Proactive response preparation contributes to contingency learning: Novel evidence from force-sensitive keyboards

Daniel H. Weissman[1] and James R. Schmidt[2]

[1]Department of Psychology, University of Michigan at Ann Arbor

[2]Université de Bourgogne

Word Count: 11,381

Acknowledgements: The authors thank Princess Ativie, Matthew Dunaway, Sophia Hopkins, Katherine Ni, Carlin Pendell, Chloe Saba, and Nancy Villagomez for assistance with data collection. This work was funded by discretionary funds from the University of Michigan to Daniel H. Weissman.

Author Contributions: Daniel H. Weissman conceptualized the study. Both authors contributed to the design. Material preparation, data collection and analysis were performed by Daniel H. Weissman. Daniel H. Weissman wrote the first draft of the manuscript and both authors commented on all versions of the paper. All authors read and approved the final manuscript.

Running Head: CONTINGENCY LEARNING AND RESPONSE PREPARATION

Correspondence: Daniel Weissman, Department of Psychology, 530 Church Street, Ann Arbor, MI, 48109, USA. Email: danweiss@umich.edu

Abstract

Contingency learning can involve learning that the identity of one stimulus in a sequence predicts the identity of the next stimulus. It remains unclear, however, whether such learning speeds responses to the next stimulus *only* by reducing the threshold for triggering the expected response after stimulus onset or *also* by preparing the expected response before stimulus onset. To distinguish between these competing accounts, we manipulated the probabilities with which each of two prime arrows (Left and Right) were followed by each of two probe arrows (Up and Down) in a prime-probe task while using force-sensitive keyboards to monitor sub-threshold finger force. Consistent with the response preparation account, two experiments revealed greater force just before probe onset on the response key corresponding to the direction in which the probe was more (versus less) likely to point (e.g., Up vs Down). Furthermore, mirroring *sequential* contingency effects in behavior, this pre-probe force effect vanished after a single low-probability trial. These findings favor the response preparation account over the threshold only account. They also suggest the possibility that contingency learning in our tasks indexes trial-by-trial expectations regarding the utility of the prime for predicting the upcoming probe.

Keywords: statistical learning, associative learning, sequential trial effects, contingency learning, response force

Introduction

Humans routinely use contingencies to aid performance. While approaching a yellow traffic light, for example, a driver presses the brake because the light usually turns red soon afterward. More broadly, such *contingency learning* – i.e., learning to predict that one event follows another with greater-than-chance probability – contributes to segmenting speech into words (Saffran et al., 1996), learning stimulus sequences (Fiser & Aslin, 2002), and proactively moving spatial attention to the likely location of an upcoming target (Huang et al., 2022).

To investigate contingency learning, researchers sometimes manipulate the probability with which an initial stimulus – or prime – is followed by a second stimulus – or probe. In each trial of the color-word contingency task, for example, one of three, color-unrelated words (e.g., search, choose, or drive) might appear briefly in light brown before turning purple, orange, or grey (Schmidt & De Houwer, 2016b). Critically, each word (e.g., search) is more likely to turn one color (e.g., purple, 80%) than to turn either of two other colors (e.g., orange or grey, 10%). Although participants are not informed of these word-color contingencies, they identify the final color more quickly in "high-contingency" trials, wherein the word (e.g., search) appears in the more frequently associated color (e.g., purple), than in "low-contingency" trials, wherein the word appears in either of the less frequently associated colors (e.g., orange or grey). Analogous effects appear in 2- and 4-choice tasks[1] (Schmidt et al., 2007, 2010) and when the initial prime is something other than a word (e.g., a letter or a shape) (Arunkumar et al., 2022; Schmidt & De Houwer, 2019).

[1] In these studies, contingency learning effects appear when the prime and probe appear simultaneously. Such effects, however, are stronger when the prime appears before the probe (Schmidt & De Houwer, 2016b).

Mechanisms of contingency learning

Although it is clear that participants can rapidly learn contingent regularities and use such regularities to facilitate performance (i.e., in high contingency trials), the mechanism(s) responsible for such contingency learning remain unclear. In the following discussion, we consider three possible mechanisms. To facilitate our explanation of these mechanisms, we frame our discussion in terms of sequential sampling (e.g., drift-diffusion) models of decision-making (Ratcliff & McKoon, 2008). Such models posit that noisy perceptual evidence (e.g., about a perceptual feature like color) drifts gradually from a starting point to a response threshold (e.g., the color is purple), after which a response is executed (**Fig. 1**). Thus, as we describe next, various accounts of behavioral facilitation arising from contingency learning can be framed as involving changes to the drift rate, the starting point, and/or the response threshold.

In Schmidt and Besner's (2008) *threshold only* account (**Fig. 1a**), contingency learning arises exclusively from a change to the response threshold. Returning to the color-word contingency task as an example, the onset of the prime (e.g., search) triggers a reduction of the response threshold corresponding to the high-contingency color (e.g., purple) without modifying the response thresholds for the low-contingency colors (e.g., orange and grey). Less evidence, therefore, is required to trigger the high-contingency response than to trigger any of the low-contingency responses. Thus, contingency learning indexes a benefit to performance in high-contingency trials with no corresponding cost to performance in low-contingency trials. We note that adjusting the response threshold for only one of two possible responses in a drift-diffusion model is atypical. Such a modification of the standard modeling approach could be justified, however, if the cognitive system (a) identifies the prime before the probe and (b) uses the prime's identity to predict the response to the probe, as is commonly assumed (Schmidt, 2018).

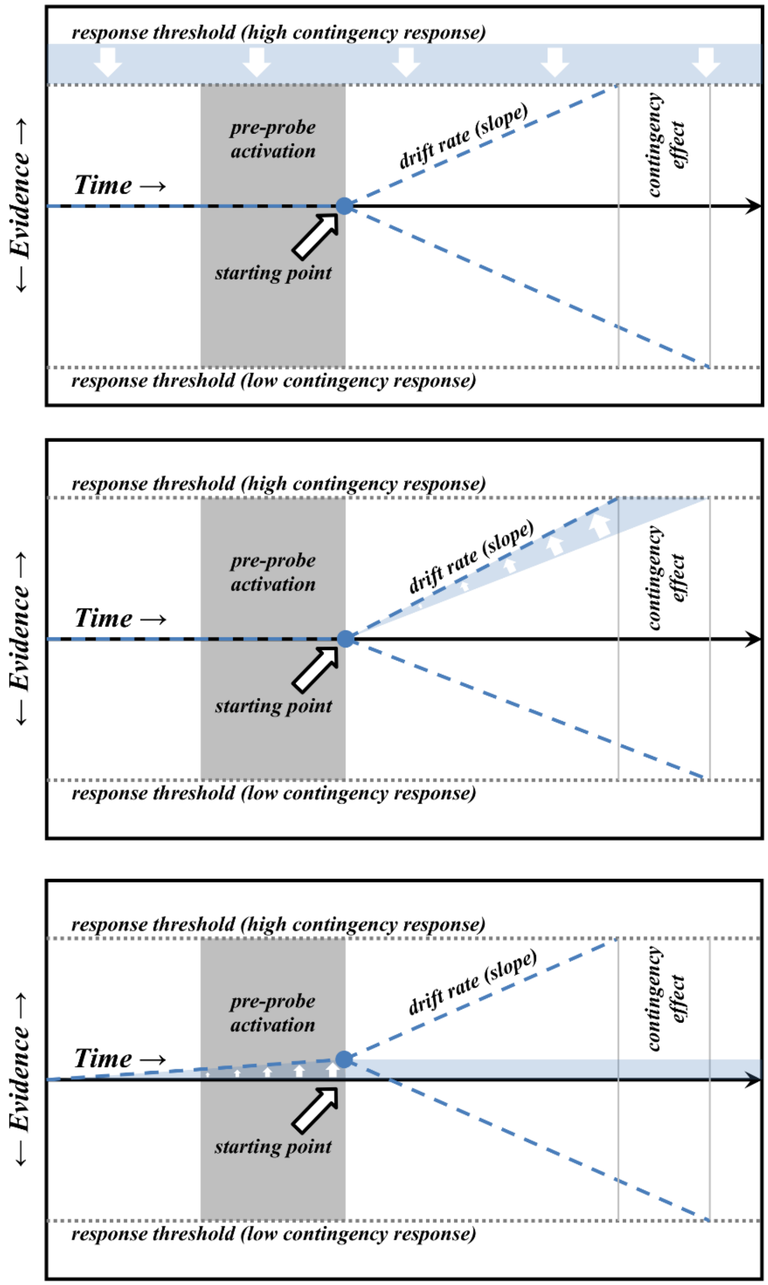

Figure 1. Illustration of how changing the (a) response threshold, (b) drift rate, or (c) starting point can lead to contingency learning effects in response time.

To test this account, Schmidt and Besner (2008, Experiment 2) included a mixture of high, medium, and low contingency trials wherein a prime word appeared with a color (e.g., blue) relatively frequently (50%), less frequently (25%), or infrequently (17%), respectively.

Several findings suggested a selective reduction of the response threshold in high-contingency trials without a corresponding increase of the response threshold in low-contingency trials[2]. For example, mean response time (RT) differed between high- and medium-contingency trials but not between low- and medium-contingency trials. Although some data suggest that contingency learning also indexes a cost to performance in low-contingency trials (Forrin & MacLeod, 2018; Lin & MacLeod, 2018), a model variant wherein low-contingency response thresholds increase as high-contingency response thresholds decrease could likely explain these data.

Other data suggest that contingency learning indexes changes in drift rate and/or starting point over and above any changes to the response threshold. To illustrate these mechanisms, we again use the color-word contingency task. First, identifying the prime (e.g., search) may trigger an increase of the rate at which evidence drifts toward the threshold for the high-contingency response (e.g., the response corresponding to the color purple) (**Fig. 1b**). That is, evidence may accumulate more rapidly for the high-contingency response than for any of the low-contingency responses. Second, identifying the prime may trigger a shift in the starting point of evidence accumulation toward the threshold for the high-contingency response (**Fig. 1c**). Relevant findings come from cueing tasks wherein an initial cue – analogous to a prime – predicts with greater-than-chance accuracy the likely perceptual category (e.g., face or house) and response (e.g., left or right key) for an upcoming target – analogous to a probe. In some tasks, the initial cue speeds performance by increasing the drift rate for the cued category *and* moving the starting point of evidence accumulation toward the cued response threshold (Dunovan et al., 2014; Dunovan & Wheeler, 2018). In other tasks, the cue speeds performance only by moving the

---

[2] The finding may alternatively be explained by the fact that the relative frequencies with which high- and medium-contingency trials appeared (50% vs 25%) differed more than the relative frequencies with which medium- and low-frequency trials appeared (25% vs 17%).

starting point of evidence accumulation toward the cued response threshold (Mulder et al., 2012, 2014). These findings motivate *drift rate* and *starting point* accounts of contingency learning. We note, however, that participants were explicitly informed of the cue-target contingencies in the studies described above. In contrast, most studies of contingency learning use *incidental learning* paradigms wherein participants must extract predictive contingencies on their own.

