Computational insights from a novel habit induction protocol

Sarah Oh (sarahoh@berkeley.edu)
Department of Psychology
University of California, Berkeley
Berkeley, CA 94720 USA

Anne G. E. Collins (annecollins@berkeley.edu)
Department of Psychology, Helen Wills Neuroscience Institute
University of California, Berkeley
Berkeley, CA 94720 USA

Abstract

Habits – automatic behavioral patterns formed through repetition – are essential for daily functioning, but can also lead to inflexible behavior. While crucial for understanding both adaptive and maladaptive decision-making, studying habits' computational and neural mechanisms has been challenging due to limited laboratory experiments demonstrating overtraining-induced inflexibility. We developed a novel task with features designed to encourage participants to engage goal-directed (GD) control between trials (interleaving extensively- and minimally-practiced contexts), then naturally release control within trials (hierarchical multi-step trial structure and opportunities to self-correct). Results showed that overtrained participants displayed stronger biases toward behaviors learned in extensively-practiced contexts, evidenced by higher Habit Index values at early response times. This effect decreased at later response times, suggesting participants could override habitual impulses with GD control. Our computational model, characterizing behavior as a mixture of reinforcement-learned policies, reproduced observed behavioral patterns, suggesting that habits can be viewed as goal-directed deployment of overtrained policies.

Keywords: habits; human behavior; reinforcement learning; computational modeling

Introduction

A habit is a behavior that has become ingrained through repetition, so that it is elicited automatically by cues that have historically signaled the availability of a desired outcome (Wood & Rünger, 2016). For frequently encountered situations, habits can be adaptive and efficient: at a familiar intersection during your commute, you might automatically enter the left-turn lane and eventually arrive at work with minimal cognitive effort. However, when goals or environmental conditions change, this automaticity can become maladaptive: when driving to a doctor's appointment instead of work, you might habitually turn left at the usual intersection, even though the quickest route requires a right turn. Understanding habits is thus crucial both for explaining adaptive behavior and for understanding psychiatric dysfunction, where habitual control may become dysregulated (Gillan et al., 2011; Everitt & Robbins, 2016; Uniacke, Timothy Walsh, Foerde, & Steinglass, 2018).

Classic animal studies using outcome devaluation paradigms demonstrated that initially goal-directed behaviors gradually become automatized with extended practice, persisting even when they no longer lead to desired outcomes (Adams, 1982; Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995). Lesion studies further revealed that goal-directed behaviors and these gradually-formed, outcome-insensitive habits are supported by distinct neural circuits (see Lingawi, Dezfouli, & Balleine, 2016 for a review). However, translating these findings to humans has proven challenging: experiments using outcome devaluation have typically failed to detect effects of practice amount on subsequent behavioral inflexibility (Tricomi, Balleine, & O'Doherty, 2009; de Wit et al., 2018; Pool et al., 2022; Gera et al., 2023), and many tasks designed to measure habits in humans appear more sensitive to goal-directed control than to putative habit measures (Gillan, Otto, Phelps, & Daw, 2015; Friedel et al., 2014; Sjoerds et al., 2016; Linnebank, Kindt, & de Wit, 2018).

One potential explanation is that human participants too readily engage goal-directed control in laboratory tasks, masking any strengthening of habitual impulses from overtraining. Supporting this interpretation, Luque, Molinero, Watson, López, and Le Pelley (2020) observed that over multiple days of training, participants become slower to respond correctly to devalued stimuli, indicating the presence of habitual impulses that conflict with, but are ultimately suppressed by, goal-directed control. Schwabe and Wolf (2009) demonstrated that acute stress leads to increased devaluation insensitivity, suggesting that limiting cognitive resources can reveal habit-like inflexibility. Indeed, Hardwick, Forrence, Krakauer, and Haith (2019) used a forced response time (RT) paradigm to show that, following a contingency reversal, participants overtrained over multiple days commit more habit-consistent slips of action than minimally-trained participants, primarily when forced to respond quickly ($<$600ms), suggesting that habits can be observed when goal-directed control has had insufficient time to prepare a response.

