1

Evidence of a goal-directed process in human Pavlovian-instrumental transfer

Tina Seabrooke[1], Mike E. Le Pelley[2], Lee Hogarth[3] & Chris J. Mitchell[1]

[1] Plymouth University, UK

[2] UNSW Sydney, Australia

[3] University of Exeter, UK

Running head: Goal-directed reward choice

Word count: 8,823

Please address correspondence to:

Dr Tina Seabrooke

School of Psychology, Portland Square

Plymouth University

Devon

PL4 8AA

United Kingdom

Email: tina.seabrooke@plymouth.ac.uk

Abstract

Cues that signal rewards can motivate reward-seeking behaviours, even for outcomes that are not currently desired. Three experiments examined this phenomenon, using an outcome-selective Pavlovian-instrumental transfer (PIT) design and an outcome devaluation procedure. In Experiment 1, participants learnt to perform one response to earn crisps points and another response to earn popcorn points. One outcome was then devalued by adulterating it to make it taste unpleasant. On test, overall response choice was biased towards the outcome that had not been devalued, indicating goal-directed control. Stimuli that signalled crisps and popcorn also biased instrumental response choice towards their respective outcomes (a PIT effect). Most importantly, the strength of this bias was not influenced by the devaluation manipulation. In contrast, Experiment 2 demonstrated that when stimuli signalled equal probability of the two outcomes, cue-elicited response choice was sensitive to the devaluation manipulation. Experiment 3 confirmed this conclusion by demonstrating a selective avoidance of the cued, devalued outcome. Together, these data support a goal-directed model of PIT in which expected outcome probability and value make independent contributions to response choice.

Keywords: Pavlovian-instrumental transfer; outcome devaluation; goal-directed action.

Contemporary theories of instrumental conditioning emphasise the role of reward-related cues in motivating and directing reward-seeking behaviour (Balleine & O'Doherty, 2010; de Wit & Dickinson, 2009; Dickinson, 2016; Hogarth, Balleine, Corbit, & Killcross, 2013). Pavlovian-instrumental transfer (PIT) tasks have played an important role in clarifying the mechanisms that underlie such cue-driven instrumental response choice (Colwill & Rescorla, 1988; Corbit & Balleine, 2005; Hogarth, Dickinson, Wright, Kouvaraki, & Duka, 2007; Kruse, Overmier, Konz, & Rokke, 1983). In an outcome-selective, appetitive PIT task, two voluntary responses (R1 and R2) are trained to predict distinct rewarding outcomes (O1 and O2), to establish R1-O1 and R2-O2 associations. In a subsequent transfer test, instrumental response choice (R1 vs R2) is then assessed in the presence of Pavlovian stimuli (S1 and S2) that are associated with each outcome (S1-O1, S2-O2). The classic result is that the Pavlovian stimuli selectively elevate the instrumental response that shares a common outcome – i.e., in the presence of Pavlovian stimulus S1, animals are more likely to make response R1 than R2; in the presence of S2, they are more likely to make response R2 than response R1. This "outcome-selective" PIT effect is robust and has been demonstrated in humans with either reward-related pictorial stimuli (e.g., Hogarth, 2012) or experimentally-trained Pavlovian stimuli as S1 and S2 (e.g., Watson, Wiers, Hommel, & de Wit, 2014).

One of the most interesting and counterintuitive aspects of PIT is that it often appears to be insensitive to outcome devaluation manipulations (Corbit, Janak, & Balleine, 2007; Hogarth, 2012; Hogarth & Chase, 2011; Holland, 2004; Rescorla, 1994; van Steenbergen, Watson, Wiers, Hommel, & de Wit, 2017; Watson et al., 2014, but see Allman, DeLeon, Cataldo, Holland, & Johnson, 2010 and Eder & Dignath, 2016a, 2016b for exceptions). Hogarth (2012), for example, trained smokers to perform two responses to earn tobacco and chocolate outcomes (R1-O1, R2-O2). Either tobacco or chocolate (O1 or O2) was then devalued, by having participants ingest nicotine replacement therapy nasal spray, or consume the chocolate to satiety. In the subsequent transfer test, participants were required to respond in the presence of pictorial stimuli that depicted the tobacco and chocolate outcomes, or a neutral stimulus. Response choice on test was biased towards the outcome that had not been devalued, suggesting that overall response choice was goal-directed (de Wit & Dickinson, 2009;

Dickinson, 1985). The tobacco and chocolate cues, by contrast, elevated responding for the outcome with which they had previously been paired, and this elevation was similar regardless of whether the outcome had been devalued. This insensitivity to devaluation is paradoxical because while the PIT effect was outcome-selective, changes in outcome value did not modulate the effect – suggesting that PIT effects do not reflect a goal-directed process.

Several theories have been put forward to explain the insensitivity to outcome devaluation that is often seen in PIT tasks (Cohen-Hatton, Haddon, George, & Honey, 2013; Eder & Dignath, 2016b; Hardy, Mitchell, Seabrooke, & Hogarth, 2017; Hogarth et al., 2014; Hogarth & Chase, 2011; Seabrooke, Hogarth, & Mitchell, 2016; Watson et al., 2014). Most commonly, PIT effects are attributed to the operation of a stimulus-outcome-response (S-O-R) associative chain (Alarcón & Bonardi, 2016; Balleine & O'Doherty, 2010; Balleine & Ostlund, 2007; de Wit & Dickinson, 2015; Hogarth, 2012; Hogarth & Chase, 2011; Holland, 2004; Watson et al., 2014; Watson, Wiers, Hommel, Ridderinkhof, & de Wit, 2016). The core idea here is that Pavlovian conditioning establishes stimulus-outcome (S-O) associations, which allows the stimulus S to bring to mind the associated outcome O. Instrumental conditioning is also suggested to produce bidirectional 'ideomotor' response-outcome (R-O) links, which allow thoughts about an outcome O to prime the associated instrumental response R (de Wit & Dickinson, 2015; Elsner & Hommel, 2001). When a stimulus S is presented during the subsequent transfer test, S-O-R theory proposes that it activates the mental representation of the associated outcome O (via the S-O link), which then triggers the associated instrumental response R (via the ideomotor R-O/O-R link). S-O-R theory predicts the selectivity of the PIT effect because it assumes that the stimulus (e.g. S1) will activate only the outcome with which it was previously paired (O1), which should in turn trigger only the associated instrumental response (R1). To explain the insensitivity of the PIT effect to outcome devaluation, S-O-R theorists have further suggested that the Pavlovian stimulus activates only the *identity* of the associated outcome (not its value), which is sufficient to trigger the associated instrumental response. In this way, S-O-R theory reconciles the outcome-selectivity of the PIT effect with its insensitivity to devaluation.

It should be noted here that several studies in the human PIT literature have recently demonstrated sensitivity to outcome devaluation. These studies demonstrated that, when one outcome had been devalued by instruction (Allman et al., 2010; Eder & Dignath, 2016a) or by making it taste unpleasant (Eder & Dignath, 2016b), a PIT effect was either attenuated or not observed for the devalued outcome. That is, the devaluation manipulations successfully reduced the capacity of an associated stimulus to increase instrumental responding for the devalued outcome. Clearly, these experiments suggest that PIT effects can be goal-directed under some circumstances. They therefore speak against strong claims that the stimulus activates the sensory properties but not the current incentive value of the associated outcome (e.g., Balleine & O'Doherty, 2010; Hogarth et al., 2013; Hogarth & Chase, 2011; Rescorla, 1994b).

Eder and Dignath (2016b) offered a simple solution to reconcile the discrepancies in the human PIT devaluation literature. They suggested that the demonstrations of insensitivity to devaluation might simply be the result of weak devaluation procedures. To support their argument, the authors used a procedure in which a lemonade drink was adulterated to make it taste unpleasant. A PIT effect for this devalued drink was weaker when participants were required to consume the earned lemonade reward periodically throughout the transfer test (i.e. the devalued drink could not be avoided), but not when the rewards were bottled for participants to take away and consume later if they wished. The authors attributed these inconsistent results to the strength of the devaluation manipulation.