Recent findings from a study of the Simon task further motivate a starting point account of contingency learning (Luo et al., 2022). Participants indicated the color of a square (red or green) on the left or the right side of the screen by pressing a key on the left or right side of the keyboard. In some of the blocks, the square's location (e.g., left) predicted the corresponding response (e.g., left) 75% of the time and the non-corresponding response (e.g., right) 25% of the time. In other blocks, the square's location predicted the non-corresponding response 75% of the time and the corresponding response 25% of the time. Consistent with prior findings (Bugg, 2014; Logan & Zbrodoff, 1979; Spinelli & Lupker, 2023), Luo and colleagues reported a smaller Simon effect (i.e., a smaller difference in performance between non-corresponding and corresponding trials) in mostly incongruent (vs mostly congruent) blocks. Most important for present purposes, the authors also reported that changes to the starting point in a drift-diffusion modeling framework best explained the contribution of contingency learning to this interaction. These findings are especially informative because, unlike in the cueing studies that we described earlier, the authors measured contingency learning effects using an incidental learning paradigm.

A response preparation account of contingency learning

Researchers often associate changes in starting point with response preparation; that is, with activating the response toward which the starting point moves before the imperative stimulus appears (de Lange et al., 2013; Urai & Donner, 2022; White & Poldrack, 2014). Changing the starting point, however, is mathematically equivalent to changing the response threshold when modelling response times in a drift-diffusion model (Luo et al., 2022; Ratcliff & Smith, 2004). Therefore, it is important to seek converging evidence for the view that changes in starting point are associated with response preparation.

Converging evidence for this view comes from neural data. First, the ability to predict an upcoming finger response via the left or right hand, which changes the starting point (White & Poldrack, 2014), biases activity toward contralateral (vs ipsilateral) regions of the motor cortex (Corbett et al., 2023; Eimer et al., 1996; Kelly et al., 2021). Second, estimating the starting point using such pre-stimulus, preparatory motor activity improves the fit of drift-diffusion models to behavioral data relative to standard approaches that do not incorporate such activity (Kelly et al., 2021). Third, and related, estimates of the starting point correlate with lateralized oscillatory activity in the motor cortex (Urai & Donner, 2022). These data support the view that changes in starting point are associated with response preparation (Urai & Donner, 2022).

The data above show that (1) contingency learning is associated with changes in starting point and (2) changes in starting point are associated with neural measures of response preparation. These findings suggest the possibility that contingency learning involves proactively preparing the probe response that is most frequently associated with the prime. Neural measures of response preparation, however, are often specific to a hand (e.g., the left hand) rather than to an effector (e.g., the left index finger). This is an important limitation of prior work because contingency learning in most keyboard-based tasks involves specific effectors. Direct support for a response preparation account of contingency learning in such tasks would come from data showing that contingency learning is associated with effector-specific response preparation.

Findings from studies of mouse tracking (Bruhn, 2013; Bruhn et al., 2014) support the response preparation account but fall short of providing evidence for effector-specific response preparation. Here, the identity of a central prime predicts with greater-than-chance accuracy the lateralized location (top left or top right) of an upcoming probe. The participant's task is to move the mouse from a central location at the bottom of the screen to the probe's location. Critically, participants start moving the mouse toward the probe location that is most often associated with the prime *before* the probe appears. These pre-probe effects are more consistent with a response preparation account than with a response threshold account. The logic here is that changing a response threshold should influence performance *after* – but not *before* – the probe appears. More specifically, reducing the threshold for the high-contingency response should reduce the amount of perceptual evidence that must accumulate to trigger the high-contingency response *after* probe onset without increasing the activation of that response *before* probe onset.

The findings above, however, are limited in two ways. First, it is unclear whether they generalize to keyboard-based tasks wherein participants use different effectors to indicate their responses. Since participants use an entire hand to make a mouse movement, contingency learning effects may index hand-related, rather than effector-specific, response preparation. Second, the experimental designs do not require participants to extract predictive relationships between the prime and probe stimuli on their own as in typical *incidental learning* paradigms (Schmidt, 2013, 2019). In one study, the prime's perceptual characteristics could have suggested a left or a right response (Bruhn, 2013). In the other study, the researchers explained to participants the manner in which the different primes predicted the different probes (Bruhn et al., 2014). The issue here is not whether participants become aware of contingencies. Indeed, prior findings indicate that contingency learning can occur either *with* conscious awareness (Arunkumar et al., 2022) or *without* conscious awareness (Schmidt et al., 2007). The issue is whether participants must detect contingencies on their own via incidental learning. The main goal of the present study, therefore, is to distinguish between the threshold only and response preparation accounts of contingency learning while overcoming the limitations described above.

Sequential contingency effects

Behavioral measures of contingency learning – as indexed by differences in performance between high- and low-contingency trials – are smaller after low-contingency trials than after high-contingency trials (Schmidt et al., 2007). These *sequential contingency effects* suggest that the previous trial's contingency (high vs low) exerts a strong influence on contingency learning over and above any additional influence of block-wide contingencies. Although such effects are rarely the subject of experimental inquiry, the broader literature suggests they may index processes related to (a) retrieving previous stimulus-response episodes from episodic memory (Giesen et al., 2019; Schmidt et al., 2020) and/or (b) adapting to violations of expectations (Schmidt et al., 2007). We will return to these accounts later. Most important for now, sequential contingency effects provide an additional opportunity to distinguish between the response preparation and threshold only accounts of contingency learning. Indeed, while a single account (e.g., the response preparation account) may explain both overall and sequential effects of contingency learning, it is also possible that different accounts explain these distinct effects.

The present study

The present goal is to distinguish between the threshold only and response preparation accounts of contingency learning while overcoming the limitations of prior mouse tracking studies. That is, we investigate whether these pre-probe force effects appear in a keyboard-based, incidental learning task wherein we do not explicitly instruct participants to learn contingent regularities. First, we investigate whether, analogous to *pre-probe mouse movements* toward the most likely location of an upcoming probe (Bruhn, 2013; Bruhn et al., 2014), contingency learning is associated with greater *pre-probe force* on the key subjects usually press (vs do not press) to indicate the most likely identity of an upcoming probe. Second, we investigate whether there are sequential contingency effects in pre-probe force that correspond to sequential contingency effects in behavior. Critically, unlike in prior mouse tracking studies, we investigate whether these pre-probe force effects appear (a) with primes whose perceptual characteristics are unlikely to suggest a specific probe response, and (b) without informing participants about the predictive relationships between primes and probes. To measure pre-probe force, we use force-sensitive keyboards that can detect small changes of force before probe onset (Weissman, 2019).

To investigate contingency learning, we employ a prime-probe task wherein participants respond to both the prime and the probe in each trial. In our view, requiring participants to attend to both the prime and the probe may facilitate the learning of contingent regularities between these successive stimuli, thereby increasing contingency learning effects and making them easier to detect, if they are present, in pre-probe force. Consistent with this view, participants learn contingent regularities better when the associated stimuli are attended (vs not attended) (Cox & Aimola Davies, 2022). Further, although participants are not required to identify the prime in the color-word contingency task, contingency learning contributes to performance in many tasks that require attending to successive stimuli including segmenting speech into words (Saffran et al., 1996), artificial grammar learning (Reber, 1967), and sequence learning (Eimer et al., 1996; Fiser & Aslin, 2002). Consequently, employing a task wherein participants attend to successive prime and probe stimuli is consistent with the broader literature on contingency learning.

The response preparation and threshold only accounts make distinct predictions regarding pre-probe force. The response preparation account posits that participants prepare the high contingency response after prime onset but before the probe appears, and that such preparation is greater after high- (vs low-) contingency trials. Therefore, this account predicts greater pre-probe force on the high-contingency response key than on the low-contingency response key, and that the magnitude of this pre-probe force effect will be greater after high-contingency trials than after low-contingency trials. The threshold only account, on the other hand, does not predict such changes in pre-probe force. Consequently, observing the pre-probe force effects described above would favor the response preparation account over the threshold only account. Of course, observing evidence for response preparation would not exclude the possibility that other mechanisms also contribute to contingency learning effects, but our method allows us to evaluate evidence for response preparation independent of other mechanisms that predict null effects.

Experiment 1

In Experiment 1, we use a prime-probe arrow task to investigate whether contingency learning is associated with changes in pre-probe finger force. Participants indicate via two independent responses in each trial whether (1) an initial prime arrow points left (50%) or right

(50%) and (2) a subsequent probe arrow points up or down. For half the subjects, a leftward pointing prime arrow usually precedes an upward pointing probe arrow (high-contingency trials) but occasionally precedes a downward pointing probe arrow (low-contingency trials). In contrast, a rightward pointing prime arrow usually precedes a downward pointing probe (high-contingency trials) but occasionally precedes an upward pointing probe arrow (low-contingency trials). For the other half of subjects, we reverse these contingencies.

In each trial, two force-sensitive keys are critical for testing our hypotheses. The high-contingency key (e.g., J) is the key subjects usually press to indicate the direction in which the probe points (e.g., Up) after they respond to a particular prime (e.g., Left). The low-contingency key (e.g., N) is the key subjects usually do *not* press to indicate the direction in which the probe points (e.g., Down) after they respond to a particular prime (e.g., Left). Which key serves as the high-contingency key (e.g., J) and which key serves as the low-contingency key (e.g., N) varies randomly across trials with whether the prime arrow points left or right. Therefore, each key (i.e., the J key and the N key) serves equally often as the high- and low-contingency key across trials. As we discussed earlier, the response preparation account predicts greater sub-threshold force on the high-contingency key than on the low-contingency key just before the probe appears. It also predicts a reduction of this pre-probe force effect following low- (versus high-) contingency trials. In contrast, the threshold only account does not predict such changes in pre-probe force.

Methods

Participants

We determined the sample size using data from an unpublished study in which we measured force 0-200 ms before the second of two arrows in a task resembling the present one. The two arrows were mostly congruent (pointed in the same direction 87.5% of the time) or mostly incongruent (pointed in opposite directions 87.5% of the time). Although "contingency" was confounded with "congruency" in this study, we measured robust changes of anticipatory response 0-200 ms before the second arrow appeared. That is, just before the second arrow appeared, subjects pressed harder on the high-contingency key than on the low-contingency key. Assuming that all of this effect (partial eta squared = 0.35[3]) indexes contingency learning, G*Power 3.1.9.7 (Faul et al., 2007) estimates that 28 subjects are needed to observe a corresponding effect with 95% power and an alpha of 0.05. We conservatively assume, however, that only some of this effect (partial eta squared = 0.32) reflects contingency learning processes. We make this assumption because congruency on its own produced a small anticipatory force effect that was associated with a partial eta squared value of 0.12. For a partial-eta-squared value of 0.32, G*Power estimates that 31 subjects are needed to observe a corresponding effect with 95% power and an alpha of 0.05. Ultimately, we decided to collect usable data from 32 subjects.

