat. Hum. Behav.* **3**(12), 1252–1262. <https://doi.org/10.1038/s41562-019-0725-0> (2019).
- Thorndike, E. L. *Animal Intelligence: Experimental Studies* (Macmillan Press, 1911).
- Watson, P., O'Callaghan, C., Perkes, I., Bradfield, L. & Turner, K. Making habits measurable beyond what they are not: A focus on associative dual-process models. *Neurosci. Biobehav. Rev.* **142**, 104869. <https://doi.org/10.1016/j.neubiorev.2022.104869> (2022).
- Adams, C. D. Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* **34**(2b), 77–98. <https://doi.org/10.1080/14640748208400878> (1982).
- Balleine, B. W. & Dickinson, A. Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**(4–5), 407–419. [https://doi.org/10.1016/S0028-3908\(98\)00033-1](https://doi.org/10.1016/S0028-3908(98)00033-1) (1998).
- Corbit, L. H. Understanding the balance between goal-directed and habitual behavioral control. *Curr. Opin. Behav. Sci.* **20**, 161–168. <https://doi.org/10.1016/j.cobeha.2018.01.010> (2018).
- Pierce-Messick, Z. J., Shipman, M. L., Desilets, G. L. & Corbit, L. H. Outcome devaluation as a method for identifying goal-directed behaviours in rats. *Nat. Protocols* <https://doi.org/10.1038/s41596-024-01054-3> (2024).
- Balleine, B. W. & O'Doherty, J. P. 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> (2010).
- Pierce-Messick, Z. J. & Corbit, L. H. Manipulations of the context-response relationship reduce the expression of response habits. *Neurobiol. Learn. Mem.* **214**, 107962. <https://doi.org/10.1016/j.nlm.2024.107962> (2024).
- Thraillkill, E. A. & Bouton, M. E. Effects of outcome devaluation on instrumental behaviors in a discriminated heterogeneous chain. *J. Exp. Psychol. Anim. Learn. Cogn.* **43**(1), 88–95. <https://doi.org/10.1037/xan0000119> (2017).
- Thraillkill, E. A., Michaud, N. L. & Bouton, M. E. Reinforcer predictability and stimulus salience promote discriminated habit learning. *J. Exp. Psychol. Anim. Learn. Cogn.* **47**(2), 183–199. <https://doi.org/10.1037/xan0000285> (2021).
- Thraillkill, E. A., Trask, S., Vidal, P., Alcalá, J. A. & Bouton, M. E. Stimulus control of actions and habits: A role for reinforcer predictability and attention in the development of habitual behavior. *J. Exp. Psychol. Anim. Learn. Cogn.* **44**(4), 370–384. <https://doi.org/10.1037/xan0000188> (2018).

14. Turner, K. M. & Balleine, B. W. Stimulus control of habits: Evidence for both stimulus specificity and devaluation insensitivity in a dual-response task. *J. Exp. Anal. Behav.* **121**(1), 52–61. <https://doi.org/10.1002/jeab.898> (2024).
15. Vandaele, Y., Pribut, H. J. & Janak, P. H. Lever insertion as a salient stimulus promoting insensitivity to outcome devaluation. *Front. Integr. Neurosci.* <https://doi.org/10.3389/fnint.2017.00023> (2017).
16. Rolls, B. J. Sensory-specific satiety. *Nutr. Rev.* **44**(3), 93–191 (1986).
17. Cartoni, E., Balleine, B. & Baldassarre, G. Appetitive Pavlovian-instrumental transfer: A review. *Neurosci. Biobehav. Rev.* **71**, 829–848. <https://doi.org/10.1016/j.neubiorev.2016.09.020> (2016).
18. Corbit, L. H. & Balleine, B. W. Learning and motivational processes contributing to Pavlovian-instrumental transfer and their neural bases: Dopamine and beyond. In *Behavioral Neuroscience of Motivation* (eds Simpson, E. H. & Balsam, P. D.) 259–289 (Springer International Publishing, 2016). https://doi.org/10.1007/7854_2015_388.
19. Dickinson, A., Balleine, B., Watt, A., Gonzalez, F. & Boakes, R. A. Motivational control after extended instrumental training. *Anim. Learn. Behav.* **23**(2), 197–206. <https://doi.org/10.3758/BF03199935> (1995).
20. Hommel, B. The Simon effect as tool and heuristic. *Acta Psychol.* **136**(2), 189–202. <https://doi.org/10.1016/j.actpsy.2010.04.011> (2011).
21. Colwill, R. M. & Rescorla, R. A. Associations between the discriminative stimulus and the reinforcer in instrumental learning. *J. Exp. Psychol. Anim. Behav. Process.* **14**(2), 155–164 (1988).
22. Colwill, R. M. & Rescorla, R. A. Effect of reinforcer devaluation on discriminative control of instrumental behavior. *J. Exp. Psychol. Anim. Behav. Process.* **16**(1), 40–47 (1990).
23. Holland, P. C. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol. Anim. Behav. Process.* **30**(2), 104–117. <https://doi.org/10.1037/0097-7403.30.2.104> (2004).
24. Kosaki, Y. & Dickinson, A. Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *J. Exp. Psychol. Anim. Behav. Process.* **36**(3), 334–342. <https://doi.org/10.1037/a0016887> (2010).
25. Arnegard, M. E., Whitten, L. A., Hunter, C. & Clayton, J. A. Sex as a biological variable: A 5-year progress report and call to action. *J. Women's Health* **29**(6), 858–864. <https://doi.org/10.1089/jwh.2019.8247> (2020).
26. Bale, T. L. & Epperson, C. N. Sex as a biological variable: Who, what, when, why, and how. *Neuropsychopharmacology* **42**(2), 386–396. <https://doi.org/10.1038/npp.2016.215> (2017).

Acknowledgements

This work was supported by a grant from the Canadian Institutes of Health Research (508660 to LHC).

Author contributions

ZPM and LHC designed experiments. ZPC performed experiments and analyses. ZPM and LHC contributed to manuscript preparation and reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-81309-x>.

Correspondence and requests for materials should be addressed to L.H.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024