Following the insight that limiting participants' ability to exert goal-directed control can help uncover their habits, we developed a task inspired by Hardwick et al. (2019) that incorporates features designed to encourage participants to naturally release goal-directed control, allowing us to observe overtraining effects within a single experimental session without artificially constraining control. Behavioral and computational results support our prediction that this approach would enable us to reveal both habit formation and, separately, control, in a short lab-based experiment, opening the door to future research on human habits.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

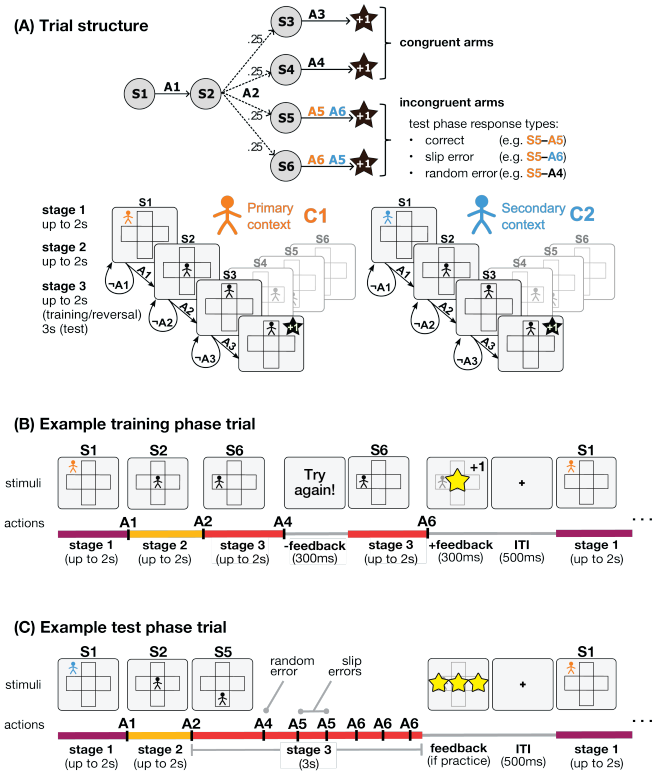

Figure 1: **Experimental design**. **(A)** Each trial has three stages: stage 1 presents state S1, stage 2 presents S2, and stage 3 presents a random terminal state (maze arms S3-S6, equiprobable). Correct choices advance to the next stage, with correct terminal state responses earning reward (1 point); incorrect responses or timeouts (>2s) return to the same state with feedback; an exception is test phase stage 3, when no feedback is shown for correct or incorrect responses. Context is indicated by avatar color at stage 1 of each trial, and hidden for stages 2-3. The correct actions for two of the terminal states ("incongruent" arms) are reversed across C1/C2; the other two terminal states ("congruent" arms) have consistent state-action mappings across C1/C2. A test phase response to an incongruent arm is either the correct response, a slip error (correct for other context), or a random error (any other error). **(B)** Example stimuli and actions during training phase. **(C)** Example stimuli and actions during the test phase. For stage 3, feedback (number of correct responses) is shown at the end of the 3s response period only for practice trials.

Task Design

Participants learned through trial and error to select correct actions from six available keypresses (A1-A6) in response to six stimuli (S1-S6) in two contexts (Fig. 1). **Training phase**: In Context 1 (C1, training phase), indicated by an orange avatar, participants learned correct keypresses for six locations on a cross-shaped maze. Each trial proceeded in three stages: 1) Starting outside the maze, the correct keypress moves the avatar to the maze center, 2) At the center, the correct keypress moves the avatar to a random maze arm, and 3) At the maze arm, the correct keypress earns a reward. **Reversal phase**: After either 200 (moderately-trained) or 600 (overtrained group) C1 trials, participants were introduced to C2, indicated by a blue avatar, wherein the stimulus–action associations for two maze arms (incongruent arms) were reversed with respect to C1. **Test phase**: After reaching a performance criterion (min 40, max 100 trials) in C2, participants experienced 48 C1 and 48 C2 trials, randomly interleaved. Participants were given 3s to freely press keys to collect multiple points in stage 3 of test phase trials, without feedback.

Throughout the task, the avatar appeared in black during stages 2 and 3, so context information was only available in stage 1 of each trial, forcing participants to hold the current context in mind while executing stages 2 and 3 during the test phase. The frequent switching between extensively- and minimally-practiced contexts during the test phase required participants to engage GD control to respond appropriately for the current context, allowing us to test their ability to respond flexibly (or conversely, how habitually they behaved). Meanwhile, the hierarchical multi-step trial structure (where stage 3 responses depend on context information from stage 1) and the allowance for multiple response attempts were designed to encourage participants to naturally release GD control during trials, enabling us to observe habitual impulses in their stage 3 responses. In what follows, we provide evidence from behavioral analyses and computational modeling that our task design is able to induce habit-like inflexible behavior as a function of training duration.