Given these discrepancies in the human PIT literature, Experiment 1 re-examined whether the PIT effect is sensitive to devaluation when a very strong devaluation procedure is used, alongside a behavioural measure of PIT that is similar to designs that have previously demonstrated the insensitivity of PIT to devaluation (e.g., Hogarth, 2012; Hogarth & Chase, 2011). The question was whether the devaluation manipulation would affect the ability of the Pavlovian stimulus S to increase choice of the associated instrumental response R, relative to a stimulus that predicted an outcome that had not been devalued. To anticipate somewhat, the standard insensitivity to outcome value was observed in Experiment 1, despite the very strong devaluation manipulation. The subsequent

experiments explored the limits of this insensitivity to devaluation. In particular, Experiment 2 examined whether the usual observation, that PIT is insensitive to outcome value, is the consequence of a ceiling effect on response choice created by the unequal value of the outcomes associated with each response. This will be more fully explained following presentation of the data from Experiment 1.

Experiment 1

Experiment 1 tested whether the PIT effect observed for a devalued outcome would be of an equivalent magnitude to the PIT effect for a valued outcome. The first row of Table 1 shows the design. Participants were first shown a bag of crisps and popcorn (outcomes O1 and O2, counterbalanced) and were told that they could win points corresponding to each type of food during the experiment. Two instrumental responses (R1 and R2) were trained to predict each of these rewards (R1-O1, R2-O2). Participants then sampled each outcome, one of which had been devalued by coating it with ground cloves and olive oil. Response choice (R1 versus R2) was then tested in the presence of pictorial stimuli that were associated with each outcome (stimulus S1, S2, or a neutral stimulus S0).

Following devaluation, we anticipated that overall response choice would be biased towards the valued outcome, indicating that overall choice is goal-directed. Furthermore, we expected that S1 and S2 would increase responding for their associated outcomes relative to baseline responding during S0. The question was whether the cue-elicited increase in choice for the devalued outcome would be smaller or of equivalent magnitude to the cue-elicited increase in choice for the valued outcome, relative to baseline responding during S0. If the PIT effect for the valued outcome is significantly larger than the PIT effect for the devalued outcome, this would suggest that the PIT effect is sensitive to the outcome devaluation manipulation, and that previous demonstrations of insensitivity to outcome devaluation were simply due to the use of weak devaluation procedures (Eder & Dignath, 2016b). If the PIT effect observed for both the valued and devalued outcome are of a similar magnitude, this would support the majority of the literature which has demonstrated that PIT effects are not sensitive to outcome devaluation.

(Table 1 about here)

Method

Participants. Sixty participants (49 female, aged between 18 and 30; M = 21.41 years, *SEM* = 0.39 years) were recruited from Plymouth University and received either £4 or course credit. Participants were screened for food allergies and intolerances. The experiment was approved by the Plymouth University Ethics Committee.

Apparatus and materials. The experiment was programmed in E-Prime 2.0 (Psychology Software Tools, Inc; pstnet.com) and was presented in a small, quiet testing room. Participants made all responses using a standard keyboard. An unopened bag of Walker's extra crunchy ready salted crisps (150g) and Tyrrell's sea salted popcorn (70g) served as visual props. These brands were also used for the devaluation manipulation; before the experiment, the outcomes were decanted into separate transparent containers, with the food name written on the lid. For the devaluation manipulation, ground cloves were combined with olive oil (11 grams oil per 5 grams cloves) to form a paste that was brushed heavily onto the devalued food. The valued outcome was simply transferred from the original packaging to its container.

A picture of crisps or popcorn (depicting the outcomes in their valued state) served as Pavlovian stimuli S1 and S2 during the transfer test. We used familiar, pre-trained Pavlovian cues to be consistent with Hogarth (2012) and Hogarth and Chase (2011), who demonstrated a PIT effect that was insensitive to devaluation using pictorial stimuli. Although the stimuli cannot be counterbalanced using this procedure, this should not affect the ability to detect a PIT effect (see Watson, Wiers, Hommel, Gerdes, & de Wit, 2017; Watson et al., 2016). We used these stimuli because they should have very well-established associations with their outcomes. These stimuli might, therefore, be more likely to produce evidence of automaticity than stimuli that are only weakly associated with the outcomes. Finally, the neutral stimulus was a grey rectangle that was of equal size to the crisps and popcorn stimuli.

Procedure. Participants were warned before the experiment that they would be required to sample foods during the experiment, that they may not match their expectations, and that they may taste unpleasant. The food props were presented, and participants were told they could win points towards them during the experiment.

Liking ratings. Participants initially rated their desire to eat each food in a random order (1 = "not at all", 7 = "very much").

Instrumental training. The props were removed, and instrumental training began with the following instructions: "In this task, you can earn the two outcomes shown before by pressing the left or right arrow keys. Your task is to learn which keys earn each outcome." There were 48 trials. Each trial began with a centrally presented choice symbol ("← or →"), which remained until participants pressed either the left or right arrow key. Each key was selectively paired with either crisps or popcorn, and this was counterbalanced between-subjects. The keys were also counterbalanced with respect to whether they earned the subsequently valued or devalued outcome. One outcome was scheduled to be available on each trial (availability of O1 or O2 on any given trial was random), and so each key had a 50% chance of yielding the associated reward. Rewards were presented as points rather than real food rewards to avoid a generalised devaluation of the outcomes through satiation (see Colagiuri & Lovibond, 2015). Following a left or right arrow key response, the choice symbol was replaced by the feedback statement "You win one [CRISPS/POPCORN] point", or "You win NOTHING" if the available outcome was not selected. The outcome (crisps/popcorn/nothing) was presented in bold text. The names of the rewards (crisps/popcorn) were presented in green or red (counterbalanced) to help participants discriminate between them. All other text was presented in black. Feedback was presented for 3000ms and the trials were separated by 750-1250ms intervals.

Instrumental knowledge test. Following instrumental training, explicit knowledge of the instrumental contingencies was tested. Two questions were presented in a random order: "Which key earned CRISPS (or POPCORN), the left or right arrow key?" Participants also rated their confidence between one (not at all confident) and seven (very confident) after each question.

Outcome devaluation. Participants sampled the valued outcome followed by the devalued outcome. They were told that the devalued outcome was past its expiry date (this was a deception), and that those were the outcomes that were now available. Liking ratings were subsequently taken as at the start of the experiment.

Transfer test. At the start of the transfer test, participants were told: "In this part of the task, you can earn the two outcomes by pressing the left or right arrow key in the same way as before. You will only be told how many of each reward you have earned at the end of the experiment. Also, sometimes pictures of the foods will be presented before you choose the left or right arrow key. NOTE: You will be required to eat all of the food you have earned at the end of the experiment, so please choose carefully." A crisps, popcorn or neutral stimulus was presented at the start of each trial for 3000ms. The choice symbol ("← or →") then appeared beneath the stimulus, until participants selected the left or right arrow key. The test phase was conducted in extinction, and so no feedback was given (i.e., participants were not told whether or which outcomes had been earned); this was done to prevent the feedback from influencing response choice. Trials were separated by an interval of 750-1250ms. There were eight cycles of six trials (48 trials in total); each stimulus was presented twice in a random order in each cycle. Before the transfer test, participants completed one practice cycle, which was identical to the test cycles but was not analysed. The outcome props were removed after the practice cycle, and participants were reminded that they would have to eat all of the food they earned at the end of the experiment[1].

Knowledge tests. After the transfer test, the two outcomes (crisps, popcorn) were presented in a random order and participants were asked whether each outcome was past its expiry date (to assess whether the devaluation cover story had been effective). Confidence ratings were also taken after each question. Participants also completed a second instrumental knowledge test and a Pavlovian knowledge test in which the crisps and popcorn stimuli were presented in a random order and

---

[1] Participants did not actually consume the outcomes after the experiment for ethical reasons. Participants were simply told that they would be required to consume the earned rewards to strengthen the devaluation manipulation (see Eder & Dignath, 2016b).

participants were required to select the outcome that the stimulus represented. In all of the reported experiments, participants also completed expectancy ratings after the transfer test (see Hardy et al., 2017 and Seabrooke et al., 2016). However, these data were inconclusive and are therefore not pursued further here. Finally, a post-experimental questionnaire was used to collect demographic information and feedback about the outcomes. Participants were fully debriefed at the end of the experiment.

Results

Knowledge tests and exclusions. Ten participants failed either the instrumental ($N = 8$) or Pavlovian ($N = 2$) knowledge tests. Confidence ratings were also recorded during the instrumental knowledge test. Unsurprisingly, participants were more confident in their instrumental knowledge before the transfer test ($M = 6.28$, $SE = 0.11$) than they were after the transfer test ($M = 5.03$, $SE = 0.15$). This difference was supported in a two-way ANOVA (pre- versus post-test, and valued versus devalued outcome). There was a main effect of pre- versus post-test, $F(1, 59) = 32.39$, $p < .001$, $\eta_p^2 = .35$, but there was no effect of outcome value and no interaction, $Fs < 1$.

