bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Skill Acquisition and Habit Formation as Distinct Effects of Practice

Robert M Hardwick*[1,2], Alexander D Forrence[1], John W Krakauer[1,3], and Adrian M Haith[1]

Departments of Neurology[1] and Neuroscience[3], Johns Hopkins University, Baltimore, USA

[2]Current Affiliation: Movement Control & Neuroplasticity Research Group, KU Leuven, Belgium

Keywords: Skill, Response Latency, Habit, Motor Learning

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Practice improves the speed at which we can perform a task, but also leads to habitual behavior. Behavioral, computational, and neurobiological evidence has suggested that these two effects of practice might be related; however, it remains unclear whether skill improvement and habit formation are two aspects of the same learning process, or are separate processes that occur in parallel. Using a visuomotor association task in human participants, we directly assessed the effects of practice on both the speed of response selection, and whether or not response selection became habitual. We found that response selection could become fully habitual within four days of practice. In contrast, the speed of response selection improved continuously with practice over twenty days. We conclude that skill learning occurs largely independently of habit formation, suggesting a distinct neural basis.

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

An essential aspect of many skills is the ability to quickly and accurately select an appropriate movement[1]. For instance, table tennis players must not only be able to execute shots with good technique, they must also be able to judge the flight of the ball and select which shot to play - all in less than a quarter of a second. Rapid action selection is critical to many everyday activities such as typing, driving, and playing sports. Selection speed, measured through the reaction time, is also used as the primary measure of performance in many prominent paradigms used to study motor learning, including sequence learning[1,2], and learning arbitrary visuomotor associations[3].

How might it be possible to improve the speed at which actions can be selected? Action selection typically depends on time-consuming computations to determine the appropriate response, but it is not always necessary to perform these costly computations every time a stimulus is encountered. By storing the outcome of common computations, the selection problem can be reduced to a direct, pre-computed stimulus-response relationship. A downside, however, of such pre-computation is inflexibility. If a change in task goals requires selection of a different action, the pre-computed stimulus-response policy will persist, leading to habitual selection of outdated responses[4,5]. The idea of storing a pre-computed policy therefore suggests a potential link between improved skill, and the tendency to become habitual following practice.

Many parallels have previously been drawn between skills and habits: both are thought to involve a qualitative change in the underlying representation of behavior[4–8], both appear to recruit the basal ganglia[9,10] and the acquisition of both is associated with dopaminergic, reward-based learning mechanisms[11,12]. Habit learning is even often studied in rodents as a model of skill acquisition[13–15].

Despite these parallels, selection skill could also improve through alternative means; for instance, by learning to execute necessary computations more efficiently[16,17], or through more

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

rapid perceptual processing of stimuli. If so, improvements in the speed of response selection could occur independently of whether a response becomes habitually selected.

Here, we examined the effects of practice on the speed and habitualness of action selection in a visuomotor association task. Participants were trained to press specific buttons as quickly as possible in response to arbitrary visual stimuli. These associations were practiced for various durations ranging from a minimal amount to 20 days. We assessed improvements in the latency of action selection through changes in the reaction time required to respond to a stimulus. To determine whether action selection had become *habitual*, we switched the stimulus-response contingencies for a subset of stimuli - if response selection were habitual, one would expect participants to persist with the initially learned mapping[5,18]. However, assessing habitual response selection is complicated by the fact that behavior is generated through an evolving competition between goal-directed and habitual processes[4,19]. A habitually selected response might be only transiently prepared, and later replaced by a more deliberately determined response. Indeed, limiting preparation time has proven to be an effective means of prohibiting deliberate, goal-directed processes from influencing behavior[17,20–22]. We therefore predicted that imposing limited preparation time would unmask such latent habitually selected responses.

In order to more precisely quantify the effects of practice, we devised a computational model that related the speed and potentially habitual nature of selecting each action to the time-varying likelihood of expressing each potential response. Fitting this model to data allowed us to identify the effects of practice on speed of response selection and the extent to which responses were selected habitually. Consequently, we were able to determine that the speed of response selection improved independently from the development of habitual response selection.

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

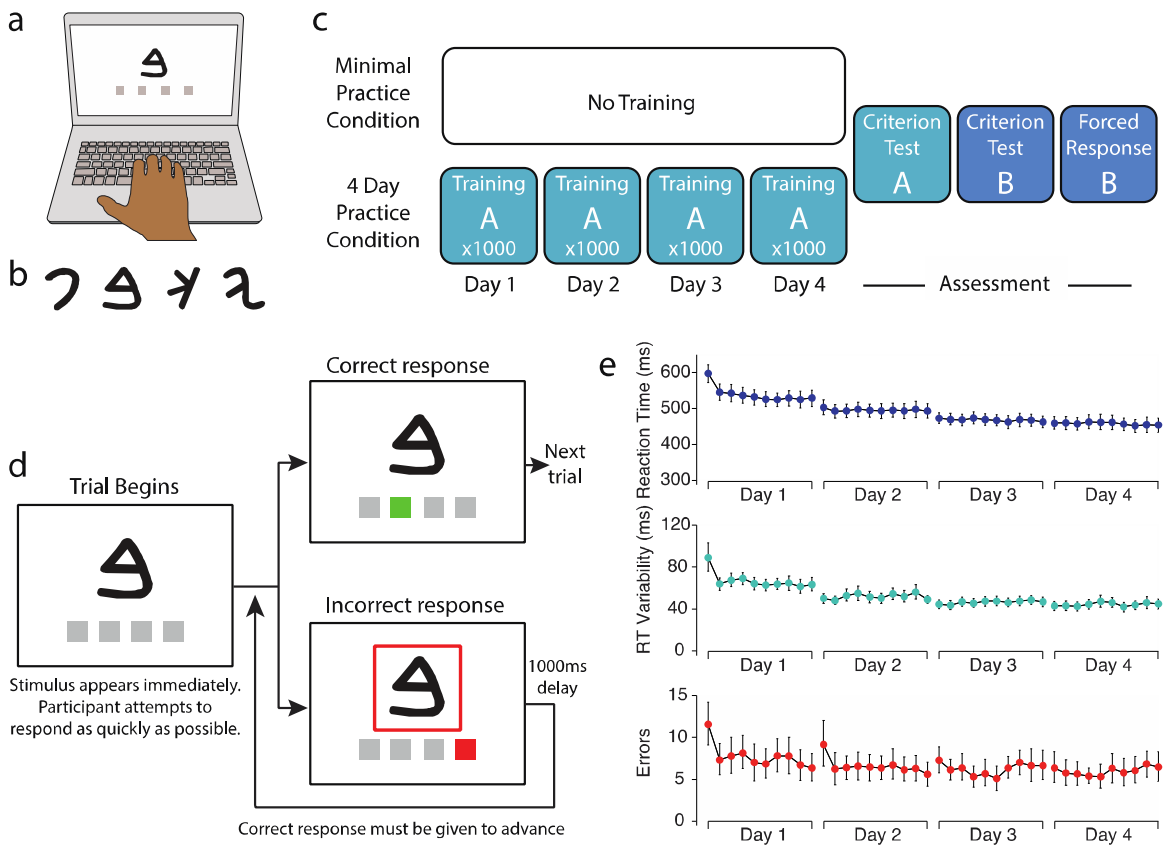


***Figure 1**. Task and training schedule for Experiment 1. a) Experimental setup. Participants responded to the appearance of a visual stimulus that cued them to press a keyboard button with a specific finger. b) Example stimuli (letters of the Phoenician alphabet). c) Experiment 1 overview. In the Minimal Practice condition participants went straight into an assessment session in which they were briefly trained on original mapping A, then briefly trained on revised mapping B (see Figure 2), after which they completed forced-response trials under this new mapping (see Figure 3). In the 4 Day Practice condition, participants completed 4,000 trials on the original mapping (1,000 trials per day) prior to this assessment procedure. Participants completed the two conditions in a counterbalanced order. d) Trial structure of the reaction time based training condition. Participants attempted to complete blocks of 100 trials as quickly as possible, incurring a time penalty for incorrect responses. e) Data from the 4-Day Practice*

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

condition. Participants' reaction times, reaction time variability (median absolute deviation) and error rates improved with training. Error bars represent bootstrapped 95% confidence intervals.

