American Psychological Association

Journal of Experimental Psychology: General

© 2019 American Psychological Association
ISSN: 0096-3445

2020, Vol. 149, No. 8, 1449–1459
http://dx.doi.org/10.1037/xge0000722

Measuring Habit Formation Through Goal-Directed Response Switching

David Luque
Autonomous University of Madrid and University of Málaga

Sara Molinero
University of Málaga

Poppy Watson
UNSW Sydney

Francisco J. López
University of Málaga

Mike E. Le Pelley
UNSW Sydney

Reward-learning theory views habits as stimulus–response links formed through extended reward training. Accordingly, animal research has shown that actions that are initially goal-directed can become habitual after operant overtraining. However, a similar demonstration is absent in human research, which poses a serious problem for translational models of behavior. We propose that response-time (RT) switch cost after operant training can be used as a new, reliable marker for the operation of the habit system in humans. Using a new method, we show that RT switch cost demonstrates the properties that would be expected of a habitual behavior: (a) it increases with overtraining, (b) it increases when rewards are larger, and (c) it increases when time pressure is added to the task, thereby hindering the competing goal-directed system. These results offer a promising new pathway for studying the operation of the habit system in humans.

Keywords: goal-directed, habit, learning, response time, response conflict

Supplemental materials: http://dx.doi.org/10.1037/xge0000722.supp

The ability to learn how to obtain rewards is one of the core building blocks of behavior and is observed across a huge range of species, from sea slugs (Brembs, Lorenzetti, Reyes, Baxter, & Byrne, 2002) to humans. Instrumental reward learning describes the process by which an organism learns to adapt its behavior to obtain valued outcomes (food, sex, money, etc.), and this ability to adapt is critical to survival—and to thriving—in a changeable world. Although reward learning is sometimes seen as the realm of behaviorists and cognitive theorists, the influence of rewards on behavior plays an important role across many areas of psychology, including decision making (Hertwig, Barron, Weber, & Erev, 2004; O'Doherty, Cockburn, & Pauli, 2017), social development (Bhanji & Delgado, 2014; Heerey, 2014), perception and attention (Anderson, 2016; Pessoa, 2015), health psychology (Verhoeven & de Wit, 2018), various aspects of psychopathology (e.g., addiction: Hyman, 2005; obsessive–compulsive disorder: Gillan et al., 2014; schizophrenia/psychosis: Kapur, Mizrahi, & Li, 2005; Tourette syndrome: Marsh et al., 2004), psycholinguistics (Ripollés et al., 2014), and many more. Given the foundational status of instrumental reward learning, the desire to gain a better understanding of its underlying psychological processes—and ways in which these processes might be manipulated—has driven substantial previous research and theorizing.

Extensive behavioral and neural research in nonhuman animals suggests that instrumental reward learning reflects the operation of two distinct neuro-cognitive systems: the goal-directed and habit systems (Balleine & O'Doherty, 2010; Verplanken, 2018). *Goal-directed actions* are those that are targeted at obtaining specific, valued outcomes: the goal-directed system "knows" that action X produces outcome Y, and hence will execute action X if outcome Y is currently desired. The *habit system*, on the other hand, operates on direct mental links between a stimulus and an operant response (S-R links), which can become habitual. Habits operate independently of the current incentive value of the outcome; hence, a habit occurs inflexibly when the stimulus is perceived,

This article was published Online First November 21, 2019.

David Luque, School of Psychology, Autonomous University of Madrid, and Department of Basic Psychology, University of Málaga; Sara Molinero, Department of Basic Psychology, University of Málaga; Poppy Watson, School of Psychology, UNSW Sydney; Francisco J. López, Department of Basic Psychology and Institute of Biomedical Research in Malaga (IBIMA), University of Málaga; Mike E. Le Pelley, School of Psychology, UNSW Sydney.

Sara Molinero is now at School of Psychology, Autonomous University of Madrid.

David Luque is supported by a Madrid Senior Fellowship (Talento-CAM 2017-T1/SOC-5147). This research was also supported by Grants PSI2014-56061-P and PGC2018-094694-B-I00 from the Spanish Ministry of Science, Innovation and Universities. We thank María José Quintero for her assistance with Experiment 2b.

Correspondence concerning this article should be addressed to David Luque, School of Psychology, Autonomous University of Madrid, Campus Cantoblanco, 28049 Madrid, Spain. E-mail: david.luque@gmail.com

even if the associated outcome is no longer desired (Dickinson & Balleine, 1994).

The balance between the goal-directed and habit systems will shape our reward-related behavior. Behavior should be goal-directed in new learning situations, because this allows us to flexibly adapt our actions; however, if a given situation is experienced as being stable over an extended period—a particular action consistently produces the same reward—then by relinquishing control to an S-R habit system ("putting it on autopilot") the organism can free up cognitive resources for other purposes (Evans & Stanovich, 2013; Wood & Rünger, 2016). A pivotal factor is, therefore, the amount of experience in a certain reward-training situation: the goal-directed system will control behavior at the beginning of training, while the habit system will gain control after prolonged, stable training. This prediction has been confirmed in research with rats (Adams, 1982; Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995) and it is at the core of neuro-computational models of reward learning (Morris, Bornstein, & Shenhav, 2019).

Laboratory studies of habits in humans have typically followed the general methodological approach taken in animal studies, by trying to index habitual responses via overt selections of responses that yield now-devalued outcomes (e.g., pressing a button to receive chocolate even after satiation on chocolate; Gillan et al., 2014; Tricomi, Balleine, & O'Doherty, 2009; Watson, Wiers, Hommel, & de Wit, 2014); there is now an extensive body of research that has taken this approach in trying to investigate—and dissociate—goal-directed and habit systems in humans (for reviews, see Knowlton & Patterson, 2016; Wood & Rünger, 2016). For instance, neuropsychological research has identified certain neural systems that appear to be relatively insensitive to changes in the current reward-value of an outcome, with the activity of these systems relating to the frequency of participants' overt selections of responses that yield now-devalued outcomes (Liljeholm, Dunne, & O'Doherty, 2015; Schwabe, Tegenthoff, Höffken, & Wolf, 2012; Soares et al., 2012; Tricomi et al., 2009; for a review, see Patterson & Knowlton, 2018). Research in this tradition also suggests that humans show more habitual responses when the goal-directed system is hindered (e.g., by manipulations designed to produce "ego-depletion," Lin, Wood, & Monterosso, 2016); on related lines, it has also been suggested that acute stress can tilt the balance in favor of the habit system (Schwabe & Wolf, 2009; Smeets, van Ruitenbeek, Hartogsveld, & Quaedflieg, 2019)—a result that mirrors findings from animal research (Braun & Hauber, 2013).

Despite the apparent promise of using overt selections of now-devalued outcomes as a measure of habitual behavior in humans, this approach has been less successful when it comes to examining the effect of the amount of training on instrumental behavior. As noted above, a central tenet of the "dual-systems" model of instrumental reward learning is that the transition from goal-directed to habitual behavior depends on experience, with the habit system coming to dominate after prolonged, stable training—a pattern that is clearly demonstrated in rodent studies (Adams, 1982; Dickinson et al., 1995). However, a corresponding pattern has not been reliably demonstrated in human studies that were intended to mirror this animal research. Although Tricomi et al. (2009) did report a transition from goal-directed to habitual behavior following extended training, a series of recent systematic attempts to replicate this critical finding have failed (de Wit et al., 2018). None of the other human studies reviewed above have shown that response selections of now-devalued outcomes are increased by overtraining of the original instrumental relationships (for discussion, see Watson & de Wit, 2018). This is problematic: an increase in habits as a result of overtraining is a diagnostic feature that is supported by animal research, and so the failure to demonstrate a similar result in humans represents a stumbling block for habit theory in general, and for translational research in particular.