Inaccurate contingency knowledge was an a priori exclusion criterion that was applied because of strong evidence to suggest that human PIT effects are only observed in participants who can verbalise the relevant contingencies (Bezzina, Lee, Lovibond, & Colagiuri, 2016; Hogarth et al., 2007; Lovibond, Satkunarajah, & Colagiuri, 2015; Talmi, Seymour, Dayan, & Dolan, 2008; Trick, Hogarth, & Duka, 2011). The ten participants that failed the contingency knowledge tests where therefore excluded from all subsequent analyses. It is worth noting, however, that none of the exclusions made in the current manuscript affected the overall pattern of results with respect to significance.

Liking ratings. Figure 1a shows the mean liking ratings for each outcome in the pre- and post-devaluation liking tests. Most importantly, there was an interaction between the liking test and outcome variables, $F(1, 49) = 313.47$, $p < .001$, $\eta_p^2 = .87$. The outcomes were rated similarly before

devaluation, $t < 1$, and the valued outcome received higher liking ratings than the devalued outcome after devaluation, $t(49) = 28.95$, $p < .001$, 95% confidence interval difference (CI_{diff}) = [4.62, 5.30].

(Figure 1 about here)

Instrumental training. Across instrumental training, choice of the two responses did not deviate from 50%, $t < 1$.

Transfer test. Figure 1b shows the transfer test results. There was an overall effect of stimulus, $F(2, 98) = 18.42$, $p < .001$, $\eta_p^2 = .27$. Planned pairwise comparisons revealed an outcome-selective PIT effect. The stimulus that signalled the valued outcome biased responding towards that outcome compared to the neutral stimulus S0, $t(49) = 4.21$, $p < .001$, 95% CI_{diff} = [9.92, 28.08], and the stimulus that signalled the devalued outcome, $t(49) = 5.33$, $p < .001$, 95% CI_{diff} = [19.94, 44.06]. The stimulus that signalled the devalued outcome increased responding for the devalued outcome relative to the neutral stimulus S0, $t(49) = 2.46$, $p < .02$, 95% CI_{diff} = [2.37, 23.63]. Finally, overall response choice was biased towards the valued outcome, $t(49) = 8.68$, $p < .001$, 95% CI_{diff} = [20.18, 32.32].

PIT scores were calculated for each outcome to compare the magnitude of the effects. For each outcome, the dependent variable was the percent choice of the response that was paired with the valued outcome. The PIT score for the valued outcome was calculated by subtracting choice on the neutral stimulus (S0) trials from choice on trials where the stimulus signalling the valued outcome was present ($S_{valued} - S0$). This calculation was reversed for the devalued outcome ($S0 - S_{devalued}$). PIT effect scores for the valued and devalued outcome did not significantly differ, $t(49) = 0.77$, $p > .05$, 95% CI_{diff} = [-9.66, 21.66]. This null result was complimented by a Bayes factor of 0.20, which supports the null hypothesis (Morey, Rouder, & Jamil, 2015).

The mean reaction time during the transfer test was 922.22ms ($SE = 111.76$ms). No significant effect of stimulus was observed on reaction times, $F(2, 98) = 1.53$, $p > .05$, $\eta_p^2 = .03$.

Discussion

Experiment 1 demonstrated a PIT effect for both a valued and a devalued outcome. These PIT effects were of a comparable magnitude. Participants reported a clear reduction in liking of the devalued outcome after the devaluation manipulation, and overall response choice was also biased towards the valued outcome in the transfer test. These results suggest that the devaluation manipulation was strong and effective throughout the transfer test. The observation of a PIT effect for the devalued outcome does not, therefore, appear to be for lack of a strong devaluation manipulation (Eder & Dignath, 2016b). Rather, the results suggest that PIT effects can be obtained for even an outcome that has been strongly devalued.

To be consistent with the previous literature, the current results should be interpreted as evidence to suggest that the PIT effect was insensitive to the devaluation manipulation. However, there is an important caveat of this interpretation. In most PIT experiments (including the present experiment), the size of the PIT effect for each outcome is assessed relative to baseline response choice in the presence of a neutral stimulus. Responding in the neutral stimulus condition is usually biased towards the valued outcome after devaluation; participants tend to choose the valued outcome. This bias was particularly apparent in the current experiment because the devaluation manipulation was very strong. However, it is usually present to some degree in PIT experiments that use other devaluation manipulations, such as selective satiation (Hogarth, 2012; Watson et al., 2014), or health warnings (Hogarth & Chase, 2011). Indeed, the non-cued bias towards the valued outcome is regarded as an important component of the overall result, because it demonstrates that the devaluation manipulation was effective throughout the transfer test. The difference between baseline and cue-elicited response choice also accords with dual-controller theories of instrumental action in which expected value and probability of an outcome independently determine responding for that outcome (de Wit & Dickinson, 2009; Hitsman et al., 2013; Hogarth, 2012; Hogarth & Chase, 2011; Watson et al., 2014). However, when responding approaches ceiling (complete responding for the valued outcome) in the neutral/non-cued baseline condition, there is less opportunity to detect a PIT effect for the valued outcome. Conversely, there is relatively *greater* scope to observe a PIT effect for the

devalued outcome. This means that the size of the PIT effect for the valued outcome may be underestimated in the standard task, and the effect for the devalued outcome exaggerated. Ironically, the stronger the devaluation manipulation, the more likely it is that the PIT effect will be seemingly insensitive to devaluation in the current design. In sum, it is possible that the apparent insensitivity of the PIT effect to outcome devaluation is simply due to an artefact in the measurement technique. Experiment 2 tested this possibility.

Experiment 2

Experiment 2 tested whether cue-elicited response choice would be modified by outcome devaluation when overall response choice was balanced close to 50%, to avoid constraint by ceiling or floor effects. The experiment was conceptually very similar to Rescorla's (1994) experiment in rats. The middle row of Table 1 shows the design. Participants first learnt to perform two responses (R1 and R2) to earn points towards four outcomes: crisps, popcorn, cashew nuts and nachos (outcomes O1-O4, counterbalanced). R1 was scheduled to produce O1 and O3 on a random half of the trials each. R2 was scheduled to produce O2 and O4 on half of the trials each. One outcome that was associated with each response (O3 and O4) was then devalued using the cloves adulteration procedure of Experiment 1. Response choice was finally tested in the presence of one of two stimulus compounds: S1 and S4, or S2 and S3. These stimulus compounds both signal one valued and one devalued outcome. They also signal one outcome that is associated with R1 and a second outcome that is associated with R2. The S1+S4 compound, for example, signals outcomes O1 and O4, which were paired with R1 and R2, respectively. Crucially, only stimulus S1 signals a valued outcome (O1), because O4 (signalled by S4) was devalued. Under these conditions, baseline response choice should not be biased in either direction because both responses were trained to produce both a valued and a devalued outcome. The most appropriate comparison would therefore be to directly compare response choice on the S1+S4 trials against the S2+S3 trials. Sensitivity to the devaluation manipulation would be indicated by a selective bias towards the cued, valued outcome. For example, the S1+S4 compound should increase R1 responses, because R1 produced the valued O1 during instrumental training (and R2 produced the devalued O4). By the same logic, the S2+S3 compound should increase R2

responses. Insensitivity to devaluation, by contrast, would be revealed if the stimulus compounds failed to bias response choice in either direction (as Rescorla, 1994, found in rats).

Method

The method was the same as Experiment 1, except in the following respects.

Participants. Thirty-three UNSW Sydney undergraduates (20 female, aged 17-24; $M = 19.00$ years, $SEM = 0.26$ years) participated for course credit. The experiment was approved by the UNSW Sydney Human Research Ethics Advisory Panel (Psychology).

Apparatus and materials. Cobs natural sea salt popcorn (80g), Doritos original salted nachos (170g), Smith's original crisps (170g) and Nobby's salted cashew nuts (300g) were used as props and for the devaluation procedure.

Procedure

Instrumental training. After completing initial liking ratings, participants were instructed that they could earn the four outcomes (crisps, popcorn, cashew nuts and nachos) by pressing the left and right arrow keys. The foods were randomised with respect to the outcomes for which they served (O1-O4). The arrow keys were also randomised with respect to the outcomes that they earned. One outcome that was associated with each response was scheduled to be available on each trial (outcome availability on any given trial was random). R1 was followed by either O1 or O3, depending on which outcome was available. R2 similarly produced either O2 or O4 on each trial. The outcomes associated with each response were presented in red or green (counterbalanced) to help discriminate them. All other aspects of instrumental training were identical to Experiment 1. Following instrumental training, participants completed an instrumental knowledge test that followed the procedure of that used in Experiment 1. That is, the four outcomes were presented in a random order and participants were required to choose the response (left or right arrow key) that produced that outcome. Confidence ratings were recorded as in Experiment 1.