Results

Response selection improved with practice

In Experiment 1, participants (n=22) completed a visuomotor association task in which arbitrary stimuli instructed them to press specific buttons on a keyboard (Figure 1). To assess the effects of practice, we contrasted behavior in two conditions, a *4-Day Practice* condition and a *Minimal Practice* condition. In the 4-Day Practice condition, participants first trained on a previously unseen stimulus-response mapping, completing 4,000 reaction-time trials (10 x 100 trial blocks for four consecutive days) in which they responded as quickly as possible to stimuli presented on the screen in rapid succession (Figure 1d). Performance, averaged over the first and last day of practice, improved (Figure 1e) as illustrated by significant reductions in reaction times (t-test, t_{21}=11.96, p<0.001), reaction time variability (t-test on reaction time median absolute deviation, t_{21}=9.38, p<0.001), and errors (t-test, t_{21}=2.18, p<0.05). Thus, practice led to a reduction in average reaction times, i.e. more rapid response selection.

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

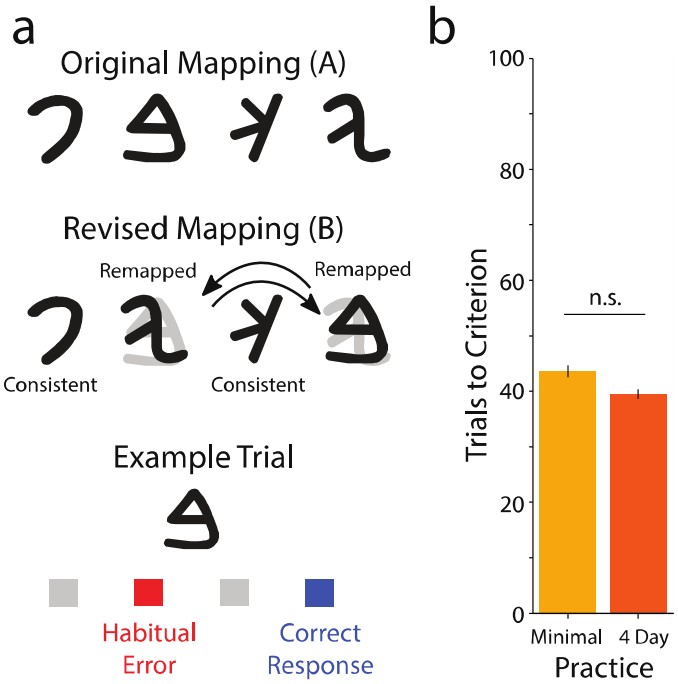


Figure 2*. Switch manipulation and retraining time. a) Following either minimal practice or four days of practice on the original stimulus-response mapping A (top: example mapping shown), two specific stimulus-response associations were switched to create a revised mapping (midde). This revised mapping allowed the identification of trials in which participants made the new correct response, or where they produced the previously practiced response (i.e. consistent with a habitual error) (bottom). d) Participants trained on the revised mapping without reaction time constraints until they reached a stable steady criterion (5 consecutive correct responses to each stimulus). Participants required approximately 40 trials to learn this revised mapping regardless of the volume of training they had completed on the original mapping. Error bars represent ±1SEM.*

We next assessed whether response selection had also become habitual. We did so by transposing the required responses for two stimuli (Figure 2b). If response selection had become habitual, we expected participants to persist in responding according to the previously-practiced mapping, rather than the revised mapping. Participants learned the revised mapping

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in a criterion assessment block; they were instructed that time constraints had been removed, and that they should focus on learning a new stimulus-response mapping. Participants trained on the revised mapping until they reached an accuracy criterion of five consecutive correct responses to each stimulus, which occurred on average within 44 (±5, SEM) trials. The number of trials required was comparable to that in the Minimal Practice control condition (40±4 SEM), in which participants barely practiced the original map (practicing it just enough to satisfy an accuracy criterion of five consecutive correct responses to each stimulus) (Figure 2d; paired samples t-test, t_{21}=0.63, p=0.53). Thus participants had no difficulty in learning to accurately respond according to the revised mapping, regardless of whether or not they had practiced the original mapping.

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

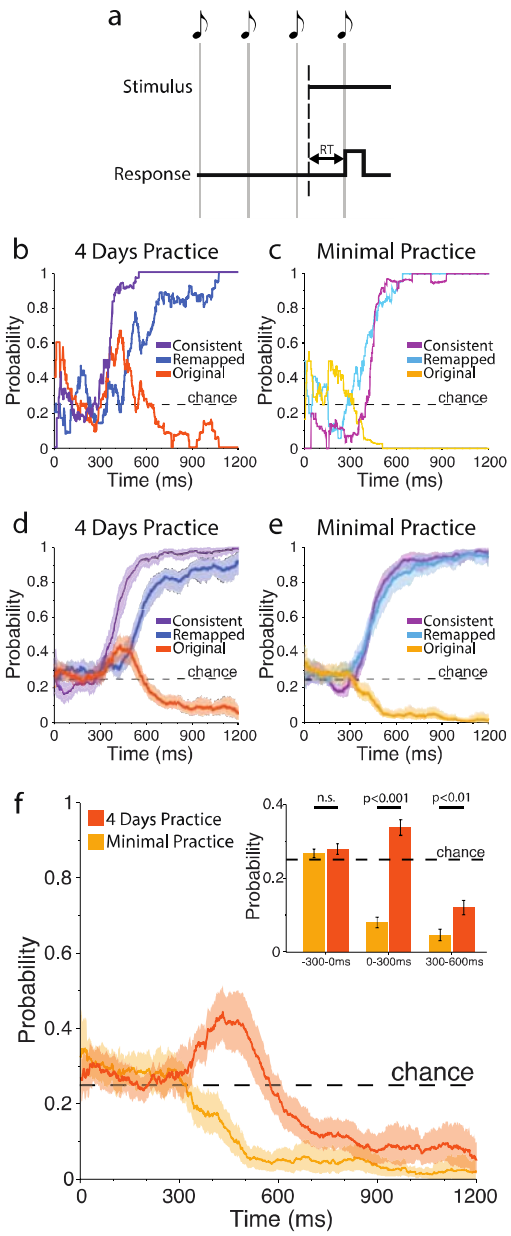


***Figure 3**. Forced response paradigm and results. a) Schematic of timed-response trial procedure. Participants were instructed that they must make a response synchronously with the final tone in a sequence of four equally spaced metronome beats. Stimulus onset was varied randomly and uniformly from 0-1200ms prior to the fourth tone, effectively controlling participant response times. b) Results for the 4 Day Practice condition for an example participant. Sliding-window probability of expressing different responses. Following initial chance performance below 300ms (when participants had to guess), the likelihood of correct responses to*

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

unchanged stimuli (purple) rose rapidly. The likelihood of correct responses to revised stimuli (blue) took longer to rise, and reached a lower plateau. The likelihood of habitual responses (red) first rose rapidly (between 300-500ms) before falling below chance. c) Results for the same participant in the Minimal Practice condition. Following initial chance performance, the rate at which participants made correct responses to both unchanged and revised stimuli then rose rapidly, while the likelihood of the original response fell monotonically. d),e), same as b,c, but behavior averaged across all participants. f) Direct comparison of the time-varying probability of expressing the original response across the groups. Inset shows comparisons of the proportion of habitual responses binned across 300ms intervals relative to the minimum time at which participants could respond to stimuli (t_{min} - see text). Shaded error regions in d-f represent bootstrapped 95% confidence intervals. Bar chart error bars represent ±1SEM.

Limiting reaction times revealed habitual selection following practice

In order to unmask potential habitual selection of the original response, we forced participants to act at different response times, ranging from 0-1200 ms, using a forced-response paradigm[17,23,24] (Figure 3a). Four tones were played, each separated by 400ms, and participants were instructed that they must make a response synchronously with the final tone. The time of stimulus presentation was varied from trial to trial relative to this fixed response time, effectively controlling the allowed preparation time in each trial.