Thirty-two students (10 male, 22 female; mean age, 19.0 years; age range: 18-23 years) from the University of Michigan's Psychology Subject Pool participated for course credit. No participants were excluded (e.g., for performing the task with less than 75% overall accuracy).

[3] We mistakenly reported this partial-eta-squared value as 0.41 in our pre-registration.

The University of Michigan's Health Sciences and Behavioral Sciences Institutional Review Board deemed the study was exempt from oversight.

Stimuli and Apparatus

Large and small arrows served as prime and probe stimuli, respectively. The set of large arrows included a leftward pointing arrow (8.0° x 4.5°), a rightward pointing arrow (8.0° x 4.5°), an upward pointing arrow (4.5° x 8.0°), and (4) a downward pointing arrow (4.5° x 8.0°). The set of small arrows included a leftward pointing arrow (2.5° x 1.7°), a rightward pointing arrow (2.5° x 1.7°), an upward pointing arrow (1.7° x 2.5°), and a downward pointing arrow (1.7° x 2.5°). We ran the experiment using PsychoPy version 2022.2.4 (Peirce et al., 2019) on a Windows 10 PC.

We used custom response boxes to collect analog and digital response force continuously (i.e., at 500 Hz) throughout the experiment. Analog force changes continuously with finger pressure. In contrast, digital response force (i.e., key actuation) is recorded only when a response key is fully pressed (threshold response force for a full keypress is 60 cN). Each response box has five keys that measure analog and digital force and two standard keys that measure only digital force. The former keys (F, G, J, K, and N) are spaced as on a QWERTY keyboard. Each of these keys reliably detects analog changes in mass as small as 100 mg. The two standard keys (space bar and escape) appear at the bottom and top left corner of the box, respectively. Weissman (2019) provides a more detailed, technical description of the response boxes. Finally, we use custom Python software to transfer information between the response box and PsychoPy.

Before the study, we calibrated each force-sensitive key as follows. First, we recorded the load cell (i.e., key) output for 1g, 2g, 5g, 10g, and 20g masses. Second, we used linear regression to determine the slope of the best-fitting line relating mass to load cell output. Third, we used this slope to convert mass to centinewtons (cN) as follows: cN = 100*(mass in kg x 9.8 m/s^2). To verify that each key remained functional throughout the study, we recorded the load cell (i.e., key) output with (a) no mass and (b) a 5g or 20g mass before running each participant.

Task

At the beginning of the block, there was a fixation cross for 1.8 seconds. Next, a blank screen replaced the fixation cross for 0.2 seconds. After this initial 2-second period, the trial sequence began. There were four sequential events in each 3-second trial (**Fig. 2**): the prime arrow (duration, 100 ms), a blank screen (duration, 1033 ms), the probe arrow (duration, 100 ms), and a second blank screen (duration, 1733 ms).

Participants had a maximum of 900 ms to respond to each arrow after it appeared. Specifically, using the custom response box described earlier (Weissman, 2019), we instructed participants to press “f” (left middle finger), “g” (left index finger), “j” (right middle finger), or “n” (right index finger) to indicate whether each arrow pointed “left”, “right”, “up”, or “down”, respectively[4]. Put another way, we instructed participants to use the same S-R mapping to make two independent responses in each trial: one to the prime arrow and one to the probe arrow. If participants responded before an arrow appeared, pressed the wrong key after an arrow appeared, took more than 900 ms to respond to either arrow, or did not respond at all to either arrow, the

[4] To be consistent with our prior studies of the prime-probe task, we used the same key-hand mapping for all participants, rather than mapping the left and right arrow-direction keys to the left hand and the up and down arrow-direction keys to the right hand in half the participants and using the opposite key-hand mapping in the other half. Using a constant mapping does not lead to a design confound because we use the same mapping in all conditions.

word “Error” appeared in white during the final 200 ms of the trial. If participants responded correctly within the 900 ms deadline, only a blank screen appeared during the final 200 ms.

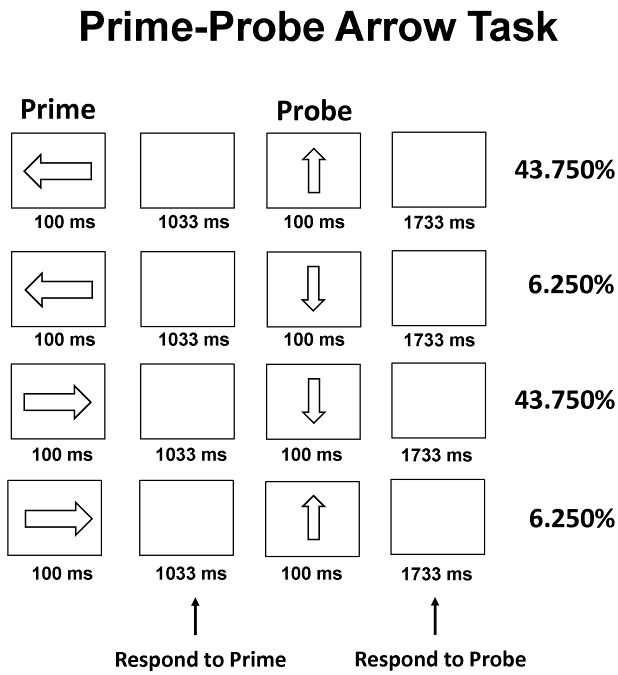

Figure 2. The four trial types in Experiment 1. Participants indicate the direction in which the first (i.e., prime) arrow and second (i.e., probe) arrow point by making two different keypresses. The time listed beneath each box in the bottom row indicates the duration of the corresponding trial component. Values in BOLD on the far right show the percentage of trials in which each trial type appears. Stimuli not drawn to scale.

Experimental Design

We used a within-participants design with a single factor: contingency (high, low). The study began with an “unbiased” block of 32 practice trials wherein left and right prime arrows preceded up and down probe arrows equally often. Next, there were eight 64-trial blocks of test trials (512 test trials in total). For half the subjects, left arrow primes usually preceded up arrow probes while right arrow primes usually preceded down arrow probes. Specifically, left arrow primes preceded up arrow probes 28 times (high contingency trials) and preceded down arrow probes four times (low contingency trials). In contrast, right arrow primes preceded down arrow probes 28 times (high contingency) and preceded up arrow probes four times (low contingency). We reversed these contingencies for the other half of subjects, such that left arrow primes usually preceded right arrow probes while right arrow primes usually preceded up arrow probes. In sum, high-contingency trials (87.5%) appeared seven times more often than low-contingency trials (12.5%) in each block. The 64 trials in each block appeared in a random order.

Procedure

Participants read a consent form on the computer screen and pressed one key to provide consent or another key to abort the experiment (no participants chose to abort the experiment). A research assistant positioned each participant's head within a chinrest such that their pupils were about 55 cm from the computer screen. The research assistant also explained the task described earlier, which involved responding to each arrow (i.e., the prime arrow and the probe arrow) immediately after its appearance as quickly as possible without making mistakes. The research assistant also explained the 900 ms response deadline for responding to each arrow. However, the research assistant did not inform participants about the contingency manipulation.

Data Analyses

We employed JASP 0.18.0.0 (JASP Team, 2023). To analyze the behavioral data, we conducted a repeated-measures analysis of variance (ANOVA) with contingency (high, low) as a factor. We analyzed mean probe response time (RT) and mean probe error rate (ER) in separate

ANOVAs. As we described earlier, high-contingency trials were those wherein the probe arrow (e.g., up) usually followed the prime arrow (e.g., left). In contrast, low-contingency trials were those wherein the probe arrow (e.g., down) usually did not follow the prime arrow (e.g., left).

To analyze the pre-probe force data, we conducted a repeated-measures ANOVA with response key (high-contingency, low-contingency) as a factor. The high-contingency key (e.g., J) is the key that subjects usually press to indicate the direction in which the probe points (e.g., Up) after they respond to a particular prime (e.g., Left). The low-contingency key (e.g., N) is the key that subjects usually do *not* press to indicate the direction in which the probe points (e.g., Down) after they respond to a particular prime (e.g., Left). As we explained earlier, which key serves as the high-contingency key (e.g., J) and which key serves as the low-contingency key (e.g., N) in any given trial varies randomly with whether the prime arrow points left or right. Thus, each key (i.e., the J key and the N key) serves equally often as the high-contingency key and as the low-contingency key across trials. For this reason, our analyses involve comparing pre-probe force on a given key (e.g., the J key) when it serves as the high-contingency key (in 50% of the trials) to force on the same key when it serves as the low-contingency key (in the other 50% of the trials).

The dependent measure for the ANOVA was average response force 0-200 ms before the probe arrow appeared (Weissman, 2019). We chose this interval because it occurs just before the probe arrow appears, which is when proactive response preparation related to the upcoming probe, if present, should be greatest. Finally, we note that this analysis compares pre-probe force on two *keys* (i.e., the key that would be correct for a high-contingency trial versus the key that would be correct for a low-contingency trial), rather than on two trial types. Whether the *trial* was (eventually) high- or low-contingency was irrelevant for this analysis, because we measured force after prime onset but *before* probe onset.

In an exploratory analysis, we analyzed sequential contingency effects in behavior and pre-probe response force. To analyze such effects in behavior, we analyzed mean RT and mean ER in separate repeated-measures ANOVAs with previous trial contingency (high, low) and current trial contingency (high, low) as factors. To analyze such effects in pre-probe response force, we analyzed mean force 0-200 ms before probe onset in a repeated-measures ANOVA with previous trial contingency (high, low) and (2) response key (high-contingency, low-contingency) as factors.