Outcome devaluation. Outcomes O3 and O4 were devalued using the procedure of Experiment 1. The valued outcomes (O1 and O2) were sampled first, followed by the devalued outcomes (O3 and O4). The outcomes were randomly sampled within this constraint.

Transfer test. Each transfer test trial presented compound cues containing S1 and S4 (pictures of outcomes O1 and O4), or S2 and S3 (pictures of O2 and O3). The cues were presented at the top centre and bottom centre of the screen, with cue location counterbalanced across trials. Participants were told that the cue location was not important. After 3000ms, the choice symbol ("← or →") was centrally presented, between the two cues, until participants performed a left or right arrow key response. There were four trial types (S1+S4 and S2+S3, with counterbalanced cue location). Each trial was presented once per cycle, and there were eight cycles (32 trials in total). Trial order was random within each cycle. Participants completed one practice cycle before continuing on to the main transfer test. After the transfer test, participants completed the knowledge tests of Experiment 1, with the four relevant outcomes.

Results

Knowledge tests and exclusions. Seven participants failed the instrumental knowledge tests. As in Experiment 1, participants were more confident in their instrumental knowledge before the transfer test ($M = 6.48$, $SE = 0.10$) than after the transfer test ($M = 5.11$, $SE = 0.23$), $F (1, 32) = 21.49$, $p < .001$, $\eta_p^2 = .40$. There was no main effect of outcome (valued versus devalued), nor was there a test $\times$ outcome interaction, $Fs < 1$. The seven participants that failed the instrumental knowledge tests were excluded from the remaining analyses, leaving 26 participants for analysis.

Liking ratings. Figure 2a shows the mean liking ratings for each outcome, at the start of the experiment and after the devaluation procedure. Most importantly, there was an interaction between the liking test and outcomes, $F (1, 25) = 136.90$, $p < .001$, $\eta_p^2 = .85$. The outcomes were rated equally before devaluation, $t < 1$, and higher liking ratings were given to the valued outcomes than the devalued outcomes after devaluation, $t (25) = 15.84$, $p < .001$, 95% $CI_{diff} = [4.08, 5.30]$.

(Figure 2 about here)

Instrumental training. Across instrumental training, choice of the two responses did not deviate from 50%, $t < 1$.

Transfer test. Figure 2b shows the transfer test results. Most importantly, the S1+S4 compound elicited more R1 responses than the S2+S3 compound, t (25) = 9.63, $p < .001$, 95% CI_{diff} = [58.21, 89.87]. Reaction times to the S1+S4 compound (M = 1314.19ms, SE = 144.16ms) did not significantly differ to the reaction times to the S2+S3 compound (M = 1513.28ms, SE = 166.38ms), t (25) = 1.13, $p > .05$, 95% CI_{diff} = [163.85, 562.03].

Discussion

When multiple outcomes and responses were cued on every transfer test trial, participants selectively responded for the cued outcome that was of the highest value. These results speak against the dominant S-O-R account of PIT, in which the stimulus activates the identity but not the value of the associated outcome (Hogarth et al., 2013; Hogarth & Chase, 2011; Holland, 2004; Martinovic et al., 2014; Watson et al., 2014). Rather, the data support a goal-directed model of PIT, where response choice reflects an integration of knowledge about the relevant Pavlovian and instrumental contingencies and the current incentive value of the outcomes (de Wit & Dickinson, 2009; Dickinson, 1985).

The goal-directed pattern of response choice observed during the transfer test could be mediated by a number of processes. One possibility is that participants responded on the basis of an expected utility function that reflected beliefs about the perceived probability and value of each outcome (Hardy et al., 2017; Hogarth, 2012; Hogarth et al., 2014; Seabrooke et al., 2016). It has been suggested that reward-related stimuli increase the perceived probability of the associated outcomes during the transfer test (Cartoni, Moretta, Puglisi-Allegra, Cabib, & Baldassarre, 2015). When both a high- and low-value outcome were cued throughout the transfer test of the current experiment, participants may have inferred that both outcomes were probable/ available, and so they chose the response that was paired with the cued outcome that was of the highest value. When only the devalued outcome is cued (such as in Experiment 1), participants may infer that only that devalued outcome is available, and so response choice is dominated by perceived availability rather than

outcome value. In other words, perceived availability may trump outcome value when only one devalued outcome is cued per trial.

Another possibility is that response choice was driven by an attentional bias towards the cues that signalled the high-value outcomes. That is, participants might have paid more attention to the stimuli that predicted the valued outcomes than those that predicted the devalued outcomes during the transfer test (see Le Pelley, Mitchell, Beesley, George, and Wills, 2016, for a review). Such an attentional bias could have promoted the observed response bias towards the stimulus that signalled the valued outcome. Experiment 3 tested these predictions.

Experiment 3

Experiment 2 provided evidence to suggest that PIT effects can be goal-directed, at least when two outcomes and responses are cued together. Experiment 3 tested this result further. The bottom row of Table 1 shows the design. Initially, participants learnt to perform two instrumental responses (R1 and R2) to earn three different outcomes (O1-O3; crisps, popcorn and cashew nuts, counterbalanced). Both responses were scheduled to produce outcome O3 on half of the trials. On the remaining trials, R1 produced O1 and R2 produced O2. The unique outcomes O1 and O2 were then devalued. In the subsequent transfer test, instrumental response choice (R1 versus R2) was tested in the presence of stimulus compounds that signalled the common, valued outcome (O3), alongside one of the devalued outcomes (O1 or O2).

The three theories outlined above make different predictions in the current design. S-O-R theory predicts that the stimulus compounds should trigger the most strongly activated response, irrespective of outcome value. This prediction is precisely why S-O-R theory is able to explain the insensitivity to devaluation that is usually seen in PIT experiments (including the current Experiment 1). According to S-O-R theory, then, the S1+S3 compound should trigger R1 more strongly than R2, because R1 is associated with both signalled outcomes O1 and O3, whereas R2 is associated with only one of the cued outcomes – O3. By the same logic, the S2+S3 compound should increase R2 responses relative to R1 responses. The expected utility theory outlined above, by contrast, predicts

the opposite result; participants should respond in such a way that they avoid the cued, devalued outcome. Hence, the S1+S3 compound should elicit few R1 responses (relative to R2 responses) because people will be motivated to avoid the cued, devalued O1. Likewise, the S2+S3 compound should elicit fewer R2 responses than R1 responses, in order to avoid the cued, devalued O2. Finally, the attentional explanation of Experiment 2 predicts a null result in the current design. Both instrumental responses are paired with the valued O3 in the instrumental training phase of the current design. An attentional bias towards S3 should not, therefore, produce any response choice bias.

Method

The method was the same as Experiment 2, except in the following respects.

Participants. Thirty participants from Plymouth University (21 females, aged 18-23; $M = 20.50$ years, $SEM = 0.25$ years) completed the experiment in exchange for £4. The experiment was approved by the Plymouth University Ethics Committee.

Apparatus and materials. Walker's extra crunchy ready salted crisps (150g), Tyrrell's sea salted popcorn (70g), and Sainsbury's salted jumbo cashew nuts (400g) were used for props and for the devaluation taste test.

Procedure. Participants were initially shown the food outcomes (crisps, popcorn and cashew nuts) in their original packaging and were told that they could win points towards them. The props were placed in front of the computer, in a consonant location to the keys that produced them (i.e., O3 was placed centrally, and O1 and O2 were placed to the left and right of O3, respectively). Participants completed liking ratings for each outcome (as in Experiments 1 and 2), before the props were removed.

The instrumental training phase followed a similar format to the previous experiments. The experimenter first read aloud the following instructions: "In this part of the task, you can earn the three outcomes shown before by pressing the A or L key. Both keys will produce [O3]. Your task is to learn which keys produce [O1] and [O2]. Press any key to begin." The terms in brackets were replaced by the respective outcome names. The instrumental responses were changed to the "A" and

"L" keys in an attempt to make them more memorable. Each trial began with the statement "Choose a key: A or L?", which remained on-screen until participants selected either the "A" or the "L" key on the computer keyboard. Outcome O3 was scheduled to be available on half of the trials, and outcomes O1 and O2 were available on the remaining trials. The trials were randomly distributed throughout training. Both responses produced O3 when it was available; the "A" key produced O1 and the "L" key produced O2 on the remaining trials. Finally, all text was presented in black during instrumental training.