We first assessed whether practice improved the speed of response selection for symbols that were consistently mapped. Figure 3 shows speed-accuracy trade-off (the probability of generating a correct response as a function of allowed preparation time; purple curve) for consistently mapped stimuli (purple curve) for an example participant (b) and averaged all participants (d). This speed-accuracy trade-off began at chance (0.25) for preparation times less than ~300 ms, indicating that participants did not have sufficient time to process the stimulus and select the appropriate response in this range, and instead had to

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

make a guess in order to meet the deadline of the fourth tone. The participants' accuracy rose gradually as preparation time increased, reaching asymptote between 700-900ms. This speed-accuracy trade-off was shifted earlier relative to the analogous curve for the Minimal Practice control condition (t-test on center of the speed-accuracy trade-off for the 4-Day Practice vs Minimal Practice conditions, t_{21}=3.84, p<0.001, mean difference 43ms), demonstrating that four days of practice led to more rapid response selection, consistent with observed reductions in reaction time during training (Figure 1e).

We next examined behavior in response to remapped stimuli. In the Minimal Practice condition, the speed-accuracy trade-offs for consistent and remapped stimuli appeared to be closely aligned (Figure 3c,e). Indeed, there was no significant difference between the center of the speed-accuracy trade-offs, assessed by fitting a cumulative Gaussian to each (See Supplementary Figure 1 for method details) (t_{21}=0.64, p=0.53), suggesting that participants were able to accommodate the remapping without any detriment to their performance. By contrast, practicing the original mapping for four days significantly slowed the ability to respond to remapped stimuli, both compared to consistently mapped stimuli (t_{21}=5.93, p<0.001, mean difference 93ms), and compared to remapped stimuli in the Minimal Practice condition (t_{21}=2.42, p<0.05, mean difference 80ms). Thus, practice with the original mapping enhanced the speed of response selection, but compromised the ability to flexibly adjust to a revised mapping, suggesting that performance may have become habitual.

Next, we directly examined whether practice led participants to habitually select practiced responses (Fig 3b,d, red curves). In the 4-Day Practice condition, the likelihood of responding according to the previously practiced mapping began at chance at low preparation times, but then briefly increased at preparation times of 300-600 ms, indicating that they habitually prepared this response. By contrast, in the Minimal Practice condition (Figure 3c,e), habitual errors began at chance, then declined monotonically as preparation time increased. We summarized these observations by analyzing the overall likelihood of habitual responses in a

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

300ms interval aligned to the minimum possible time at which participants could generate an accurate response (t_{min}, identified as the time at which the speed-accuracy trade-off for unchanged stimuli first reached 5% of its height). A significant interaction between condition and response latency ($_{RM}$ANOVA, $F_{1,21}$=58.32, p<0.001) confirmed that practicing the original mapping led to transient habitual selection of previously practiced responses.

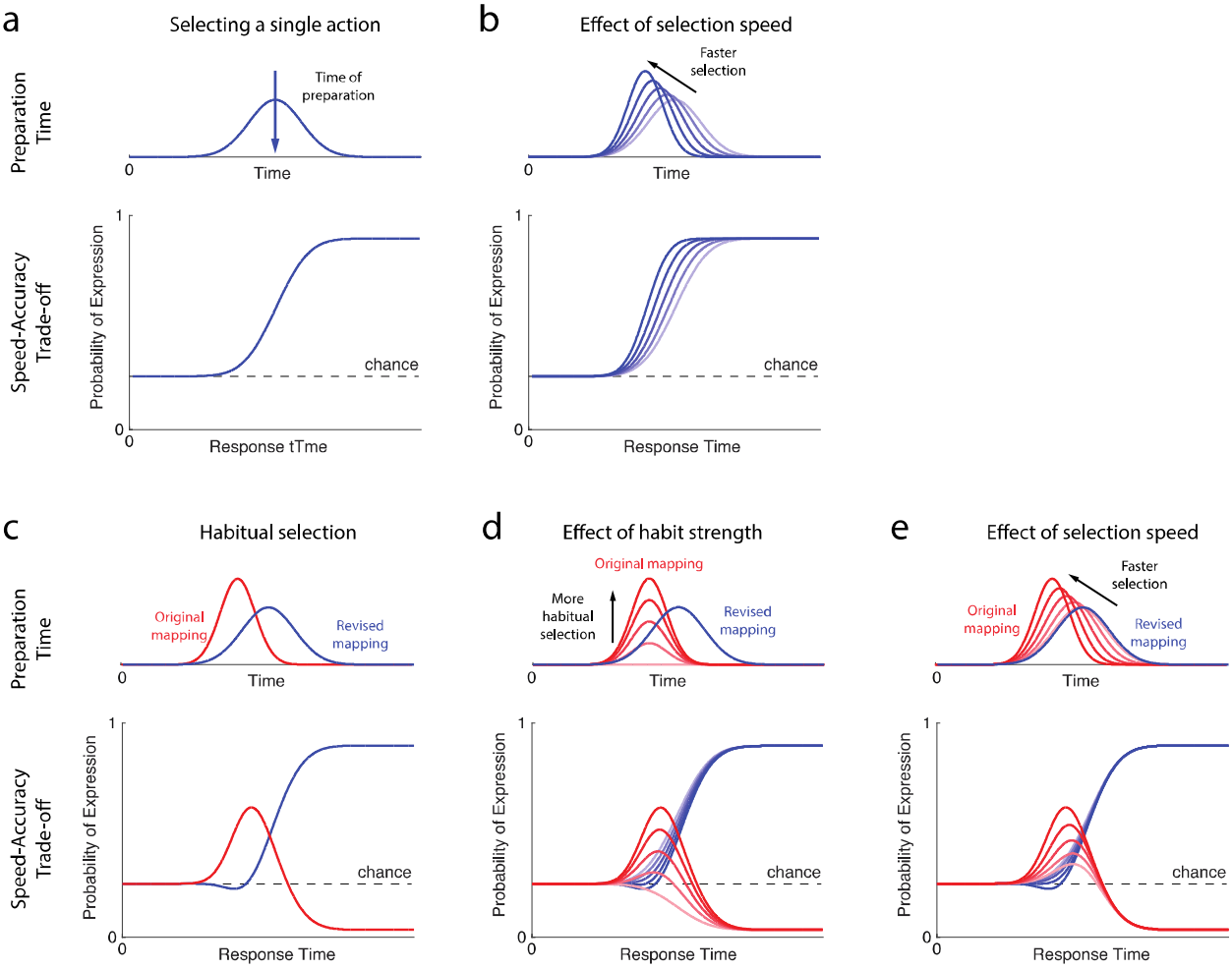


***Figure 4**. Computational model of response selection. a) We assume that, in each trial, a response is selected at a random time after stimulus onset (top panel), giving rise to the observed speed-accuracy trade-off across trials (bottom panel). b) As the mean and variance of the time of selection improve (top) the speed-accuracy trade-off becomes steeper (bottom). c) After the mapping is revised, the original mapping is habitually selected according to the same time distribution as before. The appropriate, revised response is selected at later time, at which*

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

point it replaces the habitually selected response. This leads to a particular time-varying probability of each response being expressed. d) One potential effect of practice might be to modulate the probability that a response would be habitually selected (varying shades of red in the top panel). This would be manifested behaviorally as a variable size bump in the probability of expressing the original mapping (varying shades of red/blue in the lower panel). The extreme cases are indicated by dark red (fully habitual selection) and light red (no habitual selection) e) An alternative effect of practice is that it would improve the speed at which the original response would be selected (varying shades of red in top panel). This variation in selection speed would lead to a similar modulation of the likelihood of expressing the original mapping as in d). Thus, variations in the likelihood of expressing a habitual response do not necessarily reflect variations in habit strength but might instead be attributable to more rapid response selection.

A computational model distinguished between goal-directed and habitual responding

The distribution of response times imposed on participants revealed a stereotyped time-course of habitual responding. We developed a computational model to account for this behavior and better understand how it varied across participants (Figure 4. Our model extends that proposed in our previous work[24], assuming that participants select a responses at some time T_A, which varies randomly from trial to trial (here, according to a normal distribution; Figure 4a). The speed-accuracy trade-off reflects the probability that the correct action had been selected by the time of responding. Improvements in selection speed are accounted for in the model through a shift and narrowing in the distribution of T_A (Figure 4b).