In fact, we would argue that overt response selections may not be the best place to look for evidence of habitual behavior in humans.[1] In human laboratory research, even studies using "overtraining" are typically relatively short, and any resulting tendency toward habits is unlikely to be sufficient to overcome participants' well-developed goal-directed system and produce undesired behavior. However, although a habit may not become strong enough during a typical experiment to change a response selection from one option to another, it might exert a detectable influence on the speed with which a response is made. The underlying idea is that observed behavior can reflect the simultaneous activation of both systems; this idea is part of most theoretical models of reward learning (Balleine & O'Doherty, 2010; Dickinson & Balleine, 1994) and is compatible with recent findings from human research (Lee, Shimojo, & O'Doherty, 2014; Luque et al., 2017). Consequently, situations may arise in which the two systems activate incompatible responses at the same time, creating *response conflict* (Watson, van Wingen, & de Wit, 2018) that should manifest in slower responses.

The implication is that, rather than relying on people making undesired overt responses, we can instead assess the activation of S-R habits in terms of how strongly they interfere with ongoing goal-directed behavior. The current study explored the validity of this proposed new measure of habitual behavior. Establishing a reliable way to study the progression of habit formation in humans is highly desirable from a theory-development perspective, as it would help researchers to translate, test, and extend findings of prior animal research into the neuro-cognitive and genetic processes implicated in the development of habitual behavior. This is particularly important given the suggestion that the habit system not only plays a significant role in shaping many of our daily behaviors, but also that dysfunction of this system is implicated in psychopathologies within the impulsive-compulsive spectrum (Gillan et al., 2014; Marsh et al., 2004) and may contribute to some manifestations of Parkinson's disease (Jahanshahi, Obeso, Rothwell, & Obeso, 2015).

In the following experiments participants learned different stimulus-response-outcome (S-R-O) relationships in a first phase of instrumental training. Habit formation was subsequently assessed in an outcome devaluation test. In this devaluation test, participants could still win a desired outcome, but only *if they switched their usual response.* These response switches reflect

[1] Indeed, it has recently been argued that the key measure used in most previous laboratory studies of habitual behavior in humans (overt selections of responses yielding now-devalued outcomes) may not actually reflect habits at all but may instead be a consequence of a misalignment of the participant's goals with the experimenters (De Houwer, Tanaka, Moors, & Tibboel, 2018). We remain agnostic on this issue here as our current data do not address the question directly; interested readers should refer to De Houwer et al.'s article for further discussion.

goal-directed actions, since they are targeted toward obtaining the best possible outcome (goal) and have not been trained before. However, concurrent operation of the habit system may produce competing activation of the previously trained but now devalued (and hence inappropriate) alternative response option. This competing activation would produce response conflict, which might be expected to result in slower responses. That is, conflict might produce a response time (RT) cost. Critically, we investigated whether the magnitude of this *RT switch cost* increased as a function of the amount of initial reward training of the S-R-O contingencies, as would be expected of a valid measure of habits. We also measured the influence of amount of training on participants' overt response selections during devaluation tests—this is the "standard" (Tricomi et al., 2009; Watson et al., 2014) measure of habits, which, as noted above, has to date failed to reveal a reliable increase in habitual behavior as a consequence of overtraining (de Wit et al., 2018; Watson & de Wit, 2018).

Besides amount of training, we investigated the effect of two other factors thought to be important for habits. First, habit learning is assumed to be more effective when the operant behavior leads to significant, high-value outcomes (Marien, Custers, & Aarts, 2018). Second, it has been proposed that the habit system is generally faster than the goal-directed system, and hence that the effect of habits will be more pronounced when participants must respond rapidly (because the goal-directed system has had less time to exert control; e.g., de Wit et al., 2012; Keramati, Smittenaar, Dolan, & Dayan, 2016). We aimed to further assess the validity of our dependent variables (RT switch cost, and response selection) as measures of the habit system by manipulating these factors. To summarize, a marker for the operation of the habit system should be stronger (a) after extended training, (b) when learning is boosted by high-value outcomes, and (c) when the goal-directed system is hindered by response time pressure. To anticipate, only RT switch cost complied with these predictions.

Experiment 1

Method

Participants and apparatus. The critical test in Experiment 1 was the (within-subjects) effect of amount of training on RT switch cost and response selections. We aimed for a sample size of 24 participants per group (where groups varied in time pressure: see below), which would provide power >.80 to detect a medium-size ($\eta_p^2 = .06$) within-subjects effect. In total, 55 University of Málaga students participated for course credit and possible payment (the six highest-scorers received 25€). Participants were randomly allocated to the no time pressure (NoTP, $n = 27$) and time pressure (TP, $n = 28$) groups, respectively. Testing was in a quiet room with 10 semienclosed cubicles containing standard PCs and 38.4 cm monitors; viewing distance was ~85 cm. Stimulus presentation was controlled with MATLAB using Psychophysics Toolbox extensions (Kleiner et al., 2007). Participants responded using the keyboard. Research in this article was approved by the Human Research Ethics Advisory Panel of the University of Málaga; participants provided informed consent and were treated in accordance with the Helsinki declaration. All reported normal or corrected-to-normal vision.

Stimuli and task. All materials, along with detailed instructions for their correct use, are freely available at https://osf.io/9x3tm/. Participants completed a trial-by-trial reward-learning task (see Luque et al., 2017), in which they played the role of space traders. Participants were told that on each trial they would be given a particular type of cookie and had to choose which of two aliens to trade it with. The alien would then give them one of three types of diamond, each worth a different number of points; participants' task was to earn as many points as possible.

Instrumental reward-learning trials consisted of S-R-O sequences (see Table 1). Each of these training trials began with a central fixation point presented for a random 200–300 ms, followed by one of four distinct "cookie" stimuli. These stimuli were easily discriminable colored circles (3.3° visual angle diameter), each containing several smaller circles of a different color (Figure 1A). After 800 ms, pictures of two aliens (~2.8° × 4°) appeared at either side of the screen (6.9° from the center). These aliens marked the two response options (R1 and R2, Table 1). Participants then pressed "q" or "p" to give the cookie to the alien on the left or right, respectively. Participants in the NoTP group had a time limit of 4,000 ms for responding, whereas participants in the TP group had to respond in under 500 ms. After a response, the screen blanked for 300 ms. If the participant had responded too slowly, the message "Time out, please respond faster" appeared for 1.5 s. If the participant had responded before the aliens appeared, "Too soon! No diamond for you" appeared.

If the response was timely, the outcome was presented: one of three diamonds (size 2.6° × 2.3°; yellow, blue or purple). The diamond appeared for 800ms, followed by a feedback screen. For reward-learning trials in no devaluation (No-Dev) blocks (see

Table 1
Experiment Design

S-R-O mapping			Outcome value		
S	→ R	→ O	No-Dev	Dev-O^{100}	Dev-O^{10}
S_1^{high}	→ R1	→ O^{100}	$\underline{+100}$	0	$\underline{+100}$
	→ R2	→ O^5	+5	$\underline{+5}$	+5
S_2^{high}	→ R1	→ O^5	+5	$\underline{+5}$	+5
	→ R2	→ O^{100}	$\underline{+100}$	0	$\underline{+100}$
S_1^{low}	→ R1	→ O^{10}	$\underline{+10}$	$\underline{+10}$	0
	→ R2	→ O^5	+5	+5	$\underline{+5}$
S_2^{low}	→ R1	→ O^5	+5	+5	$\underline{+5}$
	→ R2	→ O^{10}	$\underline{+10}$	$\underline{+10}$	0

Note. The first three columns show the stimulus-response-outcome (S-R-O) mappings that participants experienced. S^{high} and S^{low} denote high-value stimuli (those for which participants could earn 100 points) and low-value stimuli (those for which participants could earn at most 10 points) respectively. The four stimuli (two S^{high} and two S^{low}) were cookies differing in color. R1 and R2 denote Response 1 and Response 2; left or right button-presses depending on the counterbalancing condition. O^{100}, O^{10}, and O^5 denote the outcome of each response, typically worth 100, 10 and 5 points, respectively. The two high-value stimuli (denoted S_1^{high} and S_2^{high}) differed according to whether response R1 or R2 respectively was associated with the optimal response in No-Dev blocks (see below); ditto for the two low-value stimuli (S_1^{low} and S_2^{low}). The right-hand three columns of the table show the values associated with the different outcomes, during blocks in which neither outcome was devalued (No-Dev), blocks in which the high-value outcome was devalued (Dev-O^{100}), and blocks in which the 10-point outcome was devalued (Dev-O^{10}). The best possible outcome for each trial type is shown underlined.