The instrumental knowledge test followed a similar format to those of Experiments 1 and 2; the three outcomes were presented in a random order and participants were asked to select which response produced each outcome. They used the mouse to choose between three options - "A key", "Both" and "L key". Their choice was outlined in red for 1000ms, and participants then rated their confidence as in Experiments 1 and 2.

Outcomes O1 and O2 were then devalued using the cloves procedure of Experiments 1 and 2. As in the previous experiments, participants always sampled the valued outcome (O3) first, before the other outcomes were revealed. Outcomes O1 and O2 were then sampled randomly. In contrast to Experiments 1 and 2, participants were not told that the devalued outcomes were past their expiry date. Many of the participants reported that they did not believe this instruction in the previous experiments, so it was omitted from the procedure (the devaluation belief measure was also omitted). Post-devaluation liking ratings were completed in the same way as in Experiments 1 and 2.

During the transfer test, S3 was presented on every trial, alongside either S1 or S2. After 3000ms, the choice symbol ("Choose a key: A or L?") was presented between the two stimuli, until either the A or L key was selected. The procedure was otherwise identical to Experiment 2. Participants subsequently completed a further liking test, an instrumental knowledge test (see above) and the Pavlovian knowledge test of Experiments 1 and 2, with the three relevant stimuli.

Results

Knowledge tests and exclusions. Four participants failed the instrumental knowledge tests. As in the previous experiments, participants were more confident in their instrumental knowledge before the transfer test ($M = 6.78$, $SE = 0.09$) than after the transfer test ($M = 4.98$, $SE = 0.23$), F (1, 29) = 47.34, $p < .001$, $\eta_p^2 = .62$. Confidence ratings were also higher for the valued O3 ($M = 6.18$, $SE = 0.15$) than the devalued O1/O2 ($M = 5.58$, $SE = 0.25$), F (1, 29) = 13.65, $p < .001$, $\eta_p^2 = .32$. This result is unsurprising given that the participants were told that both responses would produce O3. There was a significant test × outcome interaction, F (1, 29) = 9.94, $p < .01$, $\eta_p^2 = .26$. Confidence ratings for the valued ($M = 6.80$, $SE = 0.12$) and devalued ($M = 6.77$, $SE = 0.14$) outcomes did not significantly differ before the transfer test, $t < 1$. Higher confidence ratings were given to the valued O3 ($M = 5.57$, $SE = 0.23$) than the devalued O1/O2 ($M = 4.40$, $SE = 0.37$) after the transfer test, t (29) = 3.22, $p < .01$, 95% CI$_{diff}$ = [0.43, 1.91]. The four participants that failed the instrumental knowledge tests were excluded from all subsequent analyses.

Liking ratings. Figure 3a shows the mean liking ratings for the devalued O1 and O2 and the valued O3 in the pre- and post-devaluation liking rating test. Most importantly, there was an interaction between the test and outcome variables, F (1, 25) = 126.93, $p < .001$, $\eta_p^2 = .84$. The outcomes were rated equally before devaluation, $t < 1$, and O1/O2 received lower ratings than O3 after devaluation, t (25) = 25.58, $p < .001$, 95% CI$_{diff}$ = [4.74, 5.57].

(Figure 3 about here)

Instrumental training. Across instrumental training, choice of the two responses did not deviate from 50%, t (25) = 1.12, $p > .05$, CI$_{diff}$ = [45.50, 65.24].

Transfer test. Figure 3b shows the transfer test results. Most importantly, the S2+S3 compound increased R1 responses compared to the S1+S3 compound, t (25) = 5.57, $p < .001$, 95% CI$_{diff}$ = [32.13, 69.79]. Reaction times to the S1+S3 compound ($M = 1181.52$ms, $SE = 126.39$ms) did

not significantly differ to the reaction times to the S2+S3 compound (*M* = 1157.67ms, *SE* = 120.18ms), *t* < 1.

Discussion

Consistent with Experiment 2, Experiment 3 demonstrated that, when a valued and a devalued outcome were cued together, response choice was biased not towards the most strongly signalled response, but towards the response that *did not* produce the cued, devalued outcome. These data support the goal-directed model of PIT, in which cue-elicited response choice is sensitive to changes in outcome value.

General Discussion

Three experiments examined the effect of a strong outcome devaluation manipulation on PIT. Experiment 1 replicated the insensitivity to devaluation that is seen in typical PIT tasks. When only one high- or low-value outcome was cued per trial during the transfer test, response choice was biased towards the response that predicted that outcome, relative to a neutral stimulus. The magnitude of this outcome-selective PIT effect was similar for each outcome, regardless of whether the outcome had been devalued. Experiments 2 and 3, by contrast, demonstrated PIT effects that were highly sensitive to the outcome devaluation manipulation. These latter experiments paired both responses with two outcomes, before one outcome associated with each response was devalued. Stimuli that signalled outcomes that were associated with each instrumental response were then presented together during the transfer test. These stimuli always signalled one valued and one devalued outcome. Under these circumstances, the PIT effect was highly sensitive to the outcome devaluation manipulation: response choice was biased away from the response that predicted the cued, devalued outcome.

The insensitivity to devaluation observed in Experiment 1 (as well as much of the previous literature) is consistent with an S-O-R model in which the Pavlovian stimuli activate the identity but not the current incentive value of the associated outcomes (Hogarth et al., 2013; Hogarth & Chase, 2011; Holland, 2004; Martinovic et al., 2014; Watson et al., 2014). Experiments 2 and 3, by contrast, speak against S-O-R theory. These experiments provide evidence to suggest that Pavlovian stimuli activate both the identities and values of the associated outcomes during PIT transfer tests. The

question, then, is whether there is a theory that is able to explain both the value-insensitive PIT effect that was observed in Experiment 1, and the value-sensitive effects that were observed in Experiments 2 and 3.

Our favoured interpretation of the current data is that PIT effects are, at least in the very specific experimental procedures used here, goal-directed. We propose that response choice is mediated by an expected utility function that reflects beliefs about the probability (Op) and the value (Ov) of each outcome (Hogarth, 2012; Hogarth et al., 2013). We further suggest that the reward cues that are presented during the transfer test increase the perceived probability of the associated outcome (Cartoni et al., 2015; Hardy et al., 2017; Hogarth et al., 2014; Seabrooke et al., 2016). This increase in perceived probability Op is assumed to underlie both the basic PIT effect and its insensitivity to outcome devaluation: participants respond for the cued outcome because it is considered to be much more available than the alternative, non-cued outcome. From this perspective, it is clear that typical PIT tasks, in which only one high- or low-value outcome is cued per trial, confound cue-elicited outcome probability Op with outcome value Ov. When the stimulus that signals the devalued outcome is presented, it signals a low-value (low Ov) outcome that is perceived to have a high probability (Op). The other outcome retains a high value Ov, but is thought to have low probability Op. Under these circumstances, participants may deliberately choose the devalued outcome when it is cued, because earning *something* is preferred to earning nothing at all. As Eder and Dignath (2016b) suggest, this decision-making process may be particularly pronounced when participants do not believe that they have to eat the devalued outcomes after the transfer test, because under these circumstances there is little cost to responding for the cued, devalued outcome. When multiple outcomes are cued and have the same perceived probability, however, response choice is biased towards the response that has the highest utility estimate: participants tend to perform the response that is paired with the cued, valued outcome. A key idea of this theory, then, is that the usual insensitivity seen in PIT tasks reflects a controlled decision-making process rather than an automatic priming mechanism.

The *associative-cybernetic* model of instrumental conditioning provides an alternative framework in which to view the goal-directed PIT effects that were observed in Experiments 2 and 3

(Balleine & Ostlund, 2007; de Wit & Dickinson, 2009, 2015, Dickinson, 1994, 2012, 2016). The associative-cybernetic model proposes that, when an outcome representation is activated, its value is assessed in an incentive system. Information about the outcome's value is then fed back to the motor program. This feedback loop allows modulation (i.e., activation or inhibition) of the instrumental response depending on the current incentive value of the outcome. Consider the S1+S3 compound presented in Experiment 3. Participants tended to perform response R2 here, which did not predict the cued, devalued outcome O1. According to the associative-cybernetic model, the S1+S3 compound should have activated the mental representations of O1 and O3. Then, the value of these outcomes would have been assessed in the incentive system, where O1 would have been recognised as devalued, and O3 as valued. This information would then be fed back to the motor programs through the feedback loop. This feedback loop would result in O3 priming R1 and R2 indiscriminately (because both responses were equally associated with O3). O1 would also inhibit the performance of R1, because O1 was devalued. Thus, the priming effect of O3 on R1 would be offset by the inhibition of R1 by O1. R2 would therefore be activated more strongly than R1, and so should be preferentially performed. This is one way in which an associative, link-based model could account for the goal-directed PIT effects that were observed in Experiments 2 and 3.