Participants

The task was administered online to undergraduate students from the University of California, Berkeley, recruited through the UC Berkeley Research Participation Program (RPP) participant pool. Students received course credit for their participation. 160 participants (118 female, 40 male, 2 other; age: *mean*=21.1, *sd*=3.40, *min*=18, *max*=43) completed the task. Of these, 17 participants were excluded based on self-reported exclusion criteria – namely, indicating in their post-task survey that they used external aids (e.g. post-it notes on the computer screen) to perform better on the task, or that they believed their data should be excluded from analysis – leaving 143 participants.

We applied the following exclusion criteria to identify careless or inattentive participants: 1) more than five missed trials during either the training (C1) or reversal (C2) phase (indicating low engagement in learning the associations), 2) failure to meet the C2 training criterion (correct first response in four out of five most recent exposures to each maze location) within 100 trials (indicating failure to learn C2, making C2 test phase errors uninterpretable), 3) fewer than two responses on more than 25% of trials in any context and congruent/incongruent arm type despite instructions to collect as many points as possible during the test phase (indicating a failure to understand the multi-response design of test

phase trials), and 4) evidence of satisficing strategies in the test phase (indicating failure to engage with the task in good faith, and undermining interpretability). Satisficing behaviors included random spamming (more than 1 unique key press in the last five responses on at least half of trials in both congruent and incongruent arms), strategic spamming (more than 1 unique key press in the last five responses on at least half of trials in incongruent arms, but not in congruent arms; typically such participants alternated between the two reversed keys), and making reliable errors (all trials ending on an incorrect response for any context and maze arm). In participants who passed all other exclusion criteria, these satisficing behaviors point to a failure to apply knowledge acquired in the training and reversal phases during the test phase.

70 participants were excluded based on these exclusion criteria (54.9% excluded from 200-trial group, 41% excluded from 600-trial group, 49% excluded overall), leaving 37 and 36 participants in the 200-trial and 600-trial groups, respectively. All included participants completed the task in under 1 hour (median 13m, range 10-23m in 200-trial group; median 32m, range 25-45m in 600-trial group).

Behavior

Defining a behavioral index of habit strength

We classified responses following the approach in Hardwick et al. (2019), designating an erroneous response as a "slip error" when it matched the correct response for the alternate context, and as a "random error" otherwise (Fig. 1A,C). Since C1 was practiced more extensively than C2, we should expect a higher rate of C1-consistent slip errors in C2, and a lower rate of C2-consistent slip errors in C1. We define a behavioral Habit Index (HI) quantifying this asymmetry:

$$\mathrm{HI} = \left(P^{C2}_{\mathrm{slip}} - P^{C2}_{\mathrm{rand}}/2\right) - \left(P^{C1}_{\mathrm{slip}} - P^{C1}_{\mathrm{rand}}/2\right) \quad (1)$$

where P_{slip} (slip error rates) are baseline-corrected by $P_{\mathrm{rand}}/2$ (the estimated rate of random errors per maze arm, since 2 maze arms contribute to P_{rand}), yielding action slip rate estimates. This baseline correction was aimed at subtracting out slip errors that are not true slips of action toward the other context, but may instead be random responses driven by confounding factors that vary with overtraining, such as fatigue. HI thus controls for overall attention or engagement differences between contexts.

If, as proposed by Hardwick et al. (2019), habitual processes rapidly prepare well-practiced responses that are subsequently overridden by goal-directed control, we would expect HI to be positive at early RTs and decrease toward zero at later RTs. We further expected that: 1) habit strength increases with training, so early-RT HI values should be higher in overtrained compared to moderately-trained participants; 2) participants in both groups successfully learn C2 contingencies and can deploy comparable levels of goal-directed control, so HI should eventually converge to similar values across groups at later RTs. Thus, we hypothesized that HI would be larger in overtrained relative to moderately-trained participants at early, but not at later RTs.

We conducted a bootstrapping analysis to ensure that any effect of overtraining on HI or its components was not observed merely by chance. For each of 10,000 iterations, we generated a bootstrap sample by drawing $N = 35$ participants with replacement from each group (200- and 600-trial); we then resampled each sampled participant's trials with replacement within each context (C1, C2). To generate a null distribution, we swapped the group labels for a random half of participants from each group. For each null distribution sample, we computed for each participant the average (across trials) proportion of slip errors falling in each combination of time bin and context, yielding $P(slip)$, and similarly to compute $P(rand)$; these average proportions were used to compute $P(slip) - P(rand)/2$ values for each participant. The contrast $C2 - C1$ was computed for each of the three measures. Statistical significance of overtraining effects was evaluated by comparing the empirical group difference (600- minus 200-trial) in a measure of interest computed from the full dataset with those computed from the null distribution samples, calculating the proportion of null distribution samples that exceeded the empirical difference.