To account for potentially habitual selection following revision of the map, we assumed that participants might habitually prepare the initially practiced response at a random time T_A, but then also prepare the correct, remapped response at some other time T_B. Participants would express the habitually selected response if the originally practiced response (A) was prepared, but this had not yet been replaced by the appropriate, remapped response (B) (Figure 4c).

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

This simple model yielded a remarkably good fit to participants' behavior in the 4 Day Practice condition (Figure 5), accounting for the data significantly better than an alternative model in which the previously practiced response was never prepared (mean difference in AIC = 6.52; Figure 5e). Examining behavior on an individual basis, we found strong evidence in favor of habitual selection in 14/21 participants. By contrast, no participants in the Minimal Practice condition showed evidence of habitual selection (mean difference in AIC = -2.81).

The overall likelihood of erroneously selecting the previously practice response, i.e. the height of the red curve in Figure 3b,d) varied considerably across individuals. Two distinct factors could have affected the shape of this curve. First, there may be varying degrees of habitual selection; the previously practiced response might not have been habitually selected on every trial, but might instead have been selected with some probability ρ. Figure 4d shows that varying ρ would affect on the probability of expressing the previously practiced response as a function of response time. Second, there may be variations in the speed at which the previously practiced response could be selected, even given a fixed probability that it would be selected habitually. As shown in Figure 4e, increasing the speed at which the practiced response can be selected leads to a very similar increase in the likelihood of it being expressed at short response times. We know from the RT data that response selection becomes faster with practice. However, can improvements in response selection alone account for variations in the varying likelihood of selecting the original response, or is it necessary to also invoke the possibility of a varying habit strength (i.e. $0 < \rho < 1$)?

We used a likelihood ratio test to assess the hypothesis that participants may have habitually selected responses with an intermediate probability ($0 \leq \rho \leq 1$), with the null hypothesis being that habitual selection was all-or-nothing for a given individual ($\rho = 0$ or $\rho = 1$). We found no evidence that habits could have an intermediate probability of being expressed

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Likelihood ratio test; p=0.914). Thus our data do not support the idea of degrees of habitual selection. Rather, habitual selection appears to be all or nothing.

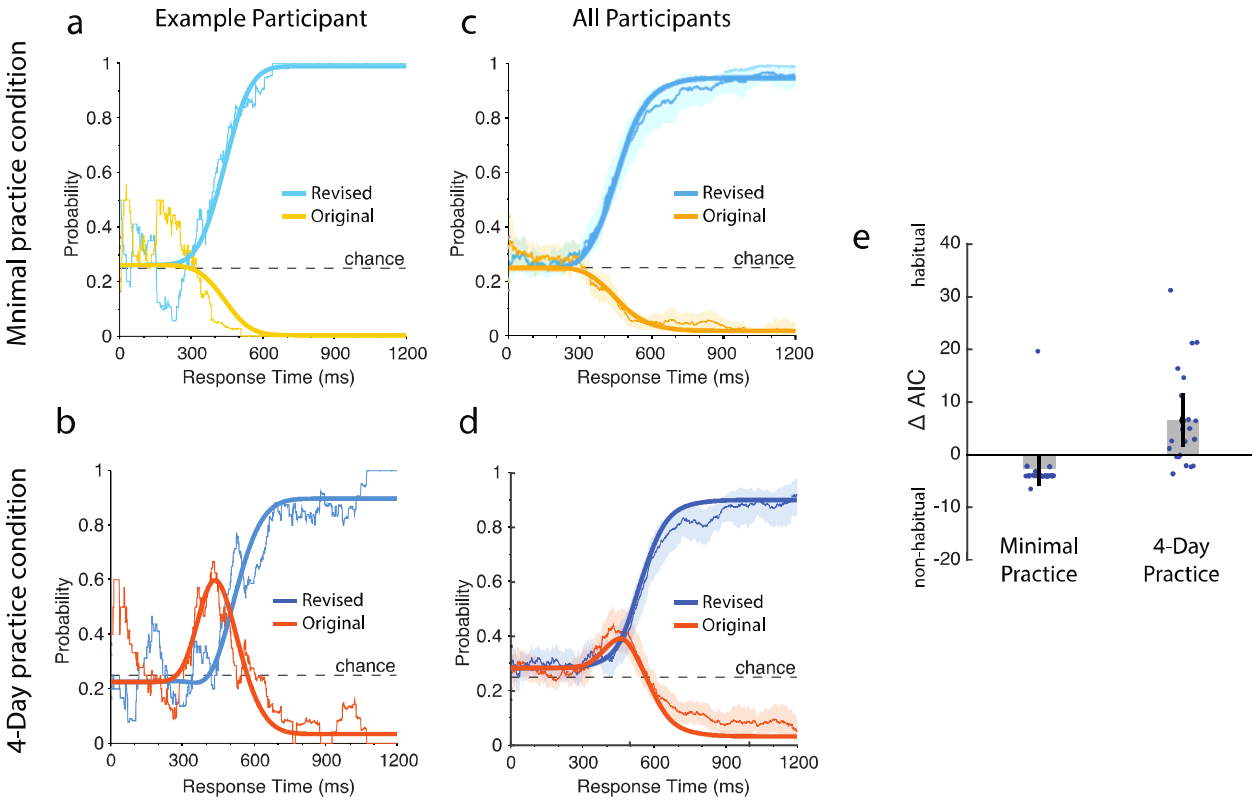


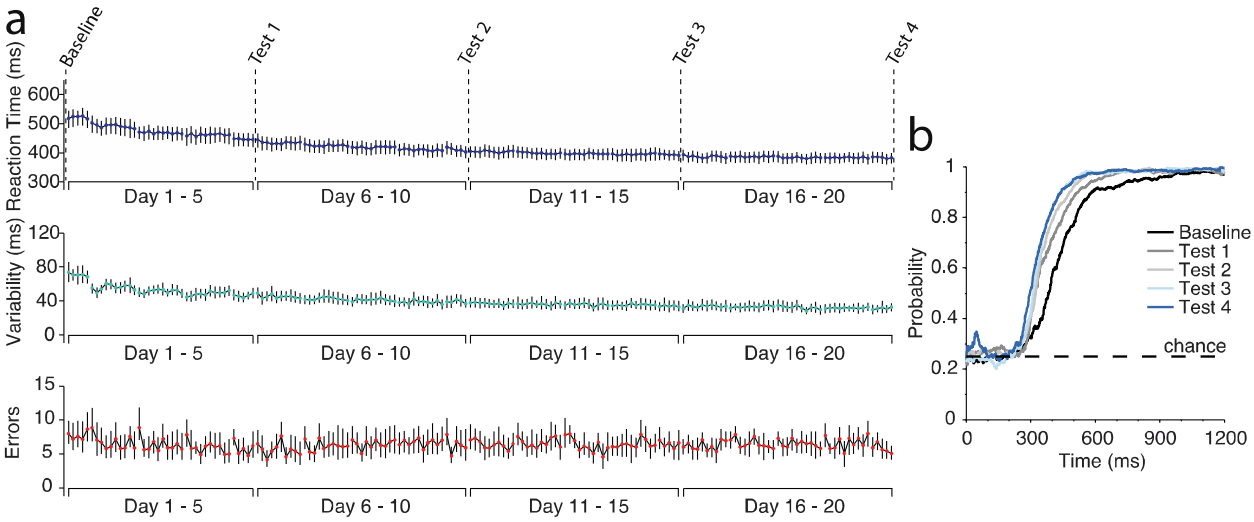
Figure 5. *Model fits to data from Experiment 1. a) Behavior (thin lines) and model fit (bold lines) for an example participant in the minimal practice condition. Fit shown here for the model with no habitual selection (α = 0), which had a lower AIC for this condition. b) Behavior and model fit the same participant in the 4-Day practice condition. c)-d) As a)-b) but averaged across participants. e) Difference in AIC between the habitual (ρ = 1) and non-habitual (ρ = 0) models in the two conditions in Experiment 1. Only one participant showed evidence of habitual selection after minimal practice, while most participants exhibited habitual selection after 4 days of practice.*

Response latency further improved following extensive practice

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The results of Experiment 1, along with the computational model, established that four days of practice led most participants to habitually select the practiced responses. Four days of practice also led to more rapid response selection. However, these results alone do not clarify whether or not improvements in the speed of response selection arise from the same process that renders response selection habitual. In Experiment 2 we extended the duration of training to 20 days, to test whether more extensive practice might enable further improvements in the speed of response selection. If the same process is responsible for responses becoming faster and habitual, further training should yield no improvements in selection speed (aside from, perhaps, rendering all participants habitual). Alternatively, if the speed of selection can be improved *independently* of it being habitual, we may see further increases in selection speed.