It is not clear, however, whether the model described above can account for the insensitivity to devaluation seen in Experiment 1 (as well as much of the previous literature). If outcome value is assessed in the incentive system, then the response that predicts the cued, devalued outcome should not be performed (because the incentive system should inhibit the response just as in Experiments 2 and 3). One way to reconcile Experiment 1 with the associative-cybernetic model would be to appeal to the ceiling effects issue discussed above. That is, to suggest that the previous demonstrations of insensitivity to devaluation were artefactual because of the bias in baseline response choice. The associative-cybernetic model would then make a clear prediction: when baseline response choice is equated, a PIT effect should not be observed for a devalued outcome, even when it is the only outcome that is cued. This prediction merits further attention.

One last issue concerns to the extent to which the current findings in humans are consistent with similar studies in the animal literature. Notably, the effect observed in Experiment 2 is quite different from that observed by Rescorla (1994), where rats showed insensitivity to devaluation in the presence of stimuli that signalled both a high- and low-value outcome. One possibility is that the differential results in rats and humans arose from procedural differences between the two experiments. As Rescorla noted, it is very difficult to ensure that outcome devaluation is complete. It is possible that the outcomes were still somewhat valued in Rescorla's study, even though the rats rejected at least some of the devalued outcomes during the devaluation procedure. Any residual value could have produced the observed indifference between the valued and devalued outcomes during the transfer test; retrieving outcome value may be more difficult when Pavlovian stimuli are presented, compared to when the outcome itself is available for consumption. An alternative possibility is that PIT effects are mediated by fundamentally different processes in rats and humans; that human PIT effects are goal-directed, and that rodent PIT effects reflect a more automatic S-O-R mechanism. This analysis would have a profound influence on our interpretation of rodent PIT experiments, because it would suggest that rodent data might translate poorly to humans. To progress this debate, it seems sensible to first replicate Rescorla's experiment in rats, perhaps with a stronger devaluation manipulation. It would also be a worthwhile endeavour to translate the procedure used in Experiment 3 for use in non-human subjects. The latter design is particularly useful because the automatic and goal-directed theories predict opposite results.

In summary, the current experiments tested the effect of a strong outcome devaluation manipulation on PIT. When only one high- or low-value outcome was cued throughout the transfer test, PIT effects were observed for both the valued and devalued outcomes, and these were of a comparable magnitude. When multiple outcomes and responses were cued on every test trial, however, response choice was biased away from the cued, devalued outcome. In our view, these data provide compelling evidence for a goal-directed process underlying PIT in humans.

References

Alarcón, D., & Bonardi, C. (2016). The effect of conditoned inhibition on the specific Pavlovian-instrumental transfer effect. *Journal of Experimental Psychology: Animal Learning and Cognition*, *42*(1), 82–94. http://doi.org/http://dx.doi.org/10.1037/xan0000087

Allman, M. J., DeLeon, I. G., Cataldo, M. F., Holland, P. C., & Johnson, A. W. (2010). Learning processes affecting human decision making: An assessment of reinforcer-selective Pavlovian-to-instrumental transfer following reinforcer devaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(3), 402–408. http://doi.org/10.1037/a0017876

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. http://doi.org/10.1038/npp.2009.131

Balleine, B. W., & Ostlund, S. B. (2007). Still at the choice-point: Action selection and initiation in instrumental conditioning. *Annals of the New York Academy of Sciences*, *1104*, 147–171. http://doi.org/10.1196/annals.1390.006

Bezzina, L., Lee, J. C., Lovibond, P. F., & Colagiuri, B. (2016). Extinction and renewal of cue-elicited reward-seeking. *Behaviour Research and Therapy*, *87*, 162–169. http://doi.org/10.1016/j.brat.2016.09.009

Cartoni, E., Moretta, T., Puglisi-Allegra, S., Cabib, S., & Baldassarre, G. (2015). The relationship between specific pavlovian instrumental transfer and instrumental reward probability. *Frontiers in Psychology*, *6*(NOV), 1–7. http://doi.org/10.3389/fpsyg.2015.01697

Cohen-Hatton, S. R., Haddon, J. E., George, D. N., & Honey, R. C. (2013). Pavlovian-to-instrumental transfer: Paradoxical effects of the Pavlovian relationship explained. *Journal of Experimental Psychology: Animal Behavior Processes*, *39*(1), 14–23. http://doi.org/10.1037/a0030594

Colagiuri, B., & Lovibond, P. F. (2015). How food cues can enhance and inhibit motivation to obtain and consume food. *Appetite*, *84*, 79–87. http://doi.org/10.1016/j.appet.2014.09.023

Colwill, R. M., & Rescorla, R. A. (1988). Associations between the discriminative stimulus and the reinforcer in instrumental learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(2), 155–164. http://doi.org/10.1037/0097-7403.14.2.155

Corbit, L. H., & Balleine, B. W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *Journal of Neuroscience*, *25*(4), 962–970. http://doi.org/10.1523/JNEUROSCI.4507-04.2005

Corbit, L. H., Janak, P. H., & Balleine, B. W. (2007). General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. *European Journal of Neuroscience*, *26*(11), 3141–3149. http://doi.org/10.1111/j.1460-9568.2007.05934.x

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. http://doi.org/10.1007/s00426-009-0230-6

de Wit, S., & Dickinson, A. (2015). *Ideomotor mechanisms of goal-directed behavior*. (T. S. Braver, Ed.), *Motivation and Cognitive Control*. Retrieved from https://books.google.co.uk/books?hl=en&lr=lang_en&id=TFFACwAAQBAJ&oi=fnd&pg=PA123&dq=%22pavlovian-instrumental+transfer%22&ots=sPpJUmww2e&sig=MCfGv5yUIn7ubOIoZhKb87oC_LE#v=onepage&q=%22pavlovian-instrumental transfer%22&f=false

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *308*(1135), 67–78. http://doi.org/10.1098/rstb.1985.0010

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

Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the*

Royal Society B: Biological Sciences, *367*(1603), 2733–2742. http://doi.org/10.1098/rstb.2012.0220

Dickinson, A. (2016). Instrumental conditioning revisited: Updating dual-process theory. In J. B. Trobalon & V. D. Chamizo (Eds.), *Associative Learning and Cognition. Homage to Professor N. J. Mackintosh. In Memoriam (1935-2015)* (pp. 177–196).

Eder, A. B., & Dignath, D. (2016a). Asymmetrical effects of posttraining outcome revaluation on outcome-selective Pavlovian-to-instrumental transfer of control in human adults. *Learning and Motivation*, *54*, 12–21. http://doi.org/10.1017/CBO9781107415324.004

Eder, A. B., & Dignath, D. (2016b). Cue-elicited food seeking is eliminated with aversive outcomes following outcome devaluation. *Quarterly Journal of Experimental Psychology*, *69*(3), 574–588. http://doi.org/10.1080/17470218.2015.1062527

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 229–240. http://doi.org/10.1037//0096-1523.27.1.229

Hardy, L., Mitchell, C. J., Seabrooke, T., & Hogarth, L. (2017). Drug cue reactivity involves hierarchical instrumental learning: Evidence from a biconditional Pavlovian to instrumental transfer task. *Psychopharmacology*.