Overtraining increases the Habit Index

Overtrained participants exhibited higher HI relative to moderately-trained participants, especially earlier in the 3s

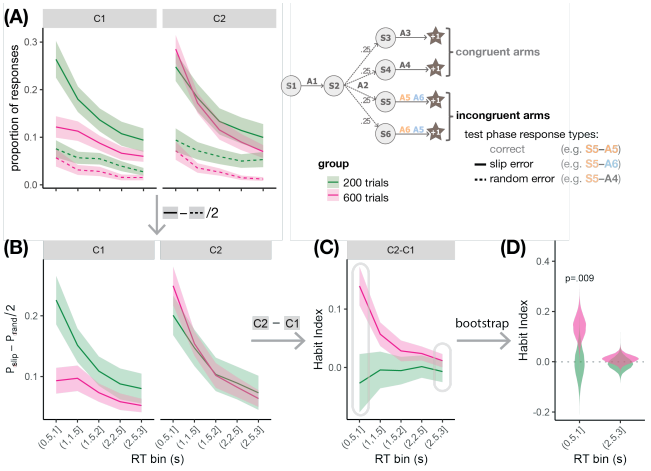

Figure 2: **Computation of the Habit Index**. Lines and ribbons (A-C) indicate mean ± s.e.m. across participants. **(A)** Proportion of responses that are slip errors (solid) and random errors (dashed) by group (200-trial, green; and 600-trial, pink), context (C1, left; and C2, right), and RT (0.5s bins spanning 0-3s). **(B)** Action slip rate by group, context, and RT. **(C)** The strength of the tendency to have a higher action slip rate in C2 vs. C1 (i.e. Habit Index, HI), by group and RT. **(D)** Distribution of HI in bootstrapped samples for early (0.5-1s) and late (2.5-3s) RT bins; p-values computed by comparing empirical group difference in HI to a null distribution of group differences in HI with shuffled group labels.

context	RT bin (s)	emp.	%>0	$p_{null>emp.}$
C2	(0.5,1]	0.057	76.0	.154
-C1	(0.5,1]	0.128	99.3	.006
C2-C1	**(0.5,1]**	**0.170**	**98.1**	**.009**
C2-C1	(2.5,3]	0.016	75.6	.275

Table 1: Summary of bootstrap analyses testing for an overtraining effect on action slip rate ($P(slip) - P(rand)/2$) in various contexts (or contrasts thereof) and RT bins. Column %>0: proportion of bootstrap samples (without shuffled group labels) greater than 0. Column $p_{null>emp.}$: significance of the empirical group difference with respect to a null distribution with shuffled group labels. Predicted effect highlighted in bold.

response period (Fig. 2C). The reliability of the effect of overtraining on HI at early RTs was confirmed by bootstrapping (Fig. 2D, Table 1). The effect was driven by overtrained participants exhibiting both a higher (although not significantly higher) baseline-corrected slip error rate in C2, and a significantly lower baseline-corrected slip error rate in C1, relative to moderately-trained participants (Fig. 2B, Table 1).

These behavioral results indicate that our task successfully induced habit-like behavior that varied with training duration. To better understand the computational mechanisms underlying these effects, we developed formal models of participants' response generation process.

Computational Modeling

We developed separate computational models to capture participants' choices during learning (training and reversal phases) and after learning (test phase). The models were refined through an iterative model comparison and validation process (Wilson & Collins, 2019): first, for each participant, parameter values that optimized the model's prediction of their choices were obtained by maximum likelihood estimation in MATLAB; the best-fitting parameters were used to simulate data from the model; behavioral patterns in the real and simulated data were compared in order to identify additional features that might improve the model's ability to capture participants' behavior; finally, model versions were compared for their quantitative (according to the Akaike Information Criterion, AIC, which penalizes models for complexity) and qualitative (based on visualizations of simulated data) fit to actual data.

Training / reversal phase model

Participants' acquisition of stimulus–action associations in C1 and C2 were modeled using a reinforcement learning (RL) model that gradually learns the expected values of each of the keypress actions for each maze location, updating its value estimates based on the difference between predicted and received outcomes. The base model represents the expected values of the 6 key-press actions available in each of the terminal 4 maze locations as a 4-by-6 "Q-table". Separate Q-tables are learned for contexts C1 and C2. At each stage 3 decision, indexed by t, the learning agent selects an action, a_t, by converting the 6 action values stored for the current maze location, s_t, in the Q-table for the current context, C_t, into action probabilities via a softmax function with an inverse choice temperature parameter β controlling how sensitive the agent is to relative expected values when choosing an action (i.e. how deterministic its choices are):