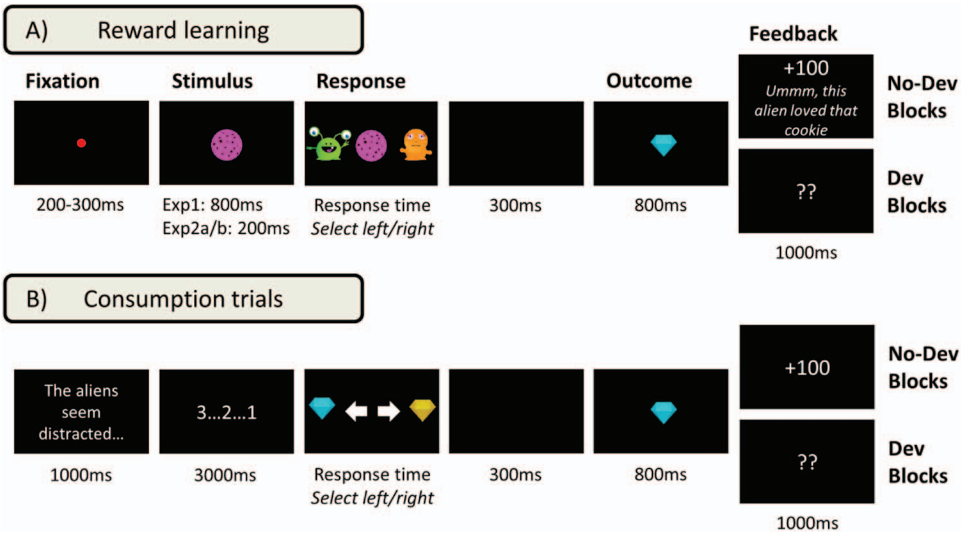

Figure 1. Paradigm used in all experiments. (A): Example of a reward-learning trial. In this example, the stimulus was a pink cookie, the participant selected a particular alien with which to trade this cookie and earned a blue diamond as a result. In No Devaluation (No-Dev) trial-blocks, feedback indicated whether the response was optimal ("Ummm, this alien loved that cookie" for optimal responses; "Puagh! This alien didn't like that cookie" for suboptimal responses), and the value of the diamond that had been earned (+100 points for optimal responses in trials with a high-value stimulus, S^{high}; +10 points for optimal responses in trials with a low-value stimulus, S^{low}; and + 5 points for all suboptimal responses). In Devaluation (Dev) blocks, feedback did not provide information regarding the correctness of the response or the value of the just-earned diamond. (B): Example of a consumption trial. Participants chose between the two diamonds, which had + 100 and + 10 points value. See the online article for the color version of this figure.

below), feedback showed the value of the just-earned diamond. One of the diamonds was worth 100 points (O^{100}), another was worth 10 points (O^{10}), and the third was worth 5 points (O^{5}).

Table 1 shows the S-R-O contingencies to be learned by participants. For each stimulus there was an optimal response. For the two high-value stimuli (S_1^{high} and S_2^{high} in Table 1), the optimal response led to the O^{100}; for the two low-value stimuli (S_1^{low} and S_2^{low}), the optimal response led to the O^{10}. For all stimuli, suboptimal responses led to the O^{5}. When participants made an optimal response (earning the O^{100} or O^{10} diamond), feedback on the diamond's value was accompanied by the message "Ummm, this alien loved that cookie"; for incorrect responses (earning the O^{5} diamond) the corresponding message was "Puagh! This alien didn't like that cookie." Feedback appeared for 1,000 ms, and the next trial then began after a blank intertrial interval of 800 ms.

To implement the outcome devaluation test, participants were instructed that the market values of the diamonds might change; therefore, they should pay attention to the ongoing value of each diamond. Participants were told the values of the diamonds in an on-screen message at the beginning of each trial-block. In blocks in which the high-value outcome was devalued (Dev-O^{100} blocks), O^{100} became worthless (0 points), whereas the other diamonds maintained their value. In blocks in which the low-value outcome was devalued (Dev-O^{10} blocks), the O^{10} was devalued to 0 points, whereas the other diamonds maintained their value. During both types of devaluation blocks (Dev blocks), feedback did not reveal the value of the diamonds or the optimality of the response ("Ummm/Puagh" message); instead, feedback was "??" on every trial (see Figure 1A). Trial-by-trial outcome values were hidden during devaluation blocks to prevent new associative learning based on the devalued outcome values (Tricomi et al., 2009; for a similar strategy, see Gillan, Otto, Phelps, & Daw, 2015). Participants were told that information about diamond values was unavailable during devaluation blocks, but this did not affect the diamonds they could earn or their values.

Thus, participants could still earn diamonds and hence gain points during the devaluation test. For trials involving a stimulus that could earn the not-devalued outcome (O^{10} in Dev-O^{100} blocks; O^{100} in Dev-O^{10} blocks), the optimal response remained the same as during previous instrumental training. By contrast, for trials with a stimulus previously associated with the now-devalued outcome, participants needed to switch their response to earn points. Consider, for example, S_1^{high} (see Table 1). In No-Dev and Dev-O^{10} blocks, the optimal response when presented with S_1^{high} was R1, since this earned 100 points. In Dev-O^{100} blocks, however, the best choice was to switch to response R2, which earned 5 points, rather than response R1, which earned the now-worthless O^{100}.

For each participant, the specific pictures used as stimuli (four cookies), responses (two aliens), and outcomes (three diamonds) were assigned to different roles for the experiment following a Latin square. The left/right position of the aliens was determined at random for each participant at the beginning of the experiment and then held constant subsequently.

Consumption trials (Figure 1B) were included in each block to assess participants' understanding of the current value of outcomes O^{100} and O^{10} (Gillan et al., 2015; Luque et al., 2017). Participants were instructed that sometimes the aliens were distracted, and participants could take a diamond without trading it for cookies. On these consumption trials, the message "The aliens seem distracted . . ." appeared for 1,000 ms, followed by a countdown (from 3 to 1) over 3 s. Then, the O^{100} and O^{10} diamonds appeared on either side of the screen (left/right positions determined randomly for each trial). Participants pressed "q" or "p" to take the left or right diamond, respectively. Responses made before the onset of the diamonds, or slower than 2,000 ms, led to "too fast" or "timeout" messages. The consumption trials continued as for No-Dev trials (but without the "Ummm/Puagh" messages).

Each block comprised 52 trials: 12 learning trials with each of the four stimuli and four consumption trials. Consumption trials always appeared as Trials 13, 26, 39, and 52 within each block. Training ran over three sessions on consecutive days, with each session comprising nine blocks. Sessions on Day 1 and Day 3 began with three No-Dev blocks. Blocks 4 and 5 were then Dev-O^{100} and Dev-O^{10} test blocks, with the order of these two devaluation tests determined randomly for each participant. Blocks 6 and 7 were No-Dev blocks. Finally, Blocks 8 and 9 were again Dev-O^{100} and Dev-O^{10} test blocks, presented in the reverse order from that used for Blocks 4 and 5. In the session on Day 2, all blocks were No-Dev blocks. Participants were reminded of instructions for the task at the beginning of each session.