Hitsman, B., Hogarth, L., Tseng, L.-J., Teige, J. C., Shadel, W. G., DiBenedetti, D. B., … Niaura, R. (2013). Dissociable effect of acute varenicline on tonic versus cue-provoked craving in non-treatment-motivated heavy smokers. *Drug and Alcohol Dependence*, *130*(1–3), 135–41. http://doi.org/10.1016/j.drugalcdep.2012.10.021

Hogarth, L. (2012). Goal-directed and transfer-cue-elicited drug-seeking are dissociated by pharmacotherapy: Evidence for independent additive controllers. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(3), 266–278. http://doi.org/10.1037/a0028914

Hogarth, L., Balleine, B. W., Corbit, L. H., & Killcross, S. (2013). Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Annals of the New York Academy of Sciences*, *1282*(1), 12–24. http://doi.org/10.1111/j.1749-6632.2012.06768.x

Hogarth, L., & Chase, H. W. (2011). Parallel goal-directed and habitual control of human drug-seeking: Implications for dependence vulnerability. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(3), 261–276. http://doi.org/10.1037/a0022913

Hogarth, L., Dickinson, A., Wright, A., Kouvaraki, M., & Duka, T. (2007). The role of drug expectancy in the control of human drug seeking. *Journal of Experimental Psychology: Animal Behavior Processes*, *33*(4), 484–496. http://doi.org/10.1037/0097-7403.33.4.484

Hogarth, L., Retzler, C., Munafò, M. R., Tran, D. M. D., Troisi, J. R., Rose, A. K., … Field, M. (2014). Extinction of cue-evoked drug-seeking relies on degrading hierarchical instrumental expectancies. *Behaviour Research and Therapy*, *59*, 61–70. http://doi.org/10.1016/j.brat.2014.06.001

Holland, P. C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(2), 104–117. http://doi.org/10.1037/0097-7403.30.2.104

Kruse, J. M., Overmier, J. B., Konz, W. A., & Rokke, E. (1983). Pavlovian conditioned stimulus effects upon instrumental choice behavior are reinforcer specific. *Learning and Motivation*, *14*, 165–181. http://doi.org/10.1016/0023-9690(83)90004-8

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin*.

Lovibond, P. F., Satkunarajah, M., & Colagiuri, B. (2015). Extinction can reduce the impact of reward cues on reward-seeking behavior. *Behavior Therapy*, *46*(4), 432–438. http://doi.org/10.1016/j.beth.2015.03.005

Martinovic, J., Jones, A., Christiansen, P., Rose, A. K., Hogarth, L., & Field, M. (2014). Electrophysiological responses to alcohol cues are not associated with Pavlovian-to-instrumental transfer in social drinkers. *PLoS ONE*, *9*(4), e94605. http://doi.org/10.1371/journal.pone.0094605

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *Package "BayesFactor."*

Rescorla, R. A. (1994). Transfer of instrumental control mediated by a devalued outcome. *Animal Learning & Behavior*, *22*(1), 27–33. http://doi.org/10.3758/BF03199953

Seabrooke, T., Hogarth, L., & Mitchell, C. J. (2016). The propositional basis of cue-controlled reward seeking. *Quarterly Journal of Experimental Psychology*, *69*(12), 2452–2470. http://doi.org/10.1080/17470218.2015.1115885

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human Pavlovian-instrumental transfer. *Journal of Neuroscience*, *28*(2), 360–368. http://doi.org/10.1523/JNEUROSCI.4028-07.2008

Trick, L., Hogarth, L., & Duka, T. (2011). Prediction and uncertainty in human Pavlovian to instrumental transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 757–765. http://doi.org/10.1037/a0022310

van Steenbergen, H., Watson, P., Wiers, R. W., Hommel, B., & de Wit, S. (2017). Dissociable corticostriatal circuits underlie goal-directed versus cue-elicited habitual food seeking after satiation: Evidence from a multimodal MRI study. *European Journal of Neuroscience*. http://doi.org/10.1111/ejn.13586

Watson, P., Wiers, R. W., Hommel, B., & de Wit, S. (2014). Working for food you don't desire. Cues interfere with goal-directed food-seeking. *Appetite*, *79*, 139–148. http://doi.org/10.1016/j.appet.2014.04.005

Watson, P., Wiers, R. W., Hommel, B., Gerdes, V. E. A., & de Wit, D. (2017). Stimulus control over action for food in obese versus healthy-weight individuals. *Frontiers in Psychology*, *8*.

http://doi.org/10.3389/fpsyg.2017.00580

Watson, P., Wiers, R. W., Hommel, B., Ridderinkhof, K. R., & de Wit, S. (2016). An associative

account of how the obesogenic environment biases adolescents' food choices. *Appetite*, *96*, 560–

571. http://doi.org/10.1016/j.appet.2015.10.008

Table 1

Design of Experiments 1, 2 and 3

	Instrumental training	Outcome devaluation	Transfer test
Experiment 1	R1 – O1	O1 *or* O2 devalued	S0: R1/R2?
	R2 – O2		S1: R1/R2?
			S2: R1/R2?
Experiment 2	R1 – O1, O3	O3 *and* O4 devalued	S1+S4: R1/R2?
	R2 – O2, O4		S2+S3: R1/R2?
Experiment 3	R1 – O1, O3	O1 *and* O2 devalued	S1+S3: R1/R2?
	R2 – O2, O3		S2+S3: R1/R2?

Note: R1 and R2 represent instrumental responses. O1-O4 denote food outcomes. S0 represents a neutral stimulus, and S1-S4 are pictorial stimuli that are associated with O1-O4, respectively.
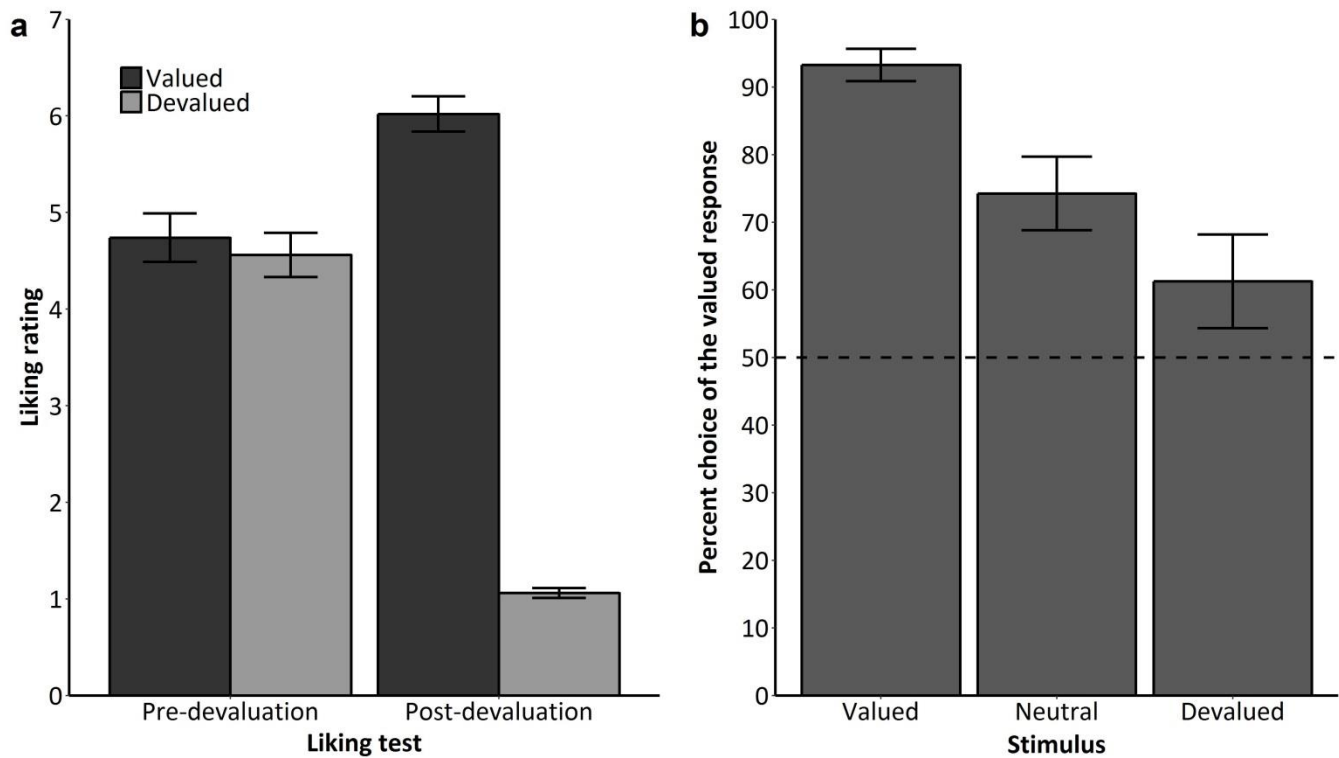
Figure 1