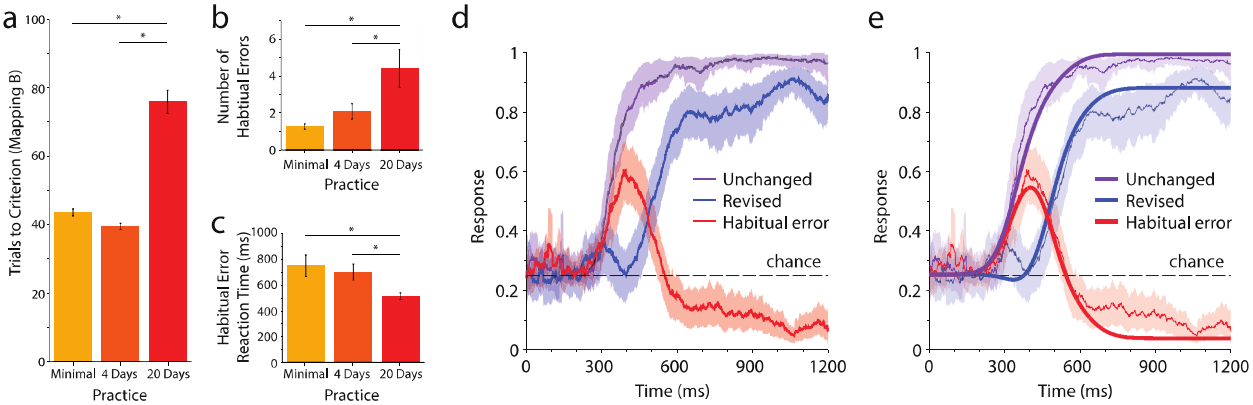
A new group of participants (n=14) trained on an original mapping over a period of 4 weeks, completing 20 days of practice in total (1000 trials per day, 20,000 total trials). Participants trained in reaction-time-based trials, but the speed of their response selection was also periodically assessed using forced-response trials to obtain a speed-accuracy trade-off for the trained mapping. We first did this immediately after they achieved the minimal accuracy criterion of 5 correct consecutive responses to each stimulus, and then again at the end of each week of training.



bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

***Figure 6**. Extensive Training data. a) Group reaction times, reaction time variability (median absolute deviation) and error rates for training blocks. Each circle presents data for one block of 100 trials. Separations in the lines between points indicate separate days. Error bars represent bootstrapped 95% confidence intervals. b) The speed-accuracy trade-off for the original stimulus-response associations (Mapping A) was assessed using forced response trials at baseline (just after achieving criterion), and tested at the end of each week of practice, identifying significant improvements over the course of training.*

Participants in the 20-Day practice continued to reduce their reaction times (t-test on first vs final day, t_{13}=13.27, p<0.001) and reaction time variability (t-test on median absolute deviation for the first vs final day, t_{13}=10.75, p<0.001) over the course of training beyond the first week (Figure 6a). The speed-accuracy trade-off, as measured using forced response trials at baseline and at the end of each week of practice, revealed similar improvement (rmANOVA on mean preparation time, $F_{4,52}$=41.81, p<0.001; Figure 6b). Notably, by the end of training, participants in the 20 Day Practice (Experiment 2) condition had reaction times ~70ms faster than participants in the 4 Day Practice (Experiment 1) condition (group-by-day (first/last) interaction, $F_{1,34}$=22.53, p<0.001); no baseline difference between groups on day 1 of training, t_{34}=0.89, p=0.38), indicating that more practice led to improved speed of selection.



bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 7*.Trials to achieve criterion across all experiments and conditions, and forced response data for the extensive training group. a) Trials to achieve criterion for the revised mapping was higher in the 20-Day Practice condition than the Minimal Practice or 4-Day Practice conditions in Experiment 1. b) Participants in the 20 Day Practice condition made significantly more habitual errors during the re-training block, and did so with significantly faster reaction times (c) compared to conditions with less practice. d) Forced response data for the 20-Day Practice condition. Speed-accuracy trade-off for consistently-mapped stimuli (purple), and remapped stimuli (blue). The probability of expressing the originally practiced response (red) showed was even greater than in the 4-Day Practice condition. e) Model fits (bold lines) to data from the 20-day condition (thin lines), averaged across participants.*

Following the 20 Day Practice (20,000 trial) condition, we tested whether responses were habitual by imposing the same switch manipulation as in Experiment 1 (see Figure 2b. Participants first practiced the new mapping until they could make 5 correct consecutive responses to each stimulus in a criterion test block (Figure 7a). Participants that practiced the original map for 20 days (Experiment 2) required more trials to achieve this criterion than were needed in either condition in Experiment 1 (t-test, 20 Day Practice vs Minimal Practice condition, t_{34}=2.74, p<0.05, and 20 Day Practice vs 4 Day Practice condition, t_{34}=3.24, p<0.01). We examined whether this reflected difficulty in acquiring the revised mapping, or could be attributed to participants habitually persisting with short reaction times that had been successful during extensive practice[25]. When attempting to learn the revised mapping, participants that completed the 20 Day Practice condition made more habitual errors (Figure 7b, Mann-Whitney tests on number of habitual errors, 20 Day Practice vs Minimal Practice, Z=2.92, p<0.01, and 20 Day Practice vs 4 Day Practice, Z=2.14, p<0.05), and did so with shorter reaction times (Figure 7c, Mann-Whitney tests on reaction times of habitual errors, 20 Day Practice vs Minimal Practice, Z=2.94, p<0.01, and 20 Day Practice vs 4 Day Practice , Z=2.46, p<0.05). The greater

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

number of trials required to reach the accuracy criterion is therefore consistent with a tendency to commit low-latency habitual errors, rather than difficulty in explicitly learning the revised mapping.

After attaining the accuracy criterion for the revised map, participants were required to respond to this revised mapping under forced-response conditions (Figure 7d). Consistent with the improvements in reaction time (Figure 4a) and speed-accuracy trade-off (Figure 4b) during practice, 20 days of practice enabled participants to improve the speed of their response selection; the center of the speed-accuracy trade-off for consistently mapped stimuli was significantly earlier than that of participants in the 4-day practice condition in Experiment 1 (t-test on mean preparation time, t_{34}=2.32, p<0.05, mean difference 48ms). As in Experiment 1, the speed at which responses to remapped stimuli could be selected was slower than for consistently mapped stimuli (t-test on mean preparation time, t_{34}=11.50, p<0.001, mean difference 162ms). Notably, the speed-accuracy trade-off for remapped stimuli was similar whether the remapping was preceded by 20 days of practice or by 4 days of practice (t-test on revised response speed-accuracy trade-offs, t_{34}=0.81, p=0.43).

As expected, the 20 Day Practice condition also led to habitual response selection. The likelihood of expressing the previously practiced response was at chance for times before participants could process the stimulus (300-0ms before t_{min}, t-test against chance, t_{13}=1.0, p=0.36), then rose above chance (0-300ms after t_{min}, t_{13}=6.0, p<0.001), before falling below chance for responses at longer response times (300-600ms after t_{min}, t_{13}=3.6, p<0.01).
When forced to respond at low latencies, participants that trained for 20 days were more likely to produce habitual responses than participants that trained for 4 days (t-test on 2-Day practice vs 4-Day practice groups for t_{min} to t_{min}+300, t_{34}=2.98, p<0.01). Our computational model again accounted for the observed behavior extremely well (Figure 7e), and demonstrated that this increased likelihood of habitual responses was attributable to the fact that practice allowed responses that were already selected habitually after 4 days of practice, to be selected more

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

rapidly. All participants in Experiment 2 exhibited habitual selection (mean difference in AIC = 19.72). Furthermore, as in the 4-Day practice group, extending the model to allow for partial habits ($0 \leq \alpha \leq 1$) did not provide a better description of the data (Likelihood ratio test; p=1.00).