$$P(a|s_t, C_t) \propto exp(\beta \times Q_{C_t}[s_t, a]) \tag{2}$$

The agent observes the outcome of the selected action (reward $R_t = 1$ if correct action, 0 otherwise), and then updates its estimated value of taking the selected action from the current state, stored in the current context's Q-table, toward the observed outcome. The magnitude of this update is the reward prediction error (the difference between outcome R_t and the expected value of the selected action a_t according to the Q-table), scaled by a learning rate parameter α controlling how strongly new observations influence the agent's expected value estimates:

$$Q_{C_t}[s_t, a_t] \leftarrow Q_{C_t}[s_t, a_t] + \alpha \times (R_t - Q_{C_t}[s_t, a_t]) \tag{3}$$

Through model comparison and model validation (Wilson & Collins, 2019), we discovered three additional features that improved the base model's quantitative (as assessed by AIC) and qualitative fit to participants' training phase choices: 1) a weight to initialize Q_{C2} to a mixture of Q_{C1} and a uniform Q-table, reflecting participants' assumption that the new context is similar to the old one, 2) separate learning rates α_{C1} and α_{C2} for C1 and C2, and 3) a multiplier applied to the learning rate to obtain a counterfactual learning rate (used to update values of unchosen actions for the current state, and values of the selected action for other states, away from the observed outcome R_t).

When generating data from the training phase model using the likelihood-maximizing parameters for each participant, the model was able to capture the learning curves and error patterns over learning well, especially later in learning for both C1 and C2 (Fig. 3A,B). This was an important prerequisite for fitting the test phase model, which assumes that the Q-tables learned for C1 and C2 by the end of the training and reversal phases guide responding during the test phase.

Test phase model

We modeled participants' test phase choices as being generated from a weighted mixture of the Q-tables learned (by the training/reversal phase model) for each context:

$$\begin{aligned} Q_{mix}(C_t, RT) &= w(C_t, RT) \times Q_{C1} + \\ &\quad (1 - w(C_t, RT)) \times Q_{C2} \end{aligned} \tag{4}$$

At response times RT corresponding to participants' actual response times, the agent samples an action $a_{t,RT}$ by converting the values in the mixed Q-table for the current state s_t

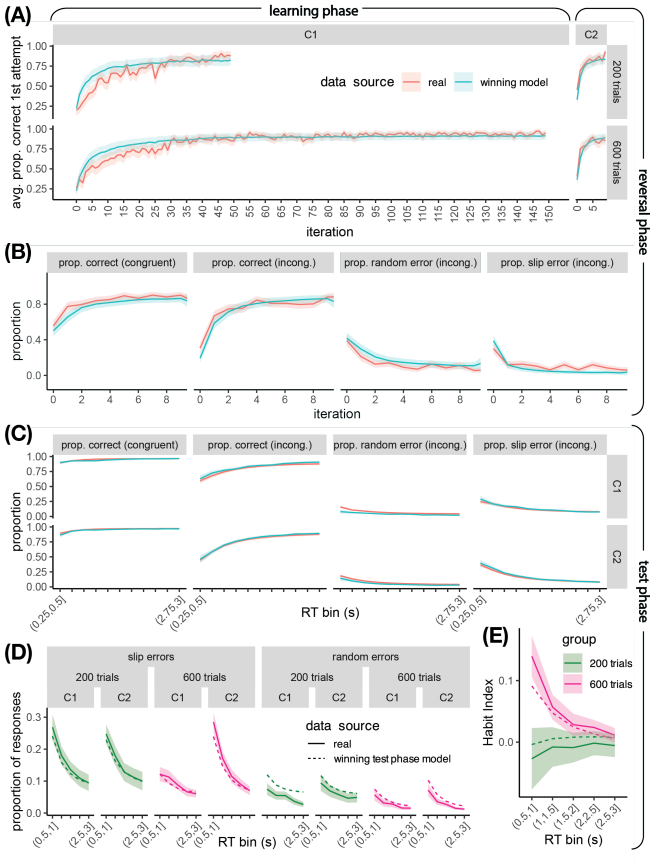

Figure 3: **Computational model validation**. Solid lines and ribbons indicate mean ± s.e.m. **(A)** Learning curves: probability of correct first response over training and reversal phases (peach: participants; turquoise: model simulations). **(B)** Response proportions (correct, slip errors, and random errors) during reversal phase. **(C)** Test phase response proportions by RT bin. **(D)** Slip and random error rates by context, RT, group (200-trial: green; 600-trial: pink; solid: data; dashed: model). **(E)** HI computed from components in (D).