Results

Data from pairs of stimuli associated with the same outcome value (e.g., S_1^{high} and S_2^{high}) were collapsed. Data were analyzed as a function of *amount of training*: Figure 1 shows data separately for Devaluation Tests 1 and 2 (the two devaluation tests of each type [Dev-O^{100} and Dev-O^{10}] on Day 1) and Tests 3 and 4 (the two tests of each type on Day 3). We corrected for multiple comparisons using the Holm-Bonferroni method; corrected $\tilde{p}$ values are shown accordingly. In all experiments, Greenhouse-Geisser corrected degrees of freedom are reported where data violated sphericity.

Consumption trials. Data from consumption trials during devaluation blocks allow us to verify that participants understood devaluation instructions. In consumption trials, participants chose between the O^{100} and O^{10} outcome. Optimal behavior was to select O^{10} in Dev-O^{100} blocks (because O^{100} was worthless), and O^{100} in Dev-O^{10} blocks. This was the clear pattern observed, with little change over the experiment (Figure 2A). These data were nonnormally distributed and were analyzed using nonparametric tests. Separate Friedman tests for each time-pressure condition (NoTP, TP) and devaluation-type (Dev-O^{100}, Dev-O^{10}) revealed no significant effect of amount of training (Tests 1–4) on the proportion of O^{100} selections, all $\tilde{p}$s $> .5$. Separate Mann–Whitney tests found no significant differences between NoTP and TP conditions in any of the devaluation tests, all $\tilde{p}$s $> .9$.

Having verified that participants understood the devaluation instructions appropriately in all tests, we turn to analyses of our primary dependent variables: (a) *response selections* and (b) *RT switch cost* during devaluation test blocks (analyses of response selections during No-Dev blocks are reported in the online supplementary materials).

Response selections. It is important to deconfound the possible effects of S-R habit formation on response selections during devaluation tests from effects that might be caused by other factors varying with amount of training (fatigue, task-familiarity, etc.). We therefore baseline-corrected data, using as baseline the data from the No-Dev block immediately preceding the devaluation test under analysis (see the report analyses of uncorrected data in the online supplementary materials). Specifically, our dependent variable was the proportion of responses in a devaluation block leading to the outcome which was devalued—that is, proportion of responses consistent with use of a habit—minus the proportion of responses leading to the same outcome for the baseline block. This subtraction was calculated independently for S^{high} and S^{low} trials (*stimulus value* factor). Values near 0 mean that responses during devaluation were similar to during no devaluation, that is, insensitivity to devaluation, as expected for habitual behavior. Values near -1 indicate effective switching of response selections because of the new outcome values, that is, sensitivity to outcome devaluation, as expected for goal-directed behavior.

The resulting data (Figure 2B) were analyzed in a Time Pressure × Amount of Training × Stimulus Value ANOVA which yielded main effects of time pressure, $F(1, 53) = 36.37$, $p < .001$, $\eta_p^2 = .41$, amount of training, $F(3, 92.78) = 26.56$, $p < .001$, $\eta_p^2 = .33$, and stimulus value, $F(1, 53) = 12.37$, $p = .001$, $\eta_p^2 = .19$, and a Significant Time Pressure × Amount of Training interaction, $F(3, 159) = 7.17$, $p < .001$, $\eta_p^2 = .12$. All other effects were nonsignificant ($ps > .1$). The significant effects arose because (a) participants in group TP were generally less sensitive to devaluation than those in group NoTP; (b) scores generally declined (indicating increasing sensitivity to devaluation) as training on S-R-O contingencies increased, with this effect being more pronounced for participants under time pressure; (c) scores were lower (indicating greater sensitivity to devaluation) for trials with high-value stimuli (S^{high}) than low-value stimuli (S^{low}). Findings (b) and (c) in particular are inconsistent with what we would expect if "incorrect" response selections reflected the operation of the habit system.

RT switch cost. *Response switches* were occasions on which participants correctly changed their response to avoid a devalued outcome; for example, for trials with S_1^{high}, response switches occurred when, during a Dev-O^{100} test, participants did not make the response that earned the (now-devalued) O^{100} (R1), but made the response that earned the O^5 (R2). RT was averaged across response switch trials separately for S^{high} and S^{low}. As for response selections, these data were then baseline-corrected (see the report analysis of uncorrected data in online supplementary materials') using RTs for trials with the same stimulus (S^{high}/S^{low}) when participants made the 'standard' response in the preceding No-Dev block (e.g., R1 for trials with S_1^{high}). We subtracted baseline RT from response-switch RT. Positive values of the resulting variable reflect a *RT switch cost* during the devaluation test, expressed in seconds. Larger values indicate more difficulty in overcoming pretrained response tendencies in devaluation blocks, consistent with greater interference from S-R habits.

Data (Figure 2C) were submitted to a 2 × 2 × 4 [Time Pressure × Amount of Training × Stimulus Value] ANOVA. This yielded main effects of time pressure, $F(1, 53) = 52.53$, $p < .001$, $\eta_p^2 = .50$, and amount of training, $F(3, 159) = 4.42$, $p = .008$, $\eta_p^2 = .08$, that were qualified by a Time Pressure × Amount of Training interaction, $F(3, 159) = 10.14$, $p < .001$, $\eta_p^2 = .16$ (all other effects were nonsignificant, $ps > .13$). Unlike for response selections, for RT switch cost the significant interaction reflected a different