Discussion

Our data and model show that practice led to both more rapid response selection, and habitual response selection. However, these developments followed a different time course. Response selection became habitual in most participants within four days of practice. By contrast, response speed improved over up to twenty days of practice. Furthermore, while the speed of response selection varied continuously with practice, being subject to habitual action selection appeared to be all or nothing. Variations in the likelihood of expressing the original response as a function of preparation time could be fully accounted for by continuous variations in the speed of response selection, without having to assume any continuum of habit strength. In other words, being habitual was a discrete state, whereas skill level could vary continuously.

Limiting reaction times unmasks habitual behavior

Our paradigm and results clearly illustrate the time-varying competition between goal-directed and habitual control processes. Varying the allowed preparation time modulated which response was expressed. This implies that both mappings were represented during each trial, demonstrating the existence of multiple components of learning. The relative expression of different components of learning has previously been shown to be influenced by limiting cognitive resources[26,27], including available preparation time[17,20,22,28,29]. However, previous research has manipulated preparation time in a relatively simple 'high-or-low' manner[17,20], or based on spontaneous variations in ‘voluntarily’ selected reaction times[29]. The forced-response paradigm used here allowed us to measure the temporal dynamics of these effects at far

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

greater resolution; by assessing responses across a continuum, we were able to track the dynamically evolving competition between habitual and goal-directed selection processes.

The behavior we observed was consistent with a model which assumed that responses were selected at a random time following movement onset. Practice reduced the mean time at which a response could be selected. Practice also led response to be habitually selected. Importantly, we suggest that *selection* of a response does not necessarily imply immediate *expression* of that response; rather, a response must be prepared and initiated separately. We have previously argued that the reaction time at which a movement is initiated is independent of preparation or selection of the required movement[24]. Participants have longer reaction times than appear necessary based on the speed-accuracy trade-off, yet also commit 'fast errors' in which they seemingly initiate movement before selecting the correct response. A similar separation between selection and initiation is particularly apparent when participants attempted to learn the revised stimulus-response relationship during the criterion training block in Experiment 2. Having practiced for 20 days, participants tended to respond rapidly, perhaps through a habitual tendency to respond at short reaction times[25]. Notably, these participants were more likely to express the previously practiced response, due it having been habitually selected.

Skills, Habits, and Automaticity

Both skills and habits are related to the notion of automaticity. Definitions of automaticity vary, but it is typically thought to involve improvements in skill, the obligatory enactment of a skill, and the ability to perform a skill with little or no conscious deliberation. Our results support links between habitual selection and automatic behavior; participants habitually chose the previously selected response, despite consciously attempting to select the revised response. There is a long-standing debate regarding whether automaticity is a continuous[30] or discrete[31] process. Our finding that habitual selection is all-or-nothing supports the idea that automaticity

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

might be discrete. However, we suggest that both the speed of response selection, and whether or not selection is habitual, both contribute to common measures of automaticity.

Although the notion of pre-computaion, or caching, or stimulus-response associations seemed to provide a plausible link between skills and habits, our data did not support this idea. Participants did not achieve more rapid response selection by becoming more habitual. The exact relationship between skills, habits and automaticity remains uncertain. However, other recent findings support the fact that skill can vary independently of habit and automaticity. Deliberate (model-based) control can end up leading to habitual[32], while goal-directed behavior can become expressible rapidly and automatically through practice[33]. The computational basis of faster response selection, habitual response selection, and automaticity remain to be precisely determined.

Recognizing that skill acquisition and habit formation are be distinct processes has significant implications for studying the neural substrates of skills, habits and automaticity. Previous research has failed to achieve any clear consensus on the neural basis of automaticity, proposing that automaticity arises either through increases in network efficiency[3,34], or through discrete shifts in the brain regions that control behavior[5,35], either within the basal ganglia[5], within cortex[36], or from the cortex to the cerebellum[3]. We propose these differing conclusions arise because the tasks they employ all involve practice, but their behavioral assays focus on only a single measure of performance. Separately measuring skill level and the extent to which behavior is habitual could therefore considerably enhance our understanding of the neural basis of performance improvement through practice.

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

Participants

A total 39 participants took part in the study. Experiment 1 included 24 individuals. Two participants were excluded from Experiment 1 as they completed only one of the two required experimental conditions, leaving 22 full datasets for the experiment (17 right handed, 13 female, mean age 21 years). Experiment 2 included 15 participants. One participant was excluded (computer hardware failure), leaving a total of 14 participants (12 right handed, 4 female, mean age 26 years) to complete the experiment. All participants gave written informed consent, and all procedures were approved by the Johns Hopkins School of Medicine Institutional Review Board. Participants received financial compensation ($15/hour) for their participation.

General Procedures

The task involved responding to the appearance of one of four stimuli (letters of the Phoenician alphabet) by pushing a specific key on a computer keyboard with the index, middle, ring, or little finger of the dominant hand. The stimulus corresponding to each response was counterbalanced across participants, controlling for potential effects whereby participants would find some stimuli easier to recognize and learn to respond to than others. As Experiment 1 comprised two conditions and used a within-subjects design, we employed two sets of distinct stimuli (see Supplementary Figure 2), and counterbalanced the condition to which they corresponded across participants. Participants in Experiment 1 also completed the two conditions in a counterbalanced order.

Participants attempted to respond to stimuli in *training, criterion test,* or *forced response* trial blocks:

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Training blocks

During training participants completed a gamified task in which they attempted to complete blocks of 100 reaction time based trials as quickly as possible (See Figure 1c). In each trial a stimulus appeared in the center of the screen, and a tone played to signal the participant that a trial had started. On correct responses a pleasant auditory tone sounded, and after a 300ms delay the task advanced to next trial. Errors were punished with an auditory buzzer sound and an enforced delay of 1000ms, after which the participant could once again respond to the same stimulus; this process repeated until the correct response was provided, at which the task progressed to the next trial. At the end of each block participants received feedback on the time taken to complete each block, and how this compared to their 'personal best' block completion time. Participants were encouraged to improve their performance by aiming to beat their personal best time each time they completed the task.

Criterion test blocks

We assessed the ability of participants to learn new, established, or revised stimulus-response associations using criterion test blocks. Participants were instructed that reaction time constraints were removed, that their goal was to learn the correct set of stimulus-response associations, and that the block would end once they had made enough correct responses in a row. These blocks ended once participants had made five consecutive correct responses to each stimulus (minimum of 20 trials), and the number of trials required to reach this steady, high-accuracy criterion was recorded.

Forced response blocks

We used forced-response trials to probe the speed of response selection and to assess whether participants habitually selected their responses. Each block comprising 100 trials. In each trial the participant heard a series of four tones, spaced 400ms apart, and was instructed

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to synchronize their response with the onset of the fourth and final tone. The stimulus appeared at a random time during the series of tones, effectively controlling the time in which participants could prepare their response. As such, in cases in which participants did not have chance to process the stimulus (e.g. when it appeared less than ~300ms before the deadline of the fourth tone), they were essentially forced to guess the correct response (and thus had a 1 in 4 chance of selecting the correct answer).

Protocol

Experiment 1

In Experiment 1 participants completed a counterbalanced, crossover design comprising two conditions. Both conditions began with a warm up/familiarization task. Participants completed 2 blocks (200 trials total) of reaction based trials in response to non-arbitrary stimuli (pictures of the hand with one finger colored black to indicate the desired response - see Supplementary Figure 1). This was followed by 2 blocks (200 trials total) of forced response trials to the same non-arbitrary stimuli. This familiarization period allowed the experimenter to explain the practice and forced response paradigms to the participant, and to ensure that the participant was capable of complying with the demands of each task.

Following this familiarization procedure, participants in the Minimal Practice condition then learned an original map of stimuli (Mapping A) in a block of criterion test trials, after which a second block of criterion test trials was used to introduce and assess the ability to learn a revised mapping (mapping B). We then probed for habitual response selection using forced response trials. The Practice condition used the same assessment, but this was completed after four consecutive days of practice (10x100 trial reaction time training blocks each day) on Mapping A.