We excluded certain trials from these analyses. In the analysis of overall contingency learning effects in *mean probe RT*, we excluded (a) practice trials, (b) trials in which a computer malfunction produced an incorrect or absent RT value (1.8%), and (c) trials after such computer malfunctions (1.4%). Among the remaining trials, we excluded (d) trials with errors or omitted responses (6.6%), (e) trials following trials with errors or omitted responses (5.7%), and (f) outliers (3.4%). We defined outliers as trials with probe RTs greater than 3*Sn (Rousseeuw & Crouz, 1993), calculated separately for each of the two conditions in our experimental design. In the analysis of overall contingency learning effects in *mean probe ER*, we excluded the same types of trials except for trials with errors or omitted responses because this criterion was the dependent measure in our analysis of mean ER. In this analysis, 3.8% of the trials were outliers[5]. In the overall analysis of mean pre-probe force (0-200 ms before probe onset), we excluded trials in which a computer malfunction produced an incorrect or absent RT value and trials after such computer malfunctions as in the other analyses. Of the remaining trials, we only included (1) trials with correct responses to the prime arrow[6] and (2) trials following trials with correct

[5] We identified outliers using both correct *and* error RTs in the analysis of the mean ER data, not only correct RTs as in the analysis of the mean RT data. Thus, these analyses produced slightly different percentages of outliers.
[6] In our pre-registration, we stated that we would analyze force in trials with correct responses to both the prime *and* the probe. An anonymous reviewer pointed out a valid concern with this approach, however. Excluding trials with

responses to both of the arrows. Finally, in the exploratory analyses of sequential effects, we excluded the same types of trials from the mean RT, mean ER, and mean pre-probe force data. The percentages of excluded trials were virtually identical to those above but differed slightly (~0.1% or less) in a few cases (e.g., because there were fewer trials per condition in each of the four trial types than in each of the two trial types in the analysis of overall contingency effects).

Transparency and Openness

In our online pre-registration document, we indicate the rationale for the sample sizes, manipulations, dependent measures, and data exclusions that we employed. We further note that we follow JARS (Kazak, 2018). The preregistration, task scripts, data analysis scripts, and raw data are freely available on the Open Science Framework (OSF) (https://osf.io/jd9ys/).

Results

Overall effects of contingency learning

Mean RT

We observed a main effect of contingency, $F(1,31) = 58.85$, $p < 0.001$, $\eta_p^2 = 0.66$: mean RT was faster in high-contingency versus low-contingency trials (294 ms vs 338 ms) (**Fig. 3a**).

---

an incorrect response to the probe could bias the results to show greater pre-probe force on the high-contingency response key than on the low-contingency response key. Indeed, randomly increasing pre-probe force on the high-contingency response key (e.g., due to guessing) could facilitate a correct probe response such that the trial is included in the analysis. In contrast, randomly increasing pre-probe force on the low-contingency response key could facilitate an incorrect probe response such that the trial is not included. To prevent such a bias, we analyzed pre-probe force regardless of whether the subsequent probe response was correct. We note that this change to our pre-registered analysis plan did not change any of the inferences that we describe below.

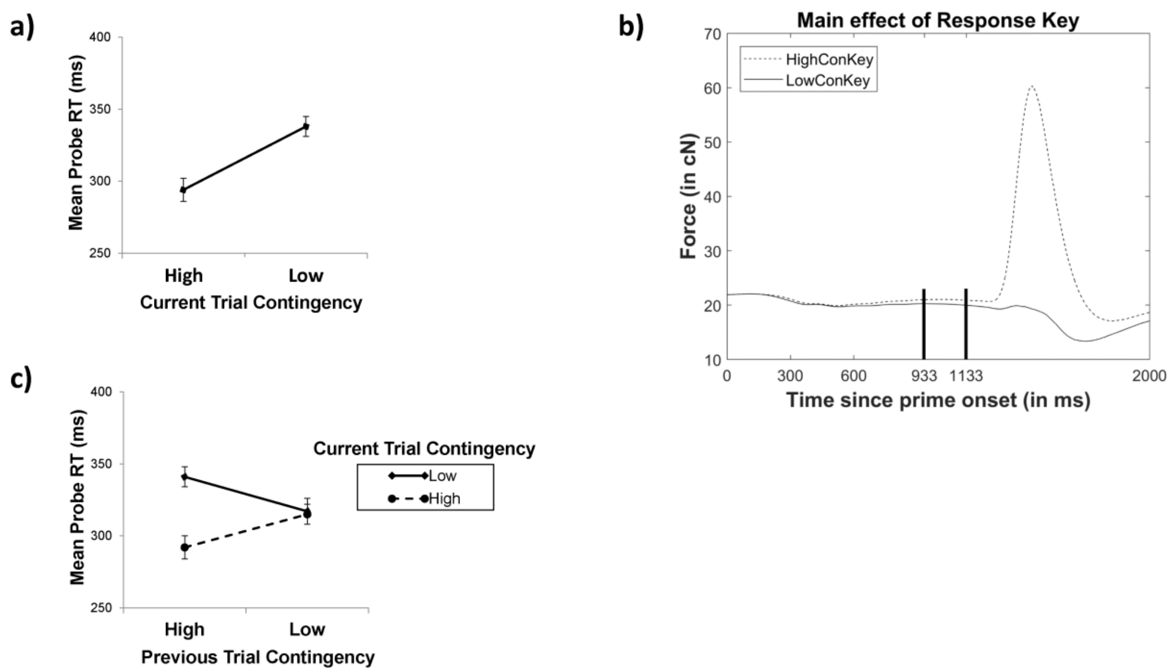

Figure 3. The main results of Experiment 1. (a) Mean probe RT on the y-axis as a function of current trial contingency (high, low) on the x-axis. (b) Force on the y-axis for the high-contingency (dashed line) and low-contingency (solid line) response keys starting at prime onset. The two vertical lines highlight the period of time (933 ms – 1133 ms) that we used to calculate mean pre-probe force (c) Mean probe RT on the y-axis as a function of previous trial contingency on the x-axis and current trial contingency (high contingency, dashed line; low contingency, solid line) on the x-axis. Error bars indicate ± 1 standard error of the mean.

Mean ER

We observed a main effect of contingency, $F(1,31) = 26.20$, $p < 0.001$, $\eta_p^2 = 0.46$. As expected, mean ER was lower in high-contingency trials (4.9%) than in low-contingency trials (13.1%).

Mean Pre-Probe Force

Figure 3b plots analog response force across time, starting at prime onset, on the high- and low-contingency response keys. The large peak for the high-contingency key, which appears about 300 ms *after* the probe onsets at 1133 ms, occurs because the high-contingency key is the correct response in most of the trials. Consequently, in most of the trials, the participant fully presses the high-contingency key, which explains the large peak. In contrast, the low-contingency key is the incorrect response in most of the trials. Therefore, the participant rarely presses the low-contingency key, which explains the absence of a large peak for this key. As described in the Methods section, which key serves as the high-contingency key (e.g., the J key) and which key serves as the low-contingency key (e.g., the N key) varies randomly across trials.

Our analyses of *pre-probe* force focus on the interval between the two vertical solid lines on the right side of Figure 3b, which represents 0-200 ms *before* probe onset. The response preparation account predicts greater sub-threshold force on the high-contingency key than on the low-contingency key during this interval. Consistent with this prediction, we observed a main effect of contingency, $F(1,31) = 25.19$, $p < 0.001$, $\eta_p^2 = 0.45$ (**Fig. 3b**). Mean pre-probe force was greater on the high-contingency key (21.01 cN) than on the low-contingency key (20.17 cN).

Exploratory analysis: Sequential effects of contingency learning

Mean RT

As in the overall analysis, we observed a main effect of contingency, $F(1,31) = 34.19$, $p < 0.001$, $\eta_p^2 = 0.52$. Critically, we also observed an interaction between previous trial contingency and current trial contingency, $F(1,31) = 22.69$, $p < 0.001$, $\eta_p^2 = 0.42$ (**Fig. 3c**). Mean RT was significantly faster in high-contingency trials (292 ms) than in low-contingency trials (341 ms) after high-contingency trials, $F(1,31) = 58.60$, $p < 0.001$, $\eta_p^2 = 0.65$, but this was not the case after low-contingency trials (315 ms vs 317 ms; $F(1,31) < 1$). No other effects were significant.

Mean ER

As in the overall analysis, we observed a main effect of contingency, $F(1,31) = 16.82$, $p < 0.001$, $\eta_p^2 = 0.35$. However, we also observed a main effect of previous trial contingency, $F(1, 31) = 4.94$, $p = 0.034$, $\eta_p^2 = 0.14$. Critically, we observed an interaction between previous trial contingency and current trial contingency, $F(1,31) = 10.02$, $p = 0.003$, $\eta_p^2 = 0.24$. Mean ER was significantly lower in high-contingency trials (4.8%) than in low- contingency trials (14.3%) after high-contingency trials, $F(1,31) = 25.53$, $p < 0.001$, $\eta_p^2 = 0.45$. However, this was not the case after low-contingency trials (5.9% vs 7.0% $F(1,31) < 1$). No other effects were significant.

Mean Pre-Probe Force

As in the overall analysis, we observed a main effect of response key, $F(1,31) = 9.53$, $p = 0.004$, $\eta_p^2 = 0.24$. Critically, we also observed an interaction between previous trial contingency and response key, $F(1,31) = 12.56$, $p < 0.001$, $\eta_p^2 = 0.30$ (**Fig. 4**). Mean pre-probe force was significantly greater on the high (21.09 cN) vs low (20.16 cN) contingency key after high-contingency trials, $F(1,31) = 26.85$, $p < 0.001$, $\eta_p^2 = 0.46$ (**Fig. 4a**). However, this effect did not occur after low-contingency trials (20.71 cN vs 20.64 cN; $F(1,31) < 1$; **Fig. 4b**). Figure 4c illustrates this interaction more specifically.