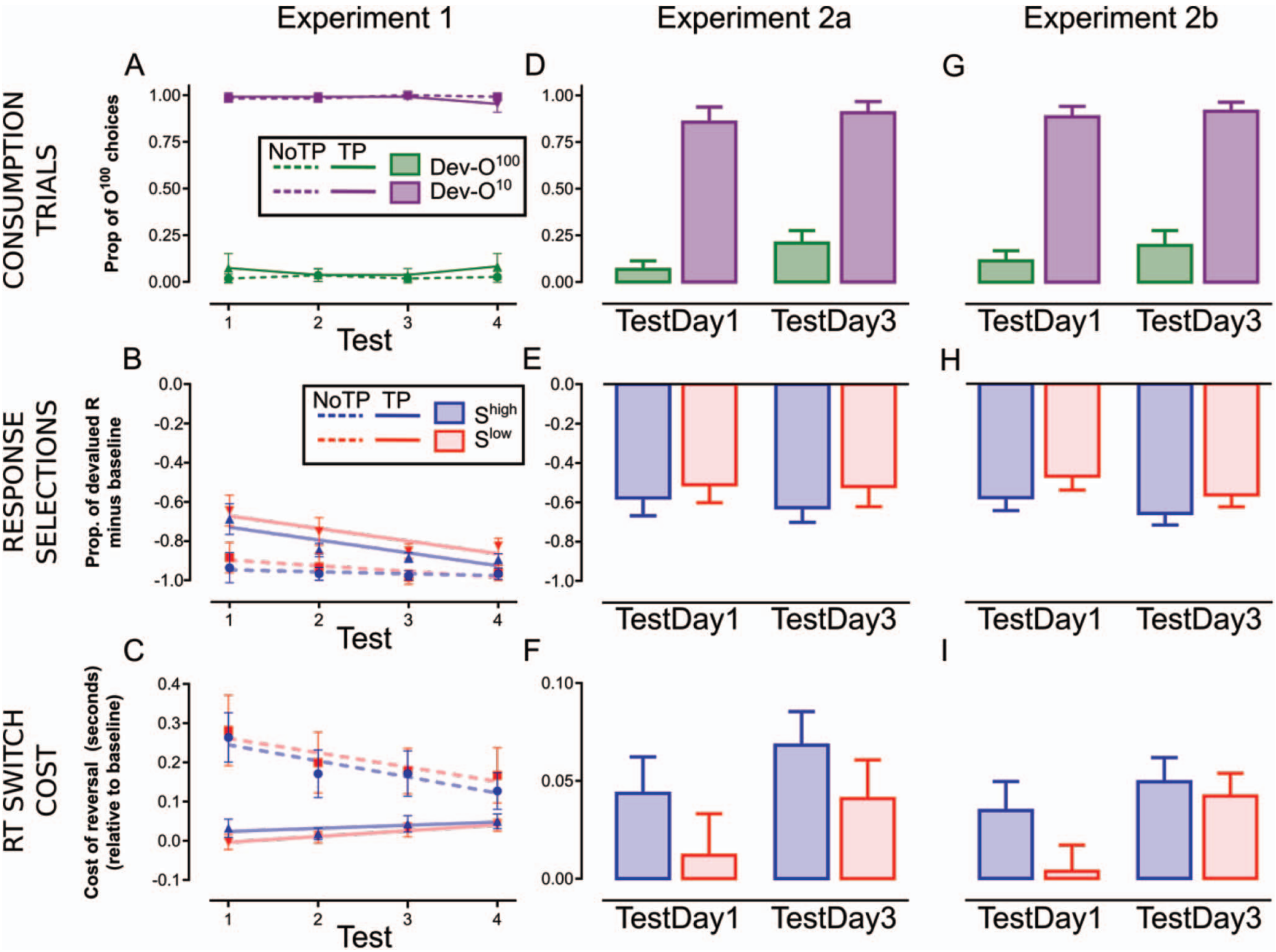

Figure 2. Main results. Top row (A, D, G): Consumption trials. Data show proportion of trials in which participants selected the high-value outcome O^{100} (+100 points) in consumption trials. Blue lines/bars show choices in Dev-O^{100} blocks; red lines/bars show choices in Dev-O^{10} blocks. Middle row (B, E, H): Response selections during devaluation. Figures show the proportion of suboptimal responses to the high-value (S^{high}) or low-value (S^{low}) stimulus during devalued blocks (relative to baseline, see main text). Larger scores indicate greater sensitivity to devaluation. Bottom row (C, F, I): Response times for optimal response switches—that is, responses avoiding the devalued outcome—during devaluation blocks (relative to baseline, see main text). Lower scores indicate less interference from previous instrumental training during devaluation. Left-hand column (A, B, C): Data from Experiment 1. TP and NoTP denote groups of participants trained under response time pressure (TP) and without time pressure (NoTP). X-axes (labeled "Test") represent devaluation tests conducted on Day 1 (Tests 1 and 2) and Day 3 (Tests 3 and 4). Lines in Panels B and C show best-fit lines for each condition over the course of learning. Middle and right-hand columns: Data from Experiments 2a and 2b, respectively. Error bars represent 95% confidence intervals. See the online article for the color version of this figure.

direction of the effect of amount of training in the two time-pressure conditions: RT switch cost reduced across training in group NoTP (indicating reduced interference from now-inappropriate responses, consistent with more goal-directed behavior), but increased across training in group TP (indicating greater interference from now-inappropriate responses, consistent with increasingly habitual behavior): see Figure 1C. To confirm this, we ran separate 2 × 4 [Amount of Training × Stimulus Value] ANOVAs for each time pressure condition.

In the NoTP group, ANOVA yielded a main effect of amount of training, $F(3, 81) = 7.94$, $p < .001$, $\eta_p^2 = .23$, with a significant linear trend, $F(1, 27) = 21.93$, $p < .001$, $\eta_p^2 = .45$, indicating a general reduction of RT switch cost across training. Other effects were not significant, $ps > .3$.

In the TP group, ANOVA yielded a main effect of stimulus value, $F(1, 26) = 7.56$, $p = .011$, $\eta_p^2 = .23$, with larger RT switch costs for high-value stimuli (S^{high}) than low-value stimuli (S^{low}). There was also a main effect of amount of training, *F*(3,

78) = 4.40, p = .013, η_p^2 = .15, with a significant linear trend, F(1, 26) = 7.91, p = .009, η_p^2 = .23, indicating an increase in RT switch cost across training. Other effects were not significant, p > .17. Hence for participants under time pressure, RT switch cost showed properties that would be expected of a valid measure of S-R habits: It increased as a function of amount of instrumental training and was larger when training was with a high-value outcome.

Experiments 2a and 2b

We postpone briefly discussion of the findings of Experiment 1 to consider a possible confound. Experiment 1 manipulated amount of training within-subjects. Consequently, differences in the amount of instrumental training were confounded with differences in the amount of experience with the outcome devaluation test, raising the possibility that the effect of amount of training might be produced by strategic adaptation to the devaluation test: for instance, participants in group TP may have learned to slow their responses without increasing their number of timeouts. Such a possibility falls within the predictions of sequential-sampling models of decision making (Forstmann, Ratcliff, & Wagenmakers, 2016). To avoid this potential confound, Experiments 2a and 2b instead manipulated amount of training between-subjects; these two experiments used the same design and procedure, with Experiment 2b (n = 95) being a better-powered replication of Experiment 2a (n = 47).

Method

Sample size determination. Effect size for the critical effect of amount of training in the time-pressure group was η_p^2 = .15 in Experiment 1. Lacking other data, we based our sample-size calculation for the (between-subjects) Experiment 2a on this previous (within-subjects) effect; this revealed 47 participants would provide power of .80. Experiment 2b was a better-powered replication of Experiment 2a in which we based sample size on the between-subjects effect of amount of training observed in Experiment 2a and aimed to achieve power ≥ .9. The resulting minimum sample size was 82.

Participants and apparatus. A total of 47 and 95 University of Málaga students participated in Experiments 2a and 2b, respectively, on the same terms as in Experiment 1. In Experiment 2, 24 participants were allocated to the TestDay1 group, and 23 to the TestDay3 group; in Experiment 3, 48 participants were allocated to TestDay1 and 47 to TestDay3. Apparatus was as for Experiment 1.

Stimuli and task. Stimuli and task were as in Experiment 1 with the following exceptions. In each trial, the stimulus (S; the cookie) was presented for only 800 ms before disappearing in the first block of each session, for 500 ms in the second block, and for 200 ms in subsequent blocks. Response time limit was 500 ms throughout (i.e., all participants completed the task under the same time pressure as Experiment 1's group TP).

Training was over three, 9-block sessions run on consecutive days, as in Experiment 1. Devaluation tests were included during the first day of training for participants in the TestDay1 group, and during the third day of training for the TestDay3 group. These devaluation tests were included as the sixth and seventh blocks within the session, in random order (Dev-O^{100} before Dev-O^{10} or vice versa). All other blocks were no-devaluation reward learning blocks.

Results

Consumption trials. Figures 2D and 2G show that participants' choices in consumption trials were highly sensitive to current outcome values during devaluation blocks. Mann–Whitney tests assessed the effect of *amount of training* (TestDay1 vs. TestDay3) on the proportion of O^{100} selections in consumption trials during Dev-O^{100} and Dev-O^{10} blocks. In Experiment 2a, there was a significant between-groups difference in Dev-O^{100} blocks (η_p^2 = .002): group TestDay3 selected the devalued outcome O^{100} more than group TestDay1. This effect, however, was not significant in Experiment 2b, η_p^2 = .214. There were no differences for Dev-O^{10} blocks in either experiment, η_p^2 > .4. The significant effect found in Experiment 2a for the Dev-O^{100} test could reflect relative inattention to the instructions regarding outcome values on Day 3. Notably, however, this effect was no longer significant in the better-powered Experiment 2b (or in Experiment 1), raising the possibility that the effect in Experiment 2a was a false positive.

Response selection. Response-selection data were baseline-corrected as for Experiment 1. Figures 2E and 2H show data from devaluation blocks (analyses of data from No-Dev blocks are reported in the online supplementary materials). For Experiment 2a, an Amount of Training (TestDay1 vs. TestDay3) × Stimulus Value (S^{high} vs. S^{low}) ANOVA yielded a significant main effect of stimulus value, F(1, 48) = 7.22, p = .010, η_p^2 = .13. For Experiment 2b, the same analysis yielded significant main effects of stimulus value, F(1, 93) = 24.65, p < .001, η_p^2 = .21 and amount of training, F(1, 93) = 4.91, p = .029, η_p^2 = .05. No other effects were significant (ps > .5). As in Experiment 1, scores were lower (participants were more sensitive to devaluation, consistent with greater goal-directed control) for trials with high-value stimuli (S^{high}) than low-value stimuli (S^{low}), and for group TestDay3 than group TestDay1 group (this effect reached significance in Experiment 2b).