into 6 action probabilities, via a softmax with inverse choice temperature parameter β_{test}. $w(C_t, RT)$ captures the degree of reliance on the Q-table learned for the current vs. the other context, as a function of current context C_t and time RT in the 3s response period. We compared multiple versions of $w(C_t, RT)$, including linear, exponential, and stepwise functions, based on their quantitative fit (as assessed by AIC) and ability to produce qualitative patterns observed in the data (Fig. 3C-E). Ultimately, the best-fitting model used two exponential curves with unique initial (w_{I1},w_{I2}) and final (w_{F1},w_{F2}) values for each context, and a rate constant (k, shared between contexts) governing the rate of exponential saturation from w_I to w_F. We further found that the model's tendency to overestimate random error probabilities at later RTs was alleviated by allowing β_{test} to change linearly from $RT = 0$ to $RT = 3$.

When generating data from the winning test phase model, the model was able to accurately reproduce the average participant's probability of making correct responses, slip errors and random errors across the 3s response window (Fig. 3C). While the effect of overtraining on the Habit Index generated by the model was weaker than that observed in the behavioral data, the expected qualitative pattern (larger HI in the overtrained group relative to the moderately-trained group at early RT, but not at late RTs) was clearly present (Fig. 3E).

A two-way mixed ANOVA testing the effect of group (between-participant) and early/late RT (within-participant) on the contrast between the weights on the incorrect Q-table when in C2 versus C1 revealed a significant main effect of RT, reflecting a tendency for participants in both groups to make slips of action toward C1 and away from C2 early in the response period, but not later; there was also an interaction between group and RT, driven by moderately-trained, but not overtrained, participants decreasing their bias toward C1 over the 3s response period (Fig. 4A).

Discussion

A central challenge in habit research is demonstrating that behavioral inflexibility truly reflects habit formation. While inflexibility is commonly observed in laboratory studies with humans, researchers have struggled to show that it reliably increases with extended practice (de Wit, Ridderinkhof, Fletcher, & Dickinson, 2013; de Wit et al., 2018; De Houwer, Tanaka, Moors, & Tibboel, 2018; Pool et al., 2022; Gera et al., 2023). Without this evidence, it remains unclear whether inflexibility stems from gradually-formed habits or simply reflects other factors, such as weak goal-directed control.

The present work addresses this challenge through a novel experimental paradigm that demonstrates increased behavioral inflexibility as a function of training duration. To quantify habit strength, we developed a Habit Index (HI) that captures two complementary aspects of habitual behavior: the tendency to inappropriately express an extensively-practiced response (i.e. inflexibility), and the enhanced ability to execute that response when it is appropriate (i.e. fluency). Using this measure, we found that overtrained participants showed higher HI values at early response times compared to moderately-trained participants, indicating stronger habit formation with extended practice. Crucially, this effect diminished at later response times, when participants could engage goal-directed control to override habitual responses. This temporal pattern – stronger habits at early RTs but successful override at later RTs – suggests that our results reflect genuine habit strengthening rather than a general degradation of goal-directed control in overtrained participants.

Our computational modeling approach provided additional insights into these behavioral findings. We developed a model that frames test phase behavior as arising from a weighted mixture of Q-tables learned during initial training. This model suggests that behavioral inflexibility emerges from the competition between these learned policies, with the extensively-practiced policy exerting stronger influence early in the response period. While the model successfully reproduced the qualitative pattern of higher HI in overtrained par-

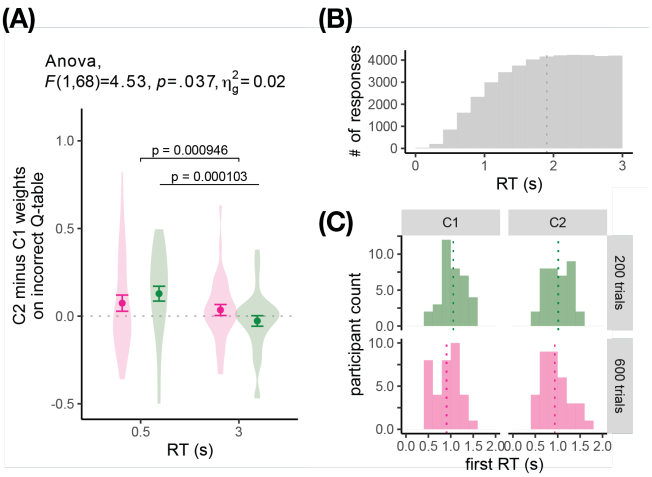

Figure 4: **Model parameters**. **(A)** Mixed ANOVA testing the effect of group (between-) and early/late RT (within-participant) on C2 minus C1 contrast of $1-w$ (weight on incorrect Q-table), and significant post-hoc t-tests. **(B)** Histogram of test phase stage 3 response RTs. **(C)** Histogram of participants' average first stage 3 response RTs, by group and context; vertical dashed lines indicate average values.