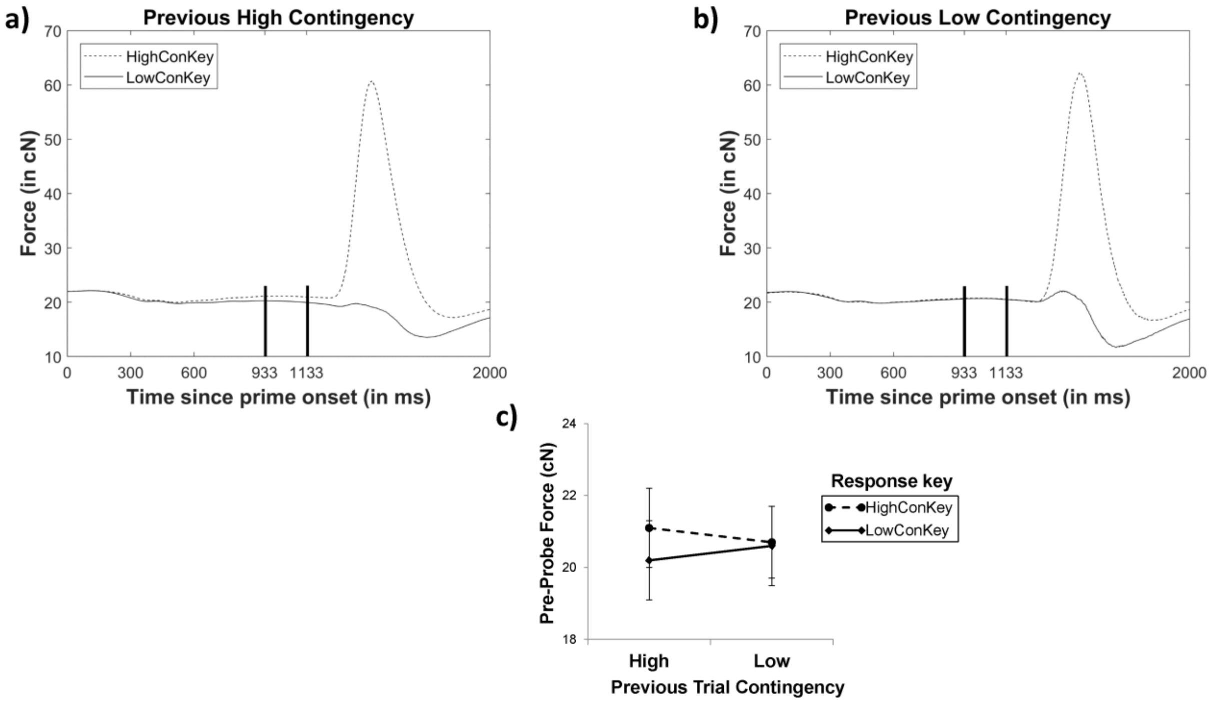

Figure 4. Sequential contingency effects in force in Experiment 1. Force on the y-axis for the high-contingency (dashed line) and low-contingency (solid line) response keys starting at prime onset after (a) high-contingency trials and (b) low-contingency trials. The two vertical lines in each panel highlight the period of time (933 ms – 1133 ms) that we used to calculate mean pre-probe force. (c) Mean pre-probe force on the y-axis as a function of previous trial contingency on the x-axis and current trial contingency (high contingency, dashed line; low contingency, solid line) on the x-axis. Error bars indicate ± 1 standard error of the mean.

Discussion

The results of Experiment 1 favor the response preparation account of contingency learning over the threshold only account. In particular, we observed greater force on the high- vs low-contingency key 0-200 ms *before* probe onset. This outcome indicates that contingency learning is associated with greater preparation of the high- (vs low-) contingency response just before probe onset, which is an effect that the threshold only account does not predict.

Exploratory analyses further revealed sequential contingency effects: behavioral and anticipatory force measures of contingency learning were greater after high-contingency trials than after low-contingency trials (Schmidt et al., 2007). In fact, these measures were no longer significant after low-contingency trials. These findings suggest that the previous trial's contingency (high vs low) exerts a strong influence on contingency learning over and above any potential influence of block-wide contingencies. Given that these findings came from an exploratory analysis, however, we sought to replicate them before drawing firm conclusions.

Experiment 2

The goal of Experiment 2 was to address two limitations of Experiment 1 that relate to the exploratory analyses of sequential contingency effects. First, we did not pre-register these analyses. Second, in our random trial sequences, one would expect two consecutive low-contingency trials to appear just 8 times across the entire experiment, which is a relatively low trial count on which to base estimates of mean RT and mean ER. In comparison, one would expect a low-contingency trial to appear after a high-contingency trial, or vice-versa, 56 times across the entire experiment and two consecutive high-contingency trials to appear 392 times.

To address these limitations in Experiment 2, we pre-registered our analyses of sequential trial effects and changed the relative proportions of high- and low-contingency trials. With respect to the latter point, high-contingency trials appeared 81.25% of the time – rather than 87.5% – while low-contingency trials appeared 18.75% of the time – rather than 12.5%. That is, high-contingency trials appeared 4.33 times more often than low-contingency trials rather than 7 times more often. We also added a block of 64 test trials to increase power. In the resulting random trial sequences, one would expect two consecutive low-contingency trials to appear about 20 times, a low-contingency trial to appear after a high-contingency trial, or vice-versa, to appear about 88 times, and two consecutive high-contingency trials to appear about 380 times.

Methods

Participants

To determine the sample size, we made use of the data from Experiment 1. Here, we observed robust contingency learning effects in mean RT and mean ER. We also observed greater mean anticipatory force on the high- vs low-contingency response key 0-200 ms before the probe arrow appeared on the screen. Finally, we observed sequential contingency effects in behavior and anticipatory response force, showing that the contingency learning effects described above were greater after high- vs low-contingency trials. The smallest of all these effect sizes (corresponding to the sequential contingency effect in anticipatory response force) had a partial-eta-squared value of 0.32. G*Power estimated that 31 subjects would be needed to observe such an effect size with 95% power (alpha = 0.05). As contingency learning effects might be smaller in Experiment 2, however, we also determined that 31 subjects would be needed to observe a partial-eta-squared value of 0.22 with 80% power (alpha = 0.05).

As we were nearing the end of the semester, we decided to use a stopping rule to determine the sample size. Specifically, we pre-registered our study to include a minimum of 32 subjects and a maximum of 51 subjects, which was the number of subjects that we had already scheduled to participate in our study before the end of the semester. We analyzed the data from the 42 subjects who showed up to their scheduled appointment to complete the experiment.

Forty-two students from the University of Michigan's Psychology Subject Pool participated for course credit. None of these subjects had participated in Experiment 1. We excluded the data from one participant who performed the task with less than 75% overall accuracy, leaving 41 participants in the final sample (17 male, 24 female; mean age, 19.1 years; age range: 18-23 years). The University of Michigan's Health Sciences and Behavioral Sciences Institutional Review Board deemed that this study was exempt from oversight.

Stimuli and Apparatus

The stimuli and apparatus were identical to those in Experiment 1.

Task

The task was identical to that in Experiment 1.

Experimental Design

The experimental design was identical to that in Experiment 1 with two exceptions. First, there were nine, rather than eight, blocks of 64 test trials. Second, in each block, high- and low-contingency trials appeared 81.25% (52 trials) and 18.75% (12 trials) of the time, rather than 87.5% (56 trials) and 12.5% (8 trials) of the time. That is, high-contingency trials appeared 4.33 times more often than low-contingency trials, rather than 7 times more often as in Experiment 1. Figure 5 shows the percentage of trials in which each of the prime-probe arrow pairs appeared.

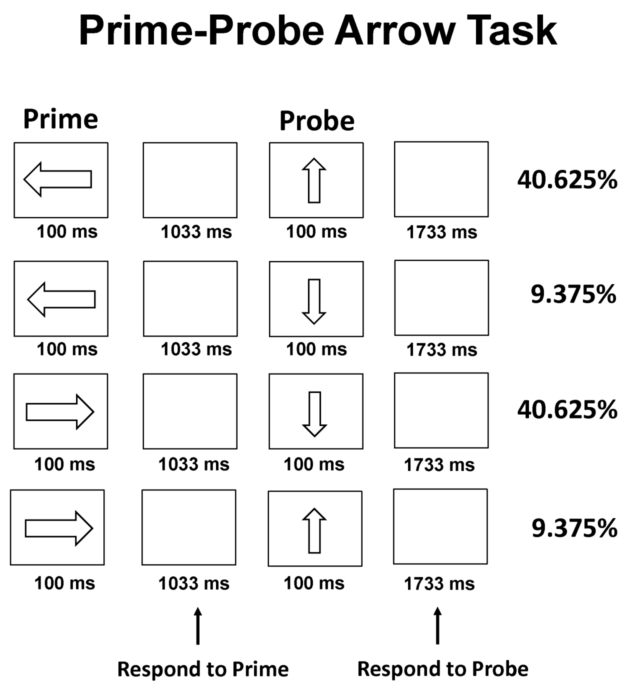

Figure 5. The four trial types in Experiment 2. Participants indicate the direction in which the first (i.e., prime) arrow and second (i.e., probe) arrow point by making two different keypresses. The time listed beneath each box in the bottom row indicates the duration of the corresponding trial component. Values in BOLD on the far right show the percentage of trials in which each trial type appears. Stimuli not drawn to scale.

Procedure

The procedure was identical to that in Experiment 1.

Data Analyses

The data analyses were identical to those described in the corresponding section of Experiment 1. However, we pre-registered not only our analyses of the overall effects of contingency learning but also our analyses of the sequential effects of contingency learning.

We excluded the same types of trials from our analyses of mean RT, mean ER, and mean anticipatory response force as in Experiment 1. In the analysis of overall contingency learning effects, we excluded (a) practice trials, (b) trials in which a computer malfunction produced an incorrect or absent RT value (2.3%), and (c) trials after computer malfunction trials (1.5%). Of the remaining trials, we excluded (d) trials with errors or omitted responses (6.0%) and (e) trials after trials with errors or omitted responses (5.2%). In the analyses of mean RT and mean ER, we further excluded 4.0% and 4.3% of the trials that were outliers, respectively. In the analyses of sequential effects, we excluded the same types of trials. The percentages of excluded trials in each category were virtually identical to those provided above but differed slightly (~0.1% or less) in a few cases (e.g., because we excluded outliers from four trial types, each of which had fewer trials than did each of the two trial types in the analysis of overall contingency effects).

Transparency and Openness

The transparency and openness were identical to those in Experiment 1. The preregistration, task scripts, data analysis scripts, and raw data for Experiment 2 are freely available on the Open Science Framework (OSF) (https://osf.io/eyzux/).

Results

<u>Overall effects of contingency learning</u>

Mean RT

We observed a main effect of contingency, $F(1,40) = 47.44$, $p < 0.001$, $\eta_p^2 = 0.54$ (**Fig. 6a**). As expected, mean RT was faster in high-contingency trials (313 ms) than in low-contingency trials (338 ms).

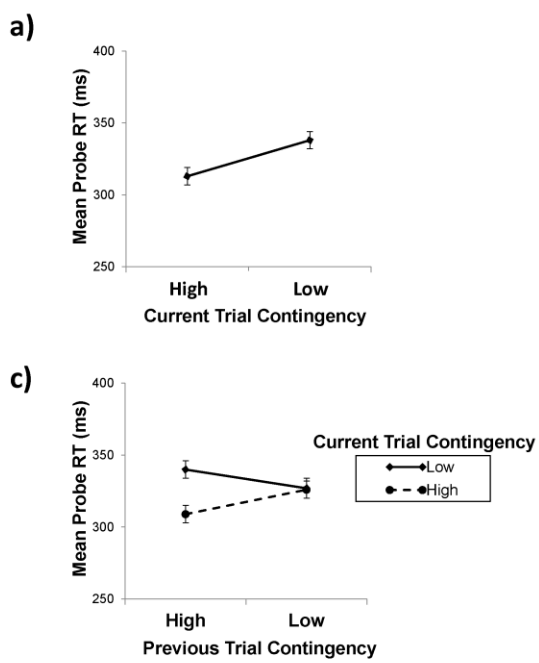

Figure 6. The main results of Experiment 2. (a) Mean probe RT on the y-axis as a function of current trial contingency (high, low) on the x-axis. (b) Force on the y-axis for the high-contingency (dashed line) and low-contingency (solid line) response keys starting at prime onset. The two vertical lines highlight the period of time (933 ms – 1133 ms) that we used to calculate mean pre-probe force (c) Mean probe RT on the y-axis as a function of previous trial contingency on the x-axis and current trial contingency (high contingency, dashed line; low contingency, solid line) on the x-axis. Error bars indicate ± 1 standard error of the mean.