RT switch cost. RT switch cost data were baseline-corrected as for Experiment 1 (Figures 2F and 2I). Data from three participants in Experiment 2a were excluded because they lacked data for at least one test. RT switch cost was analyzed by Amount of Training × Stimulus Value ANOVAs. In Experiment 2a, this yielded main effects of amount of training, F(1, 45) = 6.28, p = .016, η_p^2 = .12, and stimulus value, F(1, 45) = 14.89, p < .001, η_p^2 = .25, but no interaction, p > .7. In Experiment 2b there were main effects of amount of training, F(1, 93) = 14.48, p < .001, η_p^2 = .14, and stimulus value, F(1, 93) = 10.26, p < .001, η_p^2 = .10, and a significant interaction, F(1, 93) = 3.96, p = .050, η_p^2 = .04. In both experiments, RT switch cost was larger (greater interference from instrumental training) for participants who had undergone longer training, and for trials with a high-value stimulus than trials with a low-value stimulus, as would be expected of a valid measure of S-R habits. These results replicate findings of Experiment 1. The interaction for Experiment 2b revealed some evidence of a larger effect of stimulus value during Day 1 than Day 3. However, given that this interaction reached conventional significance only in Experiment 2b (p =

.050), and was nonsignificant in Experiments 1 and 2a, we do not draw strong claims from it.

General Discussion

That actions can eventually become habits is an idea that has exerted a deep influence across a range of scientific areas (see Verplanken, 2018). Repetition is a crucial factor for the formation of habits; indeed, it has been argued that manipulations of the amount of training are the clearest way to study the operation of the habit system (Watson & de Wit, 2018). However, recent experiments in humans using overt response selections as a measure of habits have not found evidence of increased habit strength with extended training (de Wit et al., 2018). If we are unwilling to jettison the idea that habits should become stronger with repetition, then the implication is that response selections may not (always) provide a good measure of habits. But we need a valid index of habits in human research to allow for translation from rich neurocognitive animal models—and such an index should be sensitive to manipulations of the amount of training. In the current study we tried to solve this problem by investigating an alternative behavioral marker for the operation of the habit system.

We examined the effect of overtraining on the "standard" measure of habits (overt response selections of a devalued outcome), and a novel measure, *RT switch cost*: the increase in RT for changing a previously optimal response selection. We also manipulated two other factors that should influence habits: outcome value (more valuable outcomes should increase learning rate and therefore S-R habit formation: e.g., Marien et al., 2018; Rescorla & Wagner, 1972) and time pressure (which should hinder goal-directed control).

Response selections did not show the patterns that would be expected for a valid measure of habits. Notably, overtraining did not increase frequency of response selection errors; indeed, all experiments revealed the opposite trend, and this reached significance in Experiments 1 and 2b. The implication is that response selections became more goal-directed—rather than more habitual—with extended training in this task. By contrast, the RT switch cost measure fulfilled expectations for a valid measure of habits. In all experiments, RT switch cost increased with extended training, and was larger for associations (previously) involving high-value outcomes. Finally, these effects were evident only when participants were placed under time pressure.

The implication is that, whereas S-R habits may not become strong enough to generate suboptimal overt response selections during devaluation in this procedure, they can still make their presence felt in terms of interference with optimal, goal-directed responding. Our findings are consistent with the idea that changing from a response that was previously optimal during instrumental training to a different response that is currently optimal during devaluation, puts into conflict the goal-directed system (favoring the now-optimal response) and the habit system (favoring the previously optimal response)—and this conflict manifests as a slowed response.

The idea of response conflict is central to the procedure that we used here. Our task examined a situation in which participants had a choice of two possible responses on each trial (select the left alien or select the right alien), and our index of habits was effectively a measure of the ease with which participants switched from a now-inappropriate action to an alternative, now-more-appropriate action. This procedure mirrors the situation in many real-world examples of learned, instrumental behavior in which there are alternative actions to which people should switch. For example, habit formation has been extensively studied in the context of health psychology (for review, see Verhoeven & de Wit, 2018). In this literature, there are always competing response options in terms of food choices (e.g., apple vs. cookie), alcohol choices (alcohol vs. soft drinks), activity choices (e.g., watch TV vs. go for a run), and so on. As such, a measure of habit formation based on response conflict could be straightforwardly adapted for use in such cases.

More generally, response conflict may occur in many situations in which a preponderant S-R link has to be suppressed because a new response is required. There is an extensive literature on how our cognitive control system deals with such situations in a variety of paradigms (Aron, Robbins, & Poldrack, 2004; Eagle, Bari, & Robbins, 2008; Izquierdo & Jentsch, 2012). Notably, some of this research has indicated a role of S-R learning in producing persistent RT switch costs. For instance, in the field of instrumental task-switching, Wylie and Allport (2000; see also MacLeod & Dunbar, 1988) studied the RT cost arising when participants had to switch between two tasks, as compared to repeating the same task. They used a Stroop task, in which participants had to switch between naming the color of a word presented on-screen and reading the word. Participants switched from one task to the other on every second trial (so task order was AABBAABB . . .), so task-switches were predictable. Trials were separated by 1,000 ms therefore participants had some time to prepare for the upcoming task. Nevertheless, a switch cost was observed in that RT was greater on trials in which the task had just switched (the first trial of each "run" of two) than on trials in which the task repeated (the second trial of each run). Importantly, Wylie and Allport found that the magnitude of the RT cost on "switch trials"' was a function of the degree of activation of S-R characteristics of the previous task. Specifically, in their Experiments 2 and 3, they manipulated the proportion of color-naming versus word-reading trials. Results indicated that the interfering effect of each task on the next depended on this relative amount of training: greater experience of the prior task led to more interference with the subsequent task. In the language of the current research, we could say that Wylie and Allport's amount-of-experience manipulation acted to change the relative strength of the S-R links corresponding to each Stroop task (color-naming vs. word-reading), and that they indexed the strength of these S-R links by using an RT switch cost measure.

Clearly this prior task-switching research bears important similarities to the work we have presented here. It is tempting to hypothesize that habits produced by reward learning (measured by assessing the effects of changing the reward value of outcomes in outcome-devaluation tests) and more general examples of response automatization—such as those typically studied in the task-switching literature—share the same psychological and neural systems (Dezfouli & Balleine, 2013). However, previous research on task-switching such as that by Wylie and Allport (2000; see also MacLeod & Dunbar, 1988) did not measure *habits* in the strict sense. In particular, these studies did not change the *value* of the outcome that was associated with the now-inappropriate task; instead, they changed the outcome itself (so that a response which on a preswitch trial would yield "correct" feedback now yielded "error" feedback). This distinction might be fundamental since recent research suggests that response automatization may in some cases be specifically sensitive to outcome devaluation (Garr & Delamater, 2019). In general, these ideas suggest that more research is needed to explore the possible relationships between the processes underlying response automatization and formation of S-R habits as a consequence of instrumental reward learning.