ticipants at early RTs, the magnitude of this effect was smaller than observed in the behavioral data. This quantitative discrepancy likely stems from a methodological limitation: parameter estimates were disproportionately influenced by later responses, which were far more numerous in our dataset (Fig. 4B). Nevertheless, the model's ability to capture the key behavioral pattern suggests that slips of action may arise from an initial selection of the context-appropriate policy at stage 1, followed by interference from the more extensively practiced C1 policy during execution – at least until goal-directed control can implement the correct response.

Interestingly, our modeling revealed that overtrained participants did not show a larger initial bias toward C1 compared to moderately-trained participants (Fig. 4A). Instead, the group differences appear to be driven by response timing: overtrained participants responded more quickly in both contexts, significantly so in C1 (Fig. 4C; $F(1,69) = 5.74$, $p = .019$, $\eta^2_G = 0.077$ in C1; $F(1,69) = 1.41$, $p = .239$, $\eta^2_G = 0.020$ in C2). This pattern aligns with theoretical accounts like that of Hardwick et al. (2019), who proposed that practice accelerates habitual response preparation while leaving the speed of goal-directed control unchanged. While they revealed habitual control by explicitly constraining response times, our task achieved a similar effect through participants' self-initiated response patterns.

Our study has several limitations that suggest directions for future work. While our test phase model successfully reproduces group differences through a combination of individual model parameters and yoking simulated responses to participants' actual RTs, it does not explain the process by which overtraining comes to alter subsequent test phase responding. Developing more explanatory models presents challenges: many existing computational frameworks for habits are unsuitable for our task structure – the model-based/model-free RL framework (Daw, Gershman, Seymour, Dayan, & Dolan, 2011) requires predictable state transitions for model-based planning to be viable, while free-operant models (Perez & Dickinson, 2020) assume continuous rather than trial-based responses. One potential direction would be to extend our computational model to jointly capture choices and response times, drawing inspiration from the response preparation framework proposed by Hardwick et al. (2019), where practice speeds up habitual response preparation while goal-directed response preparation remains fixed. Such a model could potentially account for both the increase in the Habit Index at early RTs and the speeding of responses with overtraining. Another promising direction is to explore value-free learning models (Miller, Shenhav, & Ludvig, 2019; Collins, 2024), which propose that habits can form through direct reinforcement of actions by repetition and contextual cues, independent of outcome values.

Another limitation of the present work is the high exclusion rate: close to half of participants were excluded from analysis, and around half of these were excluded for satisficing behaviors indicating that failed to apply their knowledge of the C1/C2 associations during the test phase. In a follow-up experiment (Exp 1b in Oh & Collins, 2025) we modified the task by adding a small cost to responding incorrectly in order to encourage participants to engage properly with the task during the test phase. This brought the proportion of participants excluded for satisficing down from 27% to 3%, but weakened our ability to detect the overtraining effect, highlighting a tension between maintaining task engagement and encouraging the release of control. Future work could explore task designs that optimize this tradeoff.

Although these challenges point to important directions for future research, our work has made substantial progress on the central challenge we set out to address: demonstrating and explaining training-dependent behavioral inflexibility in humans. We achieve this by eliciting flexible goal-directed control across trials, through frequent context switching, and facilitating a natural release of control within trials, through a hierarchical multi-step trial structure and unlimited response opportunities. Moreover, our computational framework, which successfully captures the temporal dynamics of habitual responding through a weighted mixture of learned policies, provides a novel account of how habits emerge from the interaction between training history and response timing. By providing both a reliable method for inducing and measuring training-dependent behavioral inflexibility in humans and a computational framework for understanding it, our paradigm creates new opportunities for investigating the neural and computational bases of habits in both healthy and clinical populations. Understanding these mechanisms could ultimately inform interventions for conditions where habitual control becomes dysregulated, from addiction to obsessive-compulsive disorder.