Experiment 2

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The second experiment comprised a single condition. All participants first completed the same warm up/familiarization procedure as in Experiment 1. Participants then completed a criterion test block in which they learned a set of stimulus-response associations through trial and error (Mapping A). Once they had achieved criterion, they completed 500 forced response trials on this original mapping (to allow assessment of baseline performance), followed by 500 reaction-time-based training trials. Each day thereafter participants completed 'training sessions' in which they completed 10x100 trial blocks of reaction-time-based training trials. On the final (fifth) day of training for each week of practice, participants completed a 'training and probe' session, in which they completed 500 (5x100) reaction time based training trials, followed by 500 (5x100 trial blocks) of forced-response trials. Participants completed 20 sessions in this manner (aiming to complete five sessions of practice in each seven day week), allowing us to measure changes in performance at baseline, and after one, two, three, and four weeks of practice.

On a separate day after all training sessions were complete, participants were exposed to the same assessment as in Experiment 1; they learned a revised set of stimulus-response associations in a criterion test block, and their performance on this new mapping was then probed in 5x100 trial blocks of forced-response trials.

Data Analysis

Reaction time trials

Performance for each block was measured by taking the median reaction time (measured from stimulus onset to response onset) for correct trials, the median absolute deviation of the reaction time (this is equivalent to variance but using median instead of mean averaging, and is thus more appropriate for reaction time data), and by calculating the error rate for each block (i.e. number of erroneous responses provided in each block; note that it was

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

possible for participants to make multiple errors in the same trial, as the trial did not advance until the participant provided the correct answer).

Criterion test trials

Criterion test trials were primarily analyzed by counting the number of trials required for a participant to make five consecutive correct responses to each stimulus. The reaction time for each response was recorded (although participants were made aware that there were no reaction time requirements for these trials).

Forced response trials

Preparation times were calculated as the time between the presentation of the stimulus and the first response that the participant made to it. Data were used to examine the likelihood of three types of response; correct responses to *consistently mapped* stimuli, i.e. stimuli for which the same key press was required throughout the experiment, correct responses to the *revised* associations, and responses consistent with the *original* mapping. We employed a sliding window approach to visualize the time-varying liklihood for each of these trial types and response types; responses were binned over 100ms windows, and the proportion of correct vs incorrect responses was calculated and recorded for the center of each window.

Response Selection Model

We developed a simple model to quantify participants performance and assess the relative effects of practice on the speed of response selection and whether or not response selection became habitual. We assumed that, for a given mapping A, the correct response would be selected at a random time $T_A \sim N\left(\mu_A, \sigma_A^2\right)$ following presentation of the stimulus. Responses generated prior to T_A would be random, while responses generated later than T_A

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

would be generated correctly with probability q_A. The probability of observing a correct response, $r = r_A$, given that the response was generated at time t is then given by

$$\begin{aligned} p(r_A \mid t) &= p(r_A \mid t < t_A)\, p(t < t_A) + p(r_A \mid t > t_A) \\ &= q_I \left(1 - \Phi_A(t)\right) + q_A \Phi_A(t) \end{aligned},$$

where $\Phi_A(t) = \Phi\left(t \mid \mu_A, \sigma_A^2\right)$ is the cumulative distribution of T_A. Likewise, the probability of generating any other response is given by

$$p(r = r_{\bar{A}} \mid t) = q_I \left(1 - \Phi_A(t)\right) + \frac{1 - q_A}{3} \Phi_A(t),$$

assuming that all errors after T_A would be uniformly distributed across other responses. The speed of response selection, which gives rise to the observed speed-accuracy trade-off, is therefore represented by the parameters μ_A and σ_A.

To model the impact of habitual selection when exposed to remapped stimuli, we modeled each learned response, A and B, through analogous processes, i.e. we assumed that response A could be selected at some random time T_A, and that response B became available at some stochastic time T_B after stimulus presentation. The probability of a given response being generated depended on which events (selection of A; selection of B) had occurred by the time of response initiation:

$$\begin{aligned} p(r \mid t) &= p(r \mid t < t_A, t < t_B)\, p(t < t_A, t < t_B) \\ &+ p(r \mid t > t_A, t < t_B)\, p(t > t_A, t < t_B) \\ &+ p(r \mid t < t_A, t > t_B)\, p(t < t_A, t > t_B) \\ &+ p(r \mid t > t_A, t > t_B)\, p(t > t_A, t > t_B) \\ &= \alpha_{r|\bar{A}\bar{B}} \left(1 - \Phi_A(t)\right)\left(1 - \Phi_B(t)\right) \\ &\quad + \alpha_{r|A\bar{B}} \Phi_A(t) \left(1 - \Phi_B(t)\right) \\ &\quad + \alpha_{r|\bar{A}B} \left(1 - \Phi_A(t)\right) \Phi_B(t) \\ &\quad + \alpha_{r|AB} \Phi_A(t) \Phi_B(t) \\ &= \boldsymbol{\alpha}_r^{\,T} \boldsymbol{\Psi}(t) \end{aligned},$$

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where $\mathbf{\Psi}(t) = [(1-\Phi_A(t))(1-\Phi_B(t)); \Phi_A(t)(1-\Phi_B(t)); (1-\Phi_A(t))\Phi_B(t); \Phi_A(t)\Phi_B(t)]$.

Since participants were instructed to act according to mapping B, we assumed that if the response associated with mapping B was available, then participants would generate it (with probability q_B). If, however, response A was available but not response B, then response A would be generated. If neither response was available, participants would generate a random response. We captured the fact that random responses (before selection of A or B) might not have been selected uniformly through a parameter q_I. Note that since responses were pooled across both of the two remapped stimuli, and across the two non-remapped stimuli, we only needed to include a single parameter that specified the relative baseline likelihood of selecting remapped versus non-remapped responses.

The conditional probabilities were therefore given by:

$$\boldsymbol{\alpha}_{r_A} = \begin{pmatrix} q_I \\ q_A \\ 1-q_B/3 \\ 1-q_B/3 \end{pmatrix}, \boldsymbol{\alpha}_{r_B} = \begin{pmatrix} q_I \\ 1-q_A/3 \\ q_B \\ q_B \end{pmatrix}, \boldsymbol{\alpha}_{r_{other}} = \begin{pmatrix} q_I \\ 1-q_A/3 \\ 1-q_B/3 \\ 1-q_B/3 \end{pmatrix},$$

which we express more compactly as a matrix

$$A = \begin{pmatrix} q_I & q_I & q_I \\ q_A & (1-q_A)/3 & (1-q_A)/3 \\ (1-q_B)/3 & q_B & (1-q_B)/3 \\ (1-q_B)/3 & q_B & (1-q_B)/3 \end{pmatrix}.$$

Note that the bottom two rows of the matrix are the same, reflecting the fact that the response probabilities after B is prepared are independent of whether or not A has been prepared.

Under this notation, the likelihood of a single trial with response r_i can be compactly expressed as

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$p(r = r_i \mid t) = \left(A\Psi\right)_{r_i},$$

and the overall log-likelihood is given by

$$L(r \mid t) = \sum_i \log\left(\left(A\Psi\right)_{r_i}\right).$$

We identified the parameters ($\mu_A, \sigma_A, q_A, \mu_B, \sigma_B, q_B, q_I$) that maximized this likelihood, for each participant in each condition, using the Matlab function fmincon. To achieve more robust fits to data, we regularized the fits by penalizing values of the slope parameters σ_A and σ_B that deviated from a nominal value of σ_0. Thus overall we found parameters that minimized

$$L = -\sum_i \log\left(\left(A\Psi\right)_{r_i}\right) + \lambda\left(\sigma_A - \sigma_0\right)^2 + \lambda\left(\sigma_B - \sigma_0\right)^2.$$

We set $\sigma_0 = 0.07$ and $\lambda = 1000$, though are results were not strongly affected by the specific values chosen here.