As we mentioned in the introduction, previous approaches and procedures have been successful in investigating different aspects of the goal-directed and habit systems in humans and nonhuman animals (e.g., their neural basis, their relationship with certain psychopathologies, etc.: see Verplanken, 2018). However, previous procedures developed to study habits in humans have failed to find a reliable effect of overtraining (de Wit et al., 2018), and this has stymied translational work because an increase in habit strength with overtraining is essentially axiomatic to this area. Some of these previous procedures bear important similarities to the approach used in the current study: for example, in the "fabulous fruit game" used by de Wit and colleagues (de Wit, Corlett, Aitken, Dickinson, & Fletcher, 2009; de Wit et al., 2012; Watson et al., 2018), the S-R-O relationships that participants must learn are arbitrary (e.g., pictures of fruits and keyboard responses), outcome devaluation is implemented using instructions and the test is often conducted under time pressure—just as in the current study. This suggests that the reason these prior studies (e.g., de Wit et al., 2018) did not observe an increase in habit strength as a function of overtraining is likely because of the measure of habits that was used—frequency of overt responses that yield now-devalued outcomes. The implication of the current findings is that in these existing paradigms were adapted to instead (or at least additionally) examine RT switch costs, they might provide a more sensitive and reliable measure of habit formation that is sensitive to overtraining manipulations.

In summary, our findings suggest that RT switch cost can provide a reliable marker of the operation of the habit system in humans, and potentially a more sensitive marker of such than the more standard measure of response selections of a devalued outcome. We stress, however, that we are not arguing that RT switch cost provides a *pure* measure of habit-strength to the exclusion of all other factors. For example, we noted earlier that strategic adaptation to the devaluation test may also influence switch cost in the within-subjects Experiment 1 (though this confound is ruled out in the between-subjects Experiments 2a and 2b; see also Dickinson et al., 1995). Our argument here is that RT switch cost provides a reliable *correlate* of habit strength—which is what is needed for future habit research in humans.

We have argued that our new technique for measuring habits, based on RT switch cost, provides an important advance on previous methods based on response selections, which do not show the sensitivity to overtraining that would be expected of habitual behavior. The implication is that it may be interesting to revisit previous research on the habit system, using this new strategy. For instance, future research could record neuroimaging data while participants complete the RT switch cost procedure. Such experiments might offer valuable further information about brain areas and networks involved in progressive habit formation, and could, for example, verify and complement previous findings (de Wit et al., 2012; Delorme et al., 2016; Tricomi et al., 2009). Further work could use the RT switch cost procedure to investigate differences in habit formation across various psychopathologies, because it has been claimed that dysfunction of the balance between goal-directed and habitual behavior is centrally implicated, for example, in disorders belonging to the impulsive-compulsive spectrum (Gillan et al., 2014; Marsh et al., 2004), and some forms of Parkinson's disease (Jahanshahi et al., 2015). These patient groups shown an increased frequency of overt responses that yield now-devalued outcomes (relative to healthy control participants), after a modest amount of training. However, because these data were not compared with behavior following prolonged training, it is hard to determine if differences in sensitivity to outcome devaluation between patients and healthy controls reflect dysfunction of the habit and/or the goal-directed system (Watson & de Wit, 2018); that is, whether patients have an unusually strong habit system, or an unusually weak goal-directed system. Our new procedure for measuring habit formation offers a valuable tool for probing these sorts of important questions further.

In the present study, we attempted to expose the role of the role of the habit system in instrumental reward learning by combining two different approaches: (a) we aimed to increase the strength of habits (and hence the influence of the habit system), by increasing the amount of training and the reward value of outcomes; and (b) we aimed to reduce the influence of the goal-directed system, by imposing time pressure. Future research might benefit from an approach which considers the separable effects of these (or similar) manipulations in order to isolate and distinguish between different reward-learning processes. For instance, fMRI measures could be recorded during a task similar to that used in our Experiment 1. Under conditions that should result in a strengthened habit system (e.g., during the third day of training), we would expect greater activation of the neural system supporting habits (e.g., the putamen: Knowlton & Patterson, 2016); under conditions in which the operation of the goal-directed system is hindered (e.g., under time pressure), we might expect a reduction in the activity of regions that have been implicated in goal-directed reward learning (e.g., orbitofrontal cortex: Valentin, Dickinson, O'Doherty, & Wolf, 2007). Lastly, the balance between the activity of these regions should predict the magnitude of RT switch cost during devaluation.

Finally, recent research suggests that learning about reward can produce "habit-like" patterns in the early perceptual and attentional processes that underlie information-gathering and prioritization (Anderson, 2016; Le Pelley, Mitchell, Beesley, George, & Wills, 2016; Luque et al., 2017). Future experiments measuring RT switch cost could investigate the ways in which these perceptual and attentional processes interact with mechanisms of learning and decision-making in the acquisition of habits, in healthy participants and in clinical populations in which habit-like attentional processes may be dysfunctional (see Albertella et al., 2017; Anderson, Faulkner, Rilee, Yantis, & Marvel, 2013).

Conclusion

The human goal-directed learning system is very effective at overriding preexisting habits when necessary. We have shown that the interfering effect of habits can nevertheless be detected by analyzing response times for goal-directed actions. This new marker for the activation of the habit system can be used in future human research that focuses on habit formation in healthy and clinical populations.

References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 34,* 77–98. http://dx.doi.org/10.1080/14640748208400878

Albertella, L., Copeland, J., Pearson, D., Watson, P., Wiers, R. W., & Le Pelley, M. E. (2017). Selective attention moderates the relationship between attentional capture by signals of nondrug reward and illicit drug use. *Drug and Alcohol Dependence, 175,* 99–105. http://dx.doi.org/10.1016/j.drugalcdep.2017.01.041

Anderson, B. A. (2016). The attention habit: How reward learning shapes attentional selection. *Annals of the New York Academy of Sciences, 1369,* 24–39. http://dx.doi.org/10.1111/nyas.12957

Anderson, B. A., Faulkner, M. L., Rilee, J. J., Yantis, S., & Marvel, C. L. (2013). Attentional bias for nondrug reward is magnified in addiction. *Experimental and Clinical Psychopharmacology, 21,* 499–506. http://dx.doi.org/10.1037/a0034575

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8,* 170–177. http://dx.doi.org/10.1016/j.tics.2004.02.010

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,* 48–69. http://dx.doi.org/10.1038/npp.2009.131

Bhanji, J. P., & Delgado, M. R. (2014). The social brain and reward: Social information processing in the human striatum. *WIREs Cognitive Science, 5,* 61–73. http://dx.doi.org/10.1002/wcs.1266

Braun, S., & Hauber, W. (2013). Acute stressor effects on goal-directed action in rats. *Learning & Memory, 20,* 700–709. http://dx.doi.org/10.1101/lm.032987.113

Brembs, B., Lorenzetti, F. D., Reyes, F. D., Baxter, D. A., & Byrne, J. H. (2002). Operant reward learning in Aplysia: Neuronal correlates and mechanisms. *Science, 296,* 1706–1709. http://dx.doi.org/10.1126/science.1069434

De Houwer, J., Tanaka, A., Moors, A., & Tibboel, H. (2018). Kicking the habit: Why evidence for habits in humans might be overestimated. *Motivation Science, 4,* 50–59.