Figure 1. (a) Mean liking ratings in Experiment 1. Ratings were taken for each outcome (valued, devalued) at the start of the experiment (pre-devaluation), and immediately after the devaluation procedure (post-devaluation). Ratings of one and seven represent wanting to eat the outcome "Not at all" and "Very much", respectively. **(b)** Transfer test results of Experiment 1. Response choice was tested in the presence of pictorial stimuli depicting each outcome, or a neutral stimulus. The dependent variable is the percent choice of the instrumental response that was paired with the valued outcome during instrumental training. The dotted line represents no bias in response choice. Scores above 50% demonstrate a bias towards the valued outcome. The error bars are within-subjects corrected and represent the standard error of the mean (SEM).
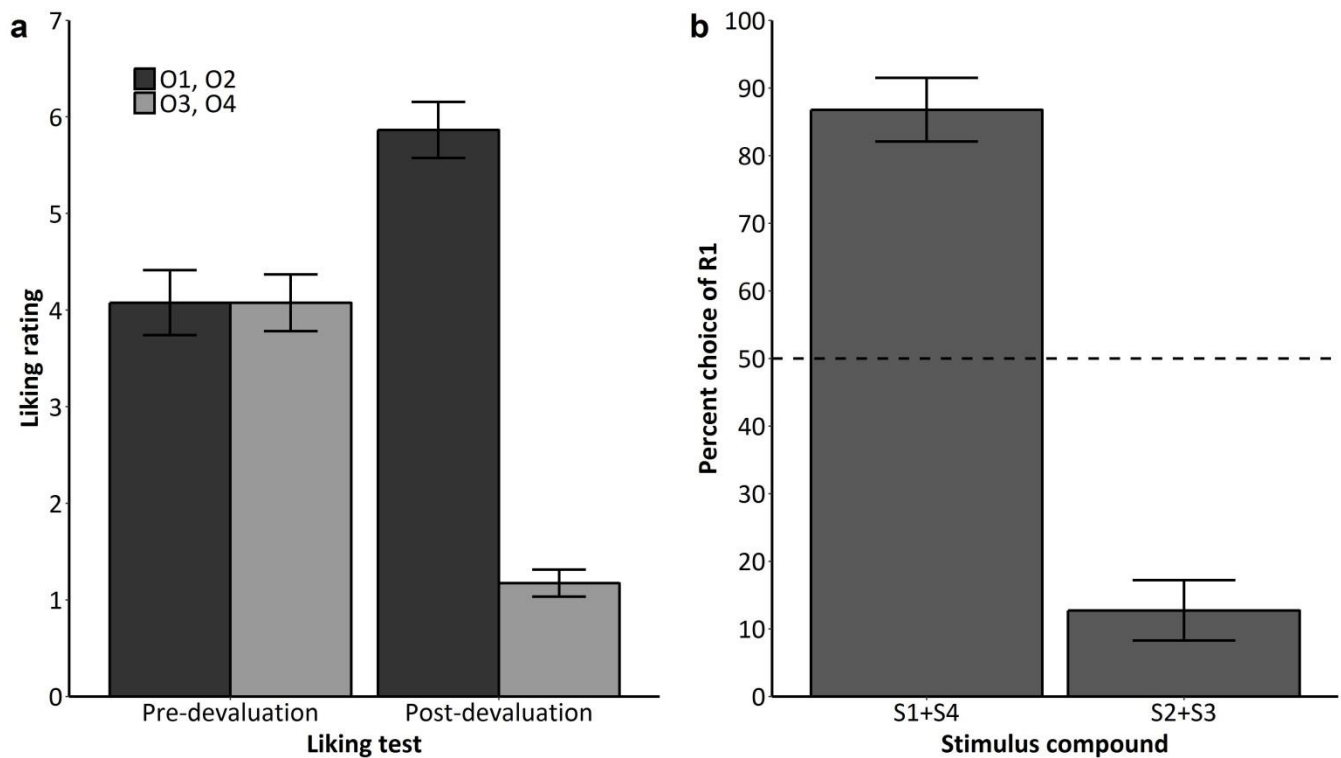
Figure 2

Figure 2. (a) Mean liking ratings in Experiment 2. Ratings were taken for each outcome (O1-O4) at the start of the experiment (pre-devaluation), and immediately after the devaluation procedure (post-devaluation). Ratings of one and seven represent wanting to eat the outcome "Not at all" and "Very much", respectively. **(b)** Transfer test results of Experiment 2. Response choice was tested in the presence of compound stimuli depicting outcomes O1 and O4 (S1+S4), and O2 and O3 (S2+S3). The dependent variable is the percent choice of the R1 response. The dotted line represents no bias in response choice. Scores above 50% demonstrate a bias towards R1, which was paired with O1 (valued) and O3 (devalued) during instrumental training. Scores below 50% represent a bias towards R2, which was paired with O2 (valued) and O4 (devalued) during instrumental training. The error bars are within-subjects corrected and represent the SEM.
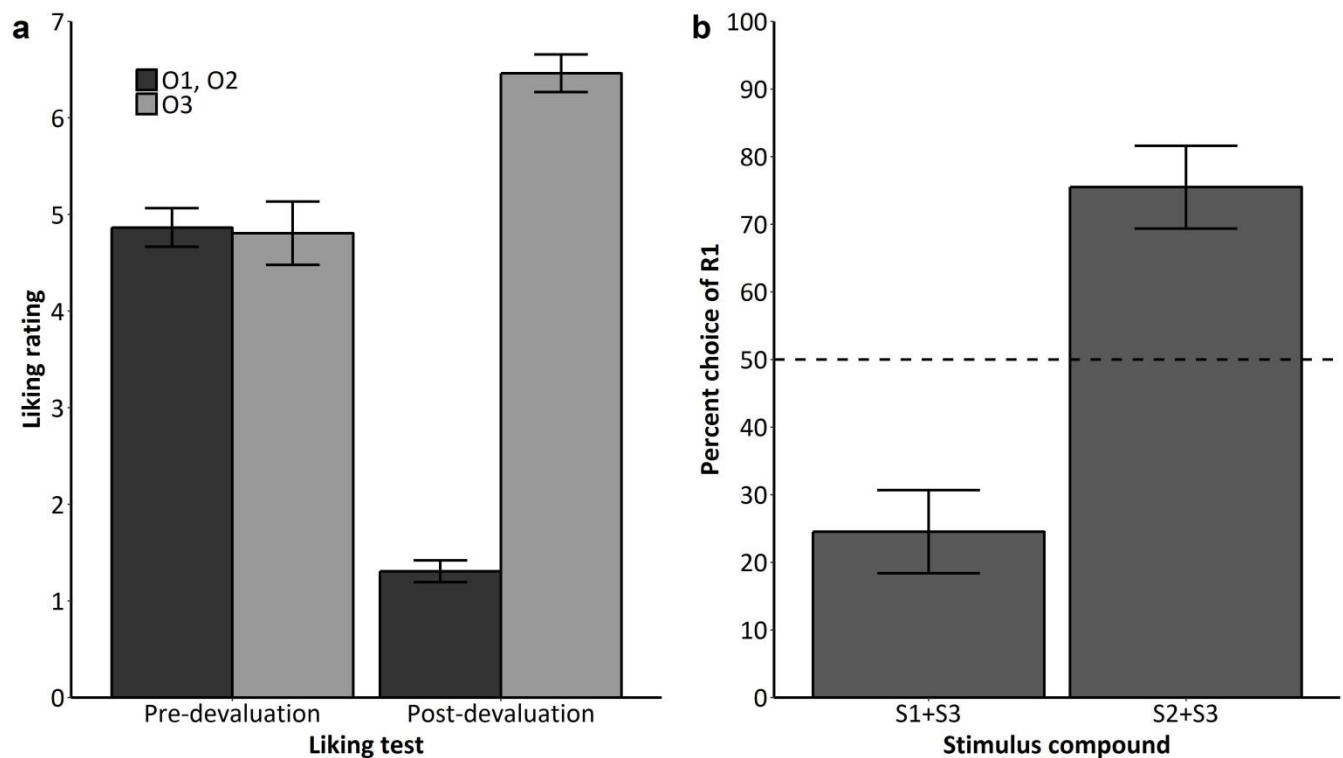
Figure 3

Figure 3. (a) Mean pre- and post-devaluation liking rating in Experiment 3 (1 = wanting to eat the outcome not at all, 7 = wanting to eat the outcome very much). **(b)** Transfer test results of Experiment 3. Response choice was tested in the presence of stimulus compounds that depicted either O1 or O2 with O3 (S1+S3, S2+S3). The dependent variable is the percent choice of the R1 response. The dotted line represents no bias in response choice. Scores above 50% demonstrate a bias towards R1, which was paired with O1 (devalued) and O3 (valued) during instrumental training. Scores below 50% represent a bias towards R2, which was paired with O2 (devalued) and O3 (valued) during instrumental training. The error bars are within-subjects corrected and represent the SEM.