We contrasted this model of habitual selection with a model in which only the response associated with the revised mapping, B, was ever selected, i.e. there was no habitual seletion of response A. This model was equivalent to the single-response model described earlier, and was implemented by setting A equal to

$$A = \begin{pmatrix} q_I & q_I & q_I \\ q_I & q_I & q_I \\ (1-q_B)/3 & q_B & (1-q_B)/3 \\ (1-q_B)/3 & q_B & (1-q_B)/3 \end{pmatrix}.$$

We similarly fit the model by finding parameter values ($\mu_B, \sigma_B, q_B, q_I$) that minimized the penalized negative log-likelihood. We compared these two models by computing the Aikake information criterion, which takes into account the relative (unpenalized) likelihood of each model while also including a term which accounts for the number of parameters in the model.

In order to describe the possibility of habitual selection that may have been only partial, we introduced a further parameter ρ which modulated the probability that A would be

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

expressed if it had been prepared. This affected only the case that which A is selected by the time of response initiation, but B is not, e.g., $p(r = r_A \mid t > t_A, t < t_B) = \rho q_I + (1-\rho) q_A$. Overall, this continuous-habit model was captured through a matrix A^ρ given by

$$A^\rho = \begin{pmatrix} q_I & q_I & q_I \\ \rho q_I + (1-\rho) q_A & \rho q_I + (1-\rho)(1-q_A)/3 & \rho q_I + (1-\rho)(1-q_A)/3 \\ (1-q_B)/3 & q_B & (1-q_B)/3 \\ (1-q_B)/3 & q_B & (1-q_B)/3 \end{pmatrix}.$$

Note that the habit and no-habit models described above are special cases of this more general model corresponding to setting $\rho = 1$ and $\rho = 0$, respectively. We identified the parameters that maximized this likelihood for each individual participant in each condition. We used a likelihood ratio test to assess whether there was any evidence that participants behaved in a manner consistent with an intermediate value of ρ.

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements

We thank E. Lesage for helpful comments on the data, and M. Adputra for producing the stimuli. This project was supported by NSF Grant 1358756. RH is supported by Marie Skłodowska Curie Individual Fellowship NEURO-AGE (702784).

Author Contributions

RH and AMH conceived and designed the experiments. RH collected the data. RH and AMH analyzed the data. RH and AMH wrote the manuscript. RH, AF, JK, and AMH reviewed and edited the manuscript.

Materials & Correspondence.

Corresponding Author: Robert Hardwick, Email: robert.hardwick@kuleuven.be.

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Diedrichsen, J. & Kornysheva, K. Motor skill learning between selection and execution. *Trends Cogn. Sci.* **19,** 227–233 (2015).

2. Robertson, E. M. The serial reaction time task: implicit motor skill learning? *J. Neurosci.* **27,** 10073–5 (2007).

3. Balsters, J. H. & Ramnani, N. Symbolic representations of action in the human cerebellum. *Neuroimage* **43,** 388–398 (2008).

4. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8,** 1704–11 (2005).

5. Ashby, F. G. & Crossley, M. J. Automaticity and multiple memory systems. *Wiley Interdiscip. Rev. Cogn. Sci.* **3,** 363–376 (2012).

6. Fitts, P. M. & POSNER, M. . *Human performance*. (Brooks and Cole, 1967).

7. Dayan, P. Goal-directed control and its antipodes. *Neural Networks* **22,** 213–219 (2009).

8. Anderson, J. R. Acquisition of cognitive skill. *Psychol. Rev.* **89,** 369–406 (1982).

9. Graybiel, A. M. & Grafton, S. T. The striatum: where skills and habits meet. *Cold Spring Harb. Perspect. Biol.* **7,** a021691 (2015).

10. Salmon, D. P. & Butters, N. Neurobiology of skill and habit learning. *Curr. Opin. Neurobiol.* **5,** 184–90 (1995).

11. Hosp, J. A., Pekanovic, A., Rioult-Pedotti, M. S. & Luft, A. R. Dopaminergic projections from midbrain to primary motor cortex mediate motor skill learning. *J. Neurosci.* **31,** 2481–7 (2011).

12. Wickens, J. R., Horvitz, J. C., Costa, R. M. & Killcross, S. Dopaminergic mechanisms in actions and habits. *J. Neurosci.* **27,** 8181–3 (2007).

13. Hikosaka, O., Yamamoto, S., Yasuda, M. & Kim, H. F. Why skill matters. *Trends Cogn. Sci.* **17,** 434–41 (2013).

14. Hikosaka, O., Kim, H. F., Yasuda, M. & Yamamoto, S. Basal ganglia circuits for reward value-guided behavior. *Annu. Rev. Neurosci.* **37,** 289–306 (2014).

15. Kawai, R. *et al.* Motor Cortex Is Required for Learning but Not for Executing a Motor Skill. *Neuron* **86,** 800–812 (2015).

16. Economides, M., Guitart-Masip, M., Kurth-Nelson, Z. & Dolan, R. J. Anterior cingulate cortex instigates adaptive switches in choice by integrating immediate and delayed components of value in ventromedial prefrontal cortex. *J. Neurosci.* **34,** 3340–9 (2014).

17. Haith, A. M., Huberdeau, D. M. & Krakauer, J. W. The Influence of Movement Preparation Time on the Expression of Visuomotor Learning and Savings. *J. Neurosci.* **35,** (2015).

18. Ashby, F. G., Ell, S. W. & Waldron, E. M. Procedural learning in perceptual categorization. *Mem. Cognit.* **31,** 1114–25 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

19. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80,** 312–25 (2013).

20. Keramati, M., Smittenaar, P., Dolan, R. J. & Dayan, P. Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 12868–12873 (2016).

21. Leow, L.-A., Gunn, R., Marinovic, W. & Carroll, T. J. Estimating the implicit component of visuomotor rotation learning by constraining movement preparation time. *J. Neurophysiol.* **118,** 666–676 (2017).

22. Katnani, H. A. & Gandhi, N. J. Time course of motor preparation during visual search with flexible stimulus-response association. *J. Neurosci.* **33,** 10057–65 (2013).

23. Ghez, C. *et al.* Discrete and continuous planning of hand movements and isometric force trajectories. *Exp. brain Res.* **115,** 217–33 (1997).

24. Haith, A. M., Pakpoor, J. & Krakauer, J. W. Independence of Movement Preparation and Movement Initiation. *J. Neurosci.* **36,** (2016).

25. Wong, A. L., Goldsmith, J., Forrence, A. D., Haith, A. M. & Krakauer, J. W. Reaction times can reflect habits rather than computations. *Elife* **6,** e28075 (2017).

26. Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 20941–6 (2013).

27. Otto, A. R., Gershman, S. J., Markman, A. B. & Daw, N. D. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* **24,** 751–61 (2013).

28. Fernandez-Ruiz, J., Wong, W., Armstrong, I. T. & Flanagan, J. R. Relation between reaction time and reach errors during visuomotor adaptation. *Behav. Brain Res.* **219,** 8–14 (2011).

29. Otto, A. R. & Daw, N. The Opportunity Cost of Time Modulates Cognitive Effort. *doi.org* 201863 (2017). doi:10.1101/201863

30. MacLeod, C. M. & Dunbar, K. Training and Stroop-like interference: evidence for a continuum of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* **14,** 126–35 (1988).

31. Schneider, W. & Shiffrin, R. M. Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention. **84,** (1977).

32. Cushman, F. & Morris, A. Habitual control of goal selection in humans. *Proc. Natl. Acad. Sci.* **112,** 13817–13822 (2015).

33. Economides, M., Kurth-Nelson, Z., Lübbert, A., Guitart-Masip, M. & Dolan, R. J. Model-Based Reasoning in Humans Becomes Automatic with Training. *PLoS Comput. Biol.* **11,** e1004463 (2015).

34. Wu, T., Kansaku, K. & Hallett, M. How Self-Initiated Memorized Movements Become Automatic: A Functional MRI Study. *J. Neurophysiol.* **91,** 1690–1698 (2004).

35. Yin, H. H. & Knowlton, B. J. The role of the basal ganglia in habit formation. *Nat. Rev.*

bioRxiv preprint doi: https://doi.org/10.1101/201095; this version posted October 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Neurosci. **7,** 464–476 (2006).

36. Grol, M. J., de Lange, F. P., Verstraten, F. A. J., Passingham, R. E. & Toni, I. Cerebral changes during performance of overlearned arbitrary visuomotor associations. *J. Neurosci.* **26,** 117–25 (2006).