Mean ER

We observed a main effect of contingency, $F(1,40) = 31.62$, $p < 0.001$, $\eta_p^2 = 0.44$. As expected, mean ER was lower in high-contingency trials (4.8%) than in low contingency trials (8.2%).

Mean Pre-Probe Force

Figure 6b plots analog response force across time, starting at prime onset, on the high- and low-contingency keys. As in Experiment 1, our analyses of *pre-probe* force focus on the interval between the two vertical solid lines on the right side of Figure 6b, which represents 0-200 ms *before* probe onset. The response preparation account predicts greater sub-threshold force on the high-contingency key than on the low-contingency key during this interval. Consistent with this prediction, we observed a main effect of contingency, $F(1,40) = 5.98$, $p = 0.019$, $\eta_p^2 = 0.13$ (**Fig. 6b**): mean pre-probe force was greater on the high-contingency key (19.73 cN) than on the low-contingency key (19.40 cN).

Sequential effects of contingency learning

Mean RT

As in the overall analysis, we observed a main effect of contingency, $F(1,40) = 27.52$, $p < 0.001$, $\eta_p^2 = 0.41$. Critically, we also observed an interaction between previous trial contingency and current trial contingency, $F(1,40) = 18.69$, $p < 0.001$, $\eta_p^2 = 0.32$ (**Fig. 6c**). Mean RT was significantly faster in high-contingency trials (309 ms) than in low-contingency trials (340 ms) after high-contingency trials, $F(1,40) = 45.09$, $p < 0.001$, $\eta_p^2 = 0.53$, but this was not the case after low-contingency trials (326 ms vs 327 ms; $F(1,40) < 1$). No other effects were significant.

Mean ER

As in the overall analysis, there was a main effect of contingency, $F(1,40) = 12.28$, $p = 0.001$, $\eta_p^2 = 0.24$. However, there was also a main effect of previous trial contingency, $F(1, 40) = 4.32$, $p = 0.044$, $\eta_p^2 = 0.10$. Critically, there was an interaction between previous trial contingency and current trial contingency, $F(1,40) = 5.66$, $p = 0.022$, $\eta_p^2 = 0.12$. Mean ER was significantly lower in high-contingency trials (4.7%) than in low-contingency trials (8.6%) following high-contingency trials, $F(1,40) = 32.74$, $p < 0.001$, $\eta_p^2 = 0.45$. This was not the case, however, following low-contingency trials (5.1% vs 5.9% $F(1,40) < 1$). No other effects were significant.

Mean Pre-Probe Force

We observed an interaction between previous trial contingency and response key, $F(1,40) = 16.43$, $p < 0.001$, $\eta_p^2 = 0.29$ (**Fig. 7**). Mean pre-probe force was significantly greater on the high (19.82 cN) vs low (19.33 cN) contingency key after high-contingency trials, $F(1,40) = 10.00.$, $p = 0.003$, $\eta_p^2 = 0.20$ (**Fig. 7a**). This pattern was significantly reversed, however, after low-contingency trials (19.36 cN vs 19.84 cN; $F(1,31) = 6.21$, $p = 0.017$, $\eta_p^2 = 0.13$ (**Fig. 7b**). Figure 7c focuses on this interaction more specifically. No other effects were significant.

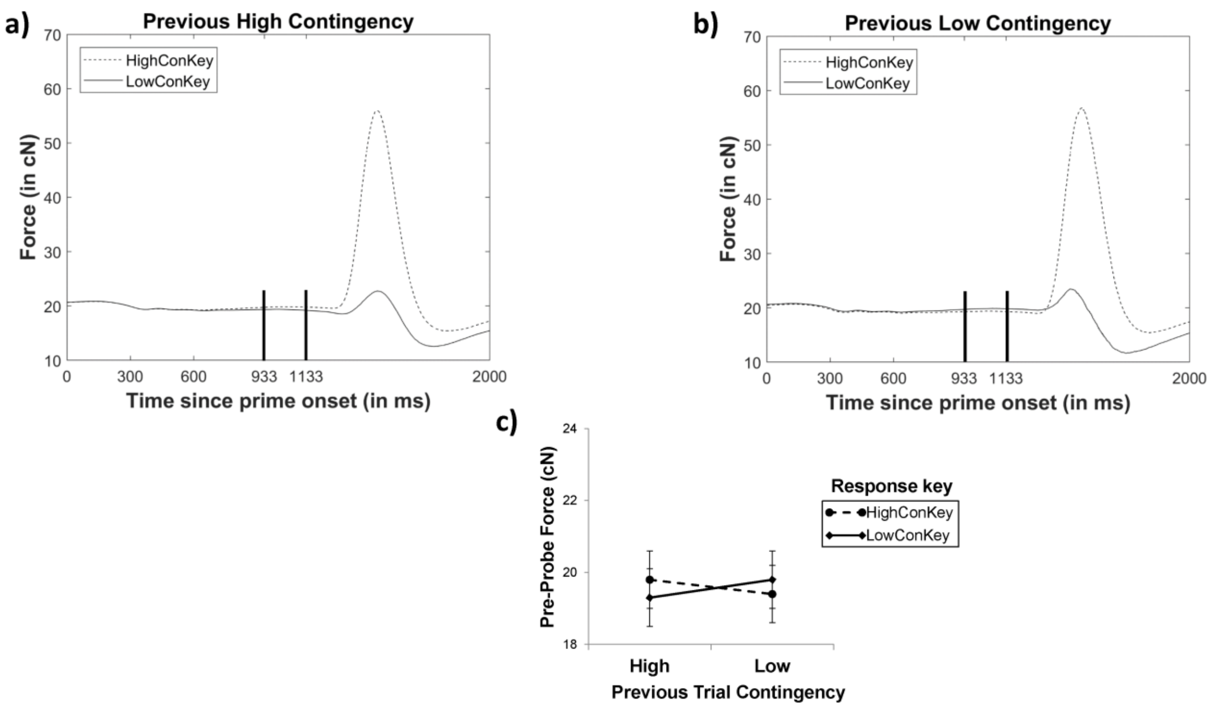

Figure 7. Sequential contingency effects in force in Experiment 2. Force on the y-axis for the high-contingency (dashed line) and low-contingency (solid line) response keys starting at prime onset after (a) high-contingency trials and (b) low-contingency trials. The two vertical lines in each panel highlight the period of time (933 ms – 1133 ms) that we used to calculate mean pre-probe force. (c) Mean pre-probe force on the y-axis as a function of previous trial contingency on the x-axis and current trial contingency (high contingency, dashed line; low contingency, solid line) on the x-axis. Error bars indicate ± 1 standard error of the mean.

Exploratory Analysis #1

One may wonder whether effects of contingency learning on mean probe RT and mean pre-probe force correlate across participants. We would not expect a perfect correlation because processes that take place following probe onset (e.g., perception, decision-making, response selection, etc.) *can* influence probe RT, but *cannot* influence pre-probe force. Nonetheless, it is possible that advance response preparation, as indexed by mean pre-probe force, predicts to some degree the size of contingency learning effects that occur later in mean probe RT.

To investigate this possibility, we employed the data from Experiment 2. Here, the number of low-contingency trials was greater than in Experiment 1. Therefore, one might expect more stable estimates of mean RT in low-contingency trials and, therefore, more stable across-participants correlations between mean RT and mean pre-probe force measures of contingency learning than in Experiment 1.

Exploratory analyses revealed such across-subject correlations in Experiment 2. First, the degree to which mean RT was faster in high- (vs low-) contingency trials correlated with the degree to which mean pre-probe force was greater on the high- (vs low-) contingency key (r = 0.31, p = 0.052). In other words, a participant with a relatively large overall contingency learning effect in mean probe RT also tended to exhibit a relatively large main effect of response key. We note, however, that this correlation did not achieve a conventional level of significance (i.e., $p <$ 0.05). Second, the degree to which the overall contingency learning effect in mean probe RT was greater after high- (vs low-) contingency trials correlated strongly with the degree to which the main effect of response key was greater after high- (vs low-) contingency trials (r = 0.58, $p <$ 0.001). In other words, a participant with a relatively large sequential contingency effect in mean probe RT also tended to exhibit a relatively large sequential contingency effect in mean pre-probe force. This finding suggests the possibility that sequential contingency effects in advance response preparation contribute to sequential contingency effects in mean RT.

Exploratory Analysis #2

One may wonder whether the present sequential contingency effects index binding and retrieval of previous stimulus-response episodes from episodic memory (e.g., Giesen et al., 2020; Schmidt et al., 2020) or processes that adapt to more abstract violations of expectations (Schmidt et al., 2007). As an initial step toward distinguishing between these possibilities, we investigated whether the sequential contingency effect is larger when the previous-trial prime arrow repeats in the current trial (e.g., left → left) than when it alternates (e.g., left → right). As we explain next, the episodic retrieval account predicts this outcome while the expectation account does not.

The episodic retrieval account predicts this outcome because it posits that repeating the previous prime arrow and its response triggers the retrieval of the previous probe arrow and its response (see Moeller & Frings, 2019 for a similar explanation). After a high-contingency trial, such retrieval will activate the correct response in a high-contingency trial but the incorrect response in a low-contingency trial, magnifying the difference in performance between these trial types. After a low-contingency trial, however, such retrieval will activate the incorrect response in a high-contingency trial but the correct response in a low-contingency trial, reducing the difference in performance between these trial types. Therefore, the sequential contingency effect should be relatively large when the prime arrow repeats from one trial to the next. In contrast, this sequential effect should be relatively small when the prime arrow alternates from one trial to the next. The idea here is that presenting a different prime arrow reduces the retrieval of information from the previous trial. For instance, if the current prime points to the left but the previous prime points to the right, then the current prime does not match the previous prime and will not strongly trigger its retrieval. This mismatch should reduce the modulatory influence of episodic retrieval on contingency learning effects that we described earlier. Thus, the sequential contingency effect should be smaller when the prime arrow alternates than when it repeats.