References

Adams, C. D. (1982, May). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *34*(2), 77–98. doi: 10.1080/14640748208400878

Collins, A. (2024, July). *RL or not RL? Parsing the processes that support human reward-based learning.* OSF. doi: 10.31234/osf.io/he3pm

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., … Gillan, C. M. (2018, July). Shifting the Balance Between Goals and Habits: Five Failures in Experimental Habit Induction. *Journal of Experimental Psychology. General*, *147*(7), 1043–1065. doi: 10.1037/xge0000402

de Wit, S., Ridderinkhof, K. R., Fletcher, P. C., & Dickinson, A. (2013, November). Resolution of outcome-induced response conflict by humans after extended training. *Psychological Research*, *77*(6), 780–793. doi: 10.1007/s00426-012-0467-3

De Houwer, J., Tanaka, A., Moors, A., & Tibboel, H. (2018, March). Kicking the habit: Why evidence for habits in humans might be overestimated. *Motivation Science*, *4*(1), 50–59. doi: 10.1037/mot0000065

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995, June). Motivational control after extended instrumental training. *Animal Learning & Behavior*, *23*(2), 197–206. doi: 10.3758/BF03199935

Everitt, B. J., & Robbins, T. W. (2016). Drug addiction: updating actions to habits to compulsions ten years on. *Annual review of psychology*, *67*, 23–50.

Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., & Schlagenhauf, F. (2014). Devaluation and sequential decisions: Linking goal-directed and model-based behavior. *Frontiers in Human Neuroscience*, *8*.

Gera, R., Bar Or, M., Tavor, I., Roll, D., Cockburn, J., Barak, S., … Schonberg, T. (2023, May). Characterizing habit learning in the human brain at the individual and group levels: A multi-modal MRI study. *NeuroImage*, *272*, 120002. doi: 10.1016/j.neuroimage.2023.120002

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015, September). Model-based learning protects against forming habits. *Cognitive, Affective & Behavioral Neuroscience*, *15*(3), 523–536. doi: 10.3758/s13415-015-0347-6

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., & de Wit, S. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *American Journal of Psychiatry*, *168*(7), 718–726.

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, *3*(12), 1252–1262.

Lingawi, N. W., Dezfouli, A., & Balleine, B. W. (2016). The Psychological and Physiological Mechanisms of Habit Formation. In *The Wiley Handbook on the Cognitive Neuroscience of Learning* (pp. 409–441). John Wiley & Sons, Ltd. doi: 10.1002/9781118650813.ch16

Linnebank, F. E., Kindt, M., & de Wit, S. (2018, September). Investigating the balance between goal-directed and habitual control in experimental and real-life settings. *Learning & Behavior*, *46*(3), 306–319. doi: 10.3758/s13420-018-0313-6

Luque, D., Molinero, S., Watson, P., López, F. J., & Le Pelley, M. E. (2020, August). Measuring habit formation through goal-directed response switching. *Journal of Experimental Psychology: General*, *149*(8), 1449–1459. doi: 10.1037/xge0000722

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019, March). Habits without values. *Psychological Review*, *126*(2), 292–311. doi: 10.1037/rev0000120

Oh, S., & Collins, A. (2025, January). *Naturally disengaging control to reveal habits.* OSF. doi: 10.31234/osf.io/zdur5

Perez, O. D., & Dickinson, A. (2020, November). A theory of actions and habits: The interaction of rate correlation and contiguity systems in free-operant behavior. *Psychological Review*, *127*(6), 945–971. doi: 10.1037/rev0000201

Pool, E. R., Gera, R., Fransen, A., Perez, O. D., Cremer, A., Aleksic, M., … O'Doherty, J. P. (2022, January). Determining the effects of training duration on the behavioral expression of habitual control in humans: A multilaboratory investigation. *Learning & Memory*, *29*(1), 16–28. doi: 10.1101/lm.053413.121

Schwabe, L., & Wolf, O. T. (2009, June). Stress Prompts Habit Behavior in Humans. *Journal of Neuroscience*, *29*(22), 7191–7198. doi: 10.1523/JNEUROSCI.0979-09.2009

Sjoerds, Z., Dietrich, A., Deserno, L., de Wit, S., Villringer, A., Heinze, H.-J., … Horstmann, A. (2016). Slips of Action and Sequential Decisions: A Cross-Validation Study of Tasks Assessing Habitual and Goal-Directed Action Control. *Frontiers in Behavioral Neuroscience*, *10*.

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*(11), 2225–2232.

Uniacke, B., Timothy Walsh, B., Foerde, K., & Steinglass, J. (2018). The role of habits in anorexia nervosa: where we are and where to go from here? *Current Psychiatry Reports*, *20*, 1–8.

Wilson, R. C., & Collins, A. G. (2019, November). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. doi: 10.7554/eLife.49547

Wood, W., & Rünger, D. (2016, January). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314. doi: 10.1146/annurev-psych-122414-033417