Delorme, C., Salvador, A., Valabrègue, R., Roze, E., Palminteri, S., Vidailhet, M., . . . Worbe, Y. (2016). Enhanced habit formation in Gilles de la Tourette syndrome. *Brain: A Journal of Neurology, 139,* 605–615. http://dx.doi.org/10.1093/brain/awv307

de Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., & Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *The Journal of Neuroscience, 29,* 11330–11338. http://dx.doi.org/10.1523/JNEUROSCI.1639-09.2009

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., . . . Gillan, C. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology: General, 147,* 1043–1065. http://dx.doi.org/10.1037/xge0000402

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *The Journal of Neuroscience, 32,* 12066–12075. http://dx.doi.org/10.1523/JNEUROSCI.1088-12.2012

Dezfouli, A., & Balleine, B. W. (2013). Actions, action sequences and habits: Evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Computational Biology, 9*(12), e1003364. http://dx.doi.org/10.1371/journal.pcbi.1003364

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior, 22,* 1–18. http://dx.doi.org/10.3758/BF03199951

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. *Animal Learning & Behavior, 23,* 197–206. http://dx.doi.org/10.3758/BF03199935

Eagle, D. M., Bari, A., & Robbins, T. W. (2008). The neuropsychopharmacology of action inhibition: Cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology, 199,* 439–456. http://dx.doi.org/10.1007/s00213-008-1127-6

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223–241. http://dx.doi.org/10.1177/1745691612460685

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology, 67,* 641–666. http://dx.doi.org/10.1146/annurev-psych-122414-033645

Garr, E., & Delamater, A. R. (2019). Exploring the relationship between actions, habits, and automaticity in an action sequence task. *Learning & Memory, 26,* 128–132. http://dx.doi.org/10.1101/lm.048645.118

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., . . . Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry, 75,* 631–638. http://dx.doi.org/10.1016/j.biopsych.2013.02.002

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective & Behavioral Neuroscience, 15,* 523–536. http://dx.doi.org/10.3758/s13415-015-0347-6

Heerey, E. A. (2014). Learning from social rewards predicts individual differences in self-reported social ability. *Journal of Experimental Psychology: General, 143,* 332–339. http://dx.doi.org/10.1037/a0031511

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15,* 534–539. http://dx.doi.org/10.1111/j.0956-7976.2004.00715.x

Hyman, S. E. (2005). Addiction: A disease of learning and memory. *The American Journal of Psychiatry, 162,* 1414–1422. http://dx.doi.org/10.1176/appi.ajp.162.8.1414

Izquierdo, A., & Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology, 219,* 607–620. http://dx.doi.org/10.1007/s00213-011-2579-7

Jahanshahi, M., Obeso, I., Rothwell, J. C., & Obeso, J. A. (2015). A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. *Nature Reviews Neuroscience, 16,* 719–732. http://dx.doi.org/10.1038/nrn4038

Kapur, S., Mizrahi, R., & Li, M. (2005). From dopamine to salience to psychosis—linking biology, pharmacology and phenomenology of psychosis. *Schizophrenia Research, 79,* 59–68. http://dx.doi.org/10.1016/j.schres.2005.01.003

Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 12868–12873. http://dx.doi.org/10.1073/pnas.1609094113

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3? *Perception, 36,* 1–16. Retrieved from https://nyuscholars.nyu.edu/en/publications/whats-new-in-psychtoolbox-3

Knowlton, B. J., & Patterson, T. K. (2016). Habit formation and the striatum. In R. E. Clark & S. Martin (Eds.), *Behavioral neuroscience of learning and memory* (Vol. 37, pp. 275–295). http://dx.doi.org/10.1007/7854_2016_451

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron, 81,* 687–699. http://dx.doi.org/10.1016/j.neuron.2013.11.028

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, 142,* 1111–1140. http://dx.doi.org/10.1037/bul0000064

Liljeholm, M., Dunne, S., & O'Doherty, J. P. (2015). Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control. *European Journal of Neuroscience, 41,* 1358–1371. http://dx.doi.org/10.1111/ejn.12897

Lin, P.-Y., Wood, W., & Monterosso, J. (2016). Healthy eating habits protect against temptations. *Appetite, 103,* 432–440. http://dx.doi.org/10.1016/j.appet.2015.11.011

Luque, D., Beesley, T., Morris, R. W., Jack, B. N., Griffiths, O., Whitford, T. J., & Le Pelley, M. E. (2017). Goal-directed and habit-like modulations of stimulus processing during reinforcement learning. *The Journal of Neuroscience, 37,* 3009–3017. http://dx.doi.org/10.1523/JNEUROSCI.3205-16.2017

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 126–135. http://www.ncbi.nlm.nih.gov/pubmed/2963892. http://dx.doi.org/10.1037/0278-7393.14.1.126

Marien, H., Custers, R., & Aarts, H. (2018). Understanding the formation of human habits: An analysis of mechanisms of habitual behaviour. In *The Psychology of Habit* (pp. 51–69). Cham, Switzerland: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-97529-0_4

Marsh, R., Alexander, G. M., Packard, M. G., Zhu, H., Wingard, J. C., Quackenbush, G., & Peterson, B. S. (2004). Habit learning in Tourette syndrome: A translational neuroscience approach to a developmental psychopathology. *Archives of General Psychiatry, 61,* 1259–1268. http://dx.doi.org/10.1001/archpsyc.61.12.1259

Morris, R., Bornstein, A., & Shenhav, A. (2019). *Goal-directed decision making: Computations and neural circuits*. Cambridge, MA: Academic Press.

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual Review of Psychology, 68,* 73–100. http://dx.doi.org/10.1146/annurev-psych-010416-044216

Patterson, T. K., & Knowlton, B. J. (2018). Subregional specificity in human striatal habit learning: A meta-analytic review of the fMRI literature. *Current Opinion in Behavioral Sciences, 20,* 75–82. http://dx.doi.org/10.1016/j.cobeha.2017.10.005

Pessoa, L. (2015). Multiple influences of reward on perception and attention. *Visual Cognition, 23,* 272–290. http://dx.doi.org/10.1080/13506285.2014.974729

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Ripollés, P., Marco-Pallarés, J., Hielscher, U., Mestres-Missé, A., Tempelmann, C., Heinze, H.-J., . . . Noesselt, T. (2014). The role of reward in word learning and its implications for language acquisition. *Current Biology, 24,* 2606–2611. http://dx.doi.org/10.1016/j.cub.2014.09.044

Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2012). Simultaneous glucocorticoid and noradrenergic activity disrupts the neural basis of goal-directed action in the human brain. *The Journal of Neuroscience, 32,* 10146–10155. http://dx.doi.org/10.1523/JNEUROSCI.1304-12.2012

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *The Journal of Neuroscience, 29,* 7191–7198. http://dx.doi.org/10.1523/JNEUROSCI.0979-09.2009

Smeets, T., van Ruitenbeek, P., Hartogsveld, B., & Quaedflieg, C. W. E. M. (2019). Stress-induced reliance on habitual behavior is moderated by cortisol reactivity. *Brain and Cognition, 133,* 60–71.

Soares, J. M., Sampaio, A., Ferreira, L. M., Santos, N. C., Marques, F., Palha, J. A., . . . Sousa, N. (2012). Stress-induced changes in human decision-making are reversible. *Translational Psychiatry, 2*(7), e131–e131. http://dx.doi.org/10.1038/tp.2012.59

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience, 29,* 2225–2232. http://dx.doi.org/10.1111/j.1460-9568.2009.06796.x

Valentin, V. V., Dickinson, A., O'Doherty, J. P., & Wolf, O. (2007). Determining the neural substrates of goal-directed learning in the human brain. *The Journal of Neuroscience, 27,* 4019–4026. http://dx.doi.org/10.1523/JNEUROSCI.0564-07.2007

Verhoeven, A., & de Wit, S. (2018). The role of habits in maladaptive behaviour and therapeutic interventions. In B. Verplanken (Ed.), *The psychology of habit* (pp. 285–303). Cham, Switzerland: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-97529-0_16

Verplanken, B. (Ed.) (2018). *The psychology of habit*. Cham, Switzerland: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-97529-0

Watson, P., & de Wit, S. (2018). Current limits of experimental research into habits and future directions. *Current Opinion in Behavioral Sciences, 20,* 33–39. http://dx.doi.org/10.1016/j.cobeha.2017.09.012

Watson, P., van Wingen, G., & de Wit, S. (2018). Conflicted between goal-directed and habitual control, an fMRI investigation. *Eneuro, 5*(4), e0240-18.2018. http://dx.doi.org/10.1523/ENEURO.0240-18.2018

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://dx.doi.org/10.1016/j.appet.2014.04.005

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology, 67,* 289–314. http://dx.doi.org/10.1146/annurev-psych-122414-033417

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs." *Psychological Research, 63,* 212–233. http://dx.doi.org/10.1007/s004269900003

Received April 3, 2019
Revision received September 4, 2019
Accepted October 15, 2019 ■