In contrast, the expectation account predicts equivalent sequential contingency effects regardless of whether the previous-trial prime arrow repeats or alternates in the current trial. In this account, sequential contingency effects reflect adjustments of control that are based on the prime arrow's usefulness to performance in the previous trial (Schmidt et al., 2007), rather than on the retrieval of stimulus- and/or response-specific information. After a high-contingency trial, wherein the prime predicts the correct probe response, participants are likely to use the prime to predict the probe response again. After a low-contingency trial, wherein the prime predicts the incorrect probe response, participants are unlikely to use the prime to predict the probe response again. Here, sequential contingency effects resemble congruency sequence effects in Stroop-like tasks, which, in some views (Gratton et al., 1992; Weissman et al., 2015), reflect adjustments of control based on the perceived utility of the distractor for predicting the target in the previous trial (i.e., perceived distractor utility is higher after congruent trials than after incongruent trials).

To distinguish between the episodic retrieval and expectation accounts, we examined whether the sequential contingency effects that we observed in mean RT and mean ER varied with whether the previous-trial prime arrow repeated or alternated in the current trial. That is, we investigated whether there was a three-way interaction among prime arrow type (repetition, alternation), previous trial contingency (high, low) and current trial contingency (high, low). We did not observe such a three-way interaction for mean RT, $F(1,40) < 1$, despite the presence of a robust interaction between previous trial contingency and current trial contingency (i.e., a robust 32 ms sequential contingency effect), $F(1,40) = 22.83$, $p < 0.001$, $\eta_p^2 = 0.36$. More specifically, sequential contingency effects did not vary with whether the previous-trial prime arrow repeated (32 ms) or alternated (31 ms) in the current trial. We also did not observe such an interaction for mean ER, $F(1,40) < 1$, although here the interaction between previous trial contingency and current trial contingency was not significant, $F(1,40) = 3.67$, $p = 0.063$, $\eta_p^2 = 0.084$, suggesting weaker sequential contingency effects in mean ER than in mean RT. These exploratory findings appear more consistent with the expectation account of the sequential contingency effect than with the episodic retrieval account. A firm conclusion, however, must await future studies that better control for stimulus and response repetitions in consecutive trials.

One may also wonder whether binding and retrieval of previous stimulus-response episodes from episodic memory influences overall contingency learning effects (e.g., Giesen et al., 2020; Schmidt et al., 2020). Any such influence should lead to larger contingency learning effects when the previous-trial prime arrow repeats in the current trial than when it does not repeat for two reasons. First, repeating the prime arrow should trigger the retrieval of the previous-trial probe response. Second, since most of the trials are high-contingency trials, both the retrieved response and the correct response to the upcoming probe will usually be the high-contingency response. For these reasons, retrieving the probe response from the previous trial should usually facilitate performance in high-contingency trials but usually impair performance in low-contingency trials, thereby increasing the magnitude of contingency learning effects.

To investigate this possibility, we determined whether there was a two-way interaction between prime arrow type (repetition, alternation) and current trial contingency (high, low). The interaction was not significant in either the mean RT data, $F(1,40) = 3.83$, $p = 0.057$, $\eta_p^2 = 0.09$, or the mean ER data, $F(1,40) < 1$. Further, we note that although there was a numerical trend toward an interaction in the mean RT data, contingency learning effects were numerically *smaller* – not larger – when the previous-trial prime arrow repeated in the current trial (11 ms) than when it changed (21 ms). This trend is exactly the opposite of what the episodic retrieval view predicts. Therefore, it provides no evidence that binding and retrieval of stimulus-response episodes from the previous trial increase overall contingency learning effects in the present task.

Discussion

In Experiment 2, we replicated the main findings from Experiment 1 while overcoming the limitations of that earlier experiment. First, we observed greater force on the high- vs low-contingency key 0-200 ms *before* probe onset. This shows once again that contingency learning is associated with proactive preparation of the high-contingency response prior to probe onset, an effect that the threshold only account does not predict. Second, we observed robust sequential contingency effects: behavioral and pre-probe force measures of contingency learning were greater after high- vs low-contingency trials. These effects further support the response preparation account and show that the previous trial's contingency (high vs low) exerts a strong influence on contingency learning over and above any influence of block-wide contingencies.

Interestingly, while behavioral measures of contingency learning were no longer significant after low-contingency trials, pre-probe force measures significantly reversed. This reversal may indicate that the cognitive system expects the current trial's contingency (e.g., low) to match the previous trial's contingency (e.g., low) and prepares accordingly. More broadly, this dissociation indicates that pre-probe force is not a perfect predictor of mean probe RT and mean probe ER. This makes intuitive sense because behavioral measures such as RT and ER index processes over and above proactive response preparation, which can occur only *after* probe onset (e.g., perception, decision-making, etc.). Consistent with this view, although the sequential contingency effect in mean probe RT correlated with the sequential contingency effect in mean pre-probe force across participants ($r = 0.58$), this correlation was far from perfect (i.e., 1).

Finally, we conducted exploratory analyses to gain insight into whether sequential contingency effects index (a) episodic retrieval of a stimulus-response binding from the previous trial (Giesen et al., 2019; Schmidt et al., 2020) or (b) stimulus-independent processes that adapt to violations of expectations (Schmidt et al., 2007). These analyses revealed that sequential contingency effects do not vary with whether the prime arrow in the current trial matches or mismatches the prime arrow in the previous trial (and further analyses yielded an analogous result for overall contingency learning effects). For the reasons described earlier, this outcome appears more consistent with the expectation account than with the episodic retrieval account.

Exploratory Across-Experiment Analyses

Prior findings indicate that the overall effects of contingency learning on performance increase as the proportions of high and low contingency trials become more unequal (Forrin & MacLeod, 2018; Schmidt & De Houwer, 2016a). This outcome suggests that effects of contingency learning might be greater in Experiment 1 (high contingency, 87.5%; low contingency, 12.5%) than in Experiment 2 (high contingency, 81.25%; low contingency, 18.75%). To investigate this possibility, we conducted exploratory across-experiment analyses.

These analyses yielded different patterns for overall vs sequential effects of contingency learning. Consistent with prior work, overall effects of contingency learning were significantly greater in Experiment 1 than in Experiment 2 (mean RT: $F(1,71) = 8.25$, $p = 0.005$, $\eta_p^2 = 0.10$; mean ER: $F(1,71) = 9.50$, $p = 0.003$, $\eta_p^2 = 0.12$; mean pre-probe force: $F(1,71) = 5.91$, $p = 0.018$, $\eta_p^2 = 0.08$). Sequential effects of contingency learning, on the other hand, did not differ between Experiment 1 and Experiment 2 (mean RT: $F(1,71) = 1.82$, $p = 0.18$, $\eta_p^2 = 0.025$; mean ER: $F(1,71) = 3.27$, $p = 0.075$, $\eta_p^2 = 0.042$; mean pre-probe force: $F(1,71) < 1$).

General Discussion

We sought to distinguish between the response preparation and threshold only accounts of contingency learning in a prime-probe task. Consistent with prior work, the behavioral data revealed faster and more accurate identification of the probe in high-contingency trials than in low-contingency trials. Critically, just before the probe appeared, we observed greater force on the high-contingency key corresponding to the direction in which the upcoming probe was more likely to point (e.g., Up) than on the low-contingency key corresponding to the direction in which the upcoming probe was less likely to point (e.g., Down). Furthermore, analogous to the behavioral effects, this pre-probe force effect (a) vanished after a single low-contingency trial and (b) increased as the proportions of high and low contingency trials became more unequal. These findings favor the response preparation account over the threshold only account.

Implications for the response preparation account

The present findings provide direct evidence for the response preparation account of contingency learning. In particular, they reveal greater pre-probe force on the key subjects usually press (vs do *not* press) to indicate the most likely identity of an upcoming probe. They also reveal sequential effects in pre-probe force that correspond well to sequential contingency effects in behavior. These data bolster the view that contingency learning involves proactively preparing the probe response that is most frequently associated with the prime (e.g., Luo et al., 2022). They also extend prior findings from studies of mouse tracking (Bruhn, 2013; Bruhn et al., 2014). Specifically, unlike these prior findings, the present data localize the pre-probe effects of interest to a specific effector (e.g., the right index finger), rather than merely to a specific hand (e.g., the right hand). This heightened specificity provides stronger evidence that the response preparation account extends to keyboard-based tasks than prior mouse tracking studies provide.

Also unlike data from mouse tracking studies (Bruhn, 2013; Bruhn et al., 2014), our data show that pre-probe effects related to contingency learning occur even when other processes do not bias participants toward a specific probe response. Indeed, the pre-probe force effects in our tasks occur (a) with primes whose perceptual characteristics are unlikely to suggest a specific probe response and (b) without informing participants about the predictive relationships between primes and probes. These task conditions are similar to those under which researchers typically measure contingency learning (see MacLeod, 2019 and Schmidt, 2019 for recent reviews). In combination with our use of keyboards to measure responses, this similarity increases the probability that our findings reflect the same contingency learning processes that influence performance in standard tasks.

Future work could assess whether the contribution of response preparation to contingency learning varies across different task conditions. In line with this possibility, prior work suggests a larger role for response preparation in contingency learning when the prime appears before (vs simultaneously with) the probe. Specifically, contingency learning effects in the color-word contingency task are larger when the prime appears before (vs with) the probe, a manipulation that should provide additional time to prepare the response most frequently associated with the prime (Schmidt & De Houwer, 2016b). Building on the present findings, one could investigate whether pre-probe force effects are smaller when the prime is task-irrelevant (vs task-relevant) and requires no response as is the case in many studies (Arunkumar et al., 2022; Forrin & MacLeod, 2018; Lin & MacLeod, 2018; Schmidt & Besner, 2008). Such an outcome would show that the prime's task-relevance (or lack thereof) influences the degree to which response preparation contributes to contingency learning. One could also investigate whether pre-probe force effects are smaller when participants cannot predict exactly which probe stimulus will appear in a high-contingency trial. This may occur if each of several equally probable stimuli is associated equally often with the same high-contingency response as in a task wherein all odd-numbered digits are mapped to one response key and all even-numbered digits are mapped to another response key (Arunkumar et al., 2022). Such an outcome would suggest that the inability to predict exactly which probe stimulus will appear in a high-contingency trial reduces the degree to which response preparation contributes to contingency learning. That is, such an outcome would suggest that contingency learning in the present tasks involves learning to predict the most likely probe *stimulus* after a given prime, not only the most likely probe *response*.

Implications for the drift rate and threshold only accounts

The changes in pre-probe force that we have observed do not appear consistent with the drift rate and threshold only accounts of contingency learning. Changes to the drift rate or the response threshold influence processes that identify or respond to a stimulus, respectively. Since each of these processes can operate only *after* the probe appears, however, they cannot easily explain the influence of contingency learning on *pre-probe* force that we have observed.

Nonetheless, changes to the drift rate and/or response threshold may influence behavioral indices of contingency learning in the present tasks. As we stated earlier, prior findings suggest that these processes often exert an influence on contingency learning effects (Dunovan et al., 2014; Dunovan & Wheeler, 2018; Forrin & MacLeod, 2018; Lin & MacLeod, 2018; Mulder et al., 2012, 2014; Schmidt & Besner, 2008). These processes may, therefore, influence performance *after* the probe appears and thereby contribute to contingency learning effects in mean probe RT and/or mean probe ER. The pre-probe force effects that we have observed simply indicate that these processes cannot *fully* explain contingency learning in our tasks.

Broader implications

The present findings highlight the tremendous value of using force-sensitive keys to investigate response preparation. Prior researchers have used such keys to investigate a variety of processes including those related to response inhibition (Ko et al., 2012), dual-task performance (Miller & Alderton, 2006), stimulus-response compatibility (Mattes et al., 2002), and coping with distraction (Miller & Rouast, 2016; Weissman, 2019). To our knowledge, however, the present study is the first to focus on response preparation in the context of contingency learning. More broadly, we note that force-sensitive keys may allow future researchers to investigate how proactive response preparation contributes to other processes such as priming, decision-making, and motor control. Indeed, such keys can measure multiple finger-specific response activations on the same hand with high temporal resolution (e.g., 500 Hz) and a high signal-to-noise ratio.

The present findings also add to a growing body of work indicating that factors other than the relative frequencies of high- and low-contingency trials influence contingency learning. In particular, they replicate prior findings indicating that contingency learning effects are smaller after low-contingency trials than after high-contingency trials (Schmidt et al., 2007). This sequential contingency effect, which is rarely the subject of experimental inquiry, complements recent data suggesting that trial-by-trial episodic retrieval of previous stimulus-response episodes influences contingency learning in some tasks (Giesen et al., 2019; Schmidt et al., 2020). It also complements recent data suggesting that conscious awareness of contingencies can lead participants to infer block-wide rules for performing a task (Arunkumar et al., 2022).

Our findings, though, go beyond bolstering the view that multiple processes influence contingency learning by providing insights into the nature of sequential contingency effects. Indeed, the present findings show that sequential contingency effects in behavioral measures (e.g., mean probe RT) do not vary with whether the prime repeats or changes in consecutive trials. This outcome appears more consistent with a process that adapts to perceived changes in the utility of the prime for predicting the probe (Schmidt et al., 2007), which may also underlie congruency sequence effects (Gratton et al., 1992; Weissman et al., 2015), than with a process that retrieves previous stimulus-response bindings from memory when the prime repeats in consecutive trials (Giesen et al., 2019; Schmidt et al., 2020). Since these findings come only from exploratory analyses, however, reaching a firm conclusion must await future studies.

More broadly, it is interesting to consider whether the changes in pre-probe force that we have observed stem from a buildup of prime-triggered, contingency-related, motor cortex activity that shifts the starting point (Corbett et al., 2023; Kelly et al., 2021)[7]. These changes in force may reflect, for example, the influence of such activity on effector-specific muscle activity. Future studies that integrate the present approach with neural measures could test this hypothesis. Confirming this hypothesis would provide a mechanistic explanation of our findings at the neural level and reveal a novel link between shifts in starting point and changes in pre-probe force.

Finally, it is interesting to consider the relationship between our findings and data showing that statistical learning influences the deployment of visual spatial attention (Huang et

[7] The authors thank Alexander Weigard at the University of Michigan for bringing this possibility to their attention.

al., 2022). Analogous to the present contingency learning effects, participants are faster to identify – via a keypress – a target that appears at a high-probability location than to identify a target that appears at a low-probability location. They also proactively allocate spatial attention to the high-probability location before target onset. The present findings suggest that such effects could index the proactive preparation of an eye movement – a type of response – to the high-probability location. That is, participants may learn that after an initial placeholder display, which appears before each trial, the target is more likely to appear at one location than at any other location. After this contingency is learned, the placeholder display may serve as a prime that triggers the preparation of an eye movement to the high-probability location. Consistent with this hypothesis, statistical learning exerts similar influences on visual spatial attention when eye movements – rather than keypresses – are required to identify a target (Godijn & Theeuwes, 2003). Future studies could investigate whether overlapping response preparation processes enable the expression of contingency learning (or statistical learning) in different tasks.

Limitations

The present study has three limitations that do not detract from our main conclusions. First, our findings do not reveal whether conscious awareness of prime-probe contingencies increases the size of pre-probe force effects (Arunkumar et al., 2022). Second, our findings do not reveal whether the present pre-probe force effects generalize to tasks wherein the prime is task-irrelevant (Forrin & MacLeod, 2018; Lin & MacLeod, 2018; Schmidt & De Houwer, 2016a). Future studies could investigate these possibilities by assessing conscious awareness of contingencies or pre-probe force in tasks wherein the prime is task-irrelevant, respectively. Third, our findings do not establish associations between pre-probe force and (a) activity in the motor cortex or (b) the starting point in sequential sampling models. Future studies could investigate such associations by integrating the present approach with neural measures and/or drift-diffusion modeling. Regardless of the results, the present findings indicate that response preparation contributes to contingency learning in the present prime-probe tasks.

Finally, we acknowledge that stimulus repetitions, which can influence contingency learning effects (e.g., Schmidt et al., 2020), are more frequent in the present two-alternative-forced-choice (i.e., 2-AFC) tasks than in 3-AFC and 4-AFC tasks. A stimulus repetition may increase contingency learning effects by triggering the retrieval of a response with which the stimulus was recently associated (Giesen et al., 2020; Schmidt et al., 2020). In the present tasks, for example, repeating the previous-trial prime (or the associated response) may trigger the retrieval of the response given to the previous-trial probe. Since most of the trials are high-contingency trials, both the retrieved response and the correct response to the upcoming probe will usually be the high-contingency response. Retrieving the previous-trial probe response should, therefore, usually facilitate performance in high-contingency trials but usually impair performance in low-contingency trials, thereby increasing the magnitude of contingency learning effects. We do not view repetition-related processes as confounds because they are consistent with models wherein episodic retrieval contributes to contingency learning (Schmidt et al., 2016). We also note, however, that the exploratory analyses in Experiment 2 suggest that there is more to contingency learning in the present tasks than retrieving an S-R binding from the previous trial. Indeed, neither the main effect of current trial contingency (i.e., overall contingency learning effects) nor the interaction between previous trial contingency and current trial contingency (i.e., sequential contingency learning effects) was larger when the prime arrow from the previous trial repeated (vs changed) in the current trial. For this reason, it appears unlikely that stimulus repetitions can fully explain the effects we have observed.

Conclusion

The present findings favor a response preparation account of contingency learning over a threshold only account. Future studies investigating the effects we have observed may provide additional novel insights into how the human mind capitalizes on statistical regularities in the environment to facilitate quick, accurate responses.

Compliance with Ethical Standards

Competing Interests

The authors have no competing financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Ethics Approval

The University of Michigan's Behavioral Sciences Internal Review Board deemed that the present study was exempt from oversight because they only involved simple cognitive tasks. The experiments were performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helskini and its later amendments or comparable ethical standards.

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Declarations

Funding

This work was funded by discretionary funds to Daniel H. Weissman from the University of Michigan.

Data Availability Statement

The preregistration, task scripts, data analysis scripts, and raw data are available on the Open Science Framework (OSF): Experiment 1 (https://osf.io/jd9ys/) and Experiment 2 (https://osf.io/eyzux/).

References

Arunkumar, M., Rothermund, K., Kunde, W., & Giesen, C. G. (2022). Being in the Know: The Role of Awareness and Retrieval of Transient Stimulus-Response Bindings in Selective Contingency Learning. *Journal of Cognition*, *5*(1), 36. https://doi.org/10.5334/joc.227

Bruhn, P. (2013). Emergence of spontaneous anticipatory hand movements in a probabilistic environment. *Advances in Cognitive Psychology*, *9*(2), 62–73. https://doi.org/10.2478/v10053-008-0132-y

Bruhn, P., Huette, S., & Spivey, M. (2014). Degree of certainty modulates anticipatory processes in real time. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(2), 525–538. https://doi.org/10.1037/a0034365

Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 567–587. https://doi.org/10.1037/a0035032

Corbett, E. A., Martinez-Rodriguez, L. A., Judd, C., O'Connell, R. G., & Kelly, S. P. (2023). Multiphasic value biases in fast-paced decisions. *eLife*, *12*, e67711. https://doi.org/10.7554/eLife.67711

Cox, J. A., & Aimola Davies, A. M. (2022). Age differences in visual statistical learning: Investigating the effects of selective attention and stimulus category. *Psychology and Aging*, *37*(6), 698–714. https://doi.org/10.1037/pag0000697

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *The Journal of Neuroscience*, *33*(4), 1400–1410. https://doi.org/10.1523/JNEUROSCI.1094-12.2013

Dunovan, K. E., Tremel, J. J., & Wheeler, M. E. (2014). Prior probability and feature predictability interactively bias perceptual decisions. *Neuropsychologia*, *61*, 210–221. https://doi.org/10.1016/j.neuropsychologia.2014.06.024

Dunovan, K., & Wheeler, M. E. (2018). Computational and neural signatures of pre and post-sensory expectation bias in inferior temporal cortex. *Scientific Reports*, *8*(1), 13256. https://doi.org/10.1038/s41598-018-31678-x

Eimer, M., Goschke, T., Schlaghecken, F., & Stürmer, B. (1996). Explicit and implicit learning of event sequences: Evidence from event-related brain potentials. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(4), 970–987. https://doi.org/10.1037//0278-7393.22.4.970

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037//0278-7393.28.3.458

Forrin, N. D., & MacLeod, C. M. (2018). Contingency proportion systematically influences contingency learning. *Attention, Perception & Psychophysics*, *80*(1), 155–165. https://doi.org/10.3758/s13414-017-1424-4

Giesen, C. G., Schmidt, J. R., & Rothermund, K. (2019). The Law of Recency: An Episodic Stimulus-Response Retrieval Account of Habit Acquisition. *Frontiers in Psychology, 10,* 2927. https://doi.org/10.3389/fpsyg.2019.02927

Godijn, R., & Theeuwes, J. (2003). Parallel allocation of attention prior to the execution of saccade sequences. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(5), 882–896. https://doi.org/10.1037/0096-1523.29.5.882

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation and responses. *Journal of Experimental Psychology. General*, *4*, 480–506.

Huang, C., Donk, M., & Theeuwes, J. (2022). Proactive enhancement and suppression elicited by statistical regularities in visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *48*(5), 443–457. https://doi.org/10.1037/xhp0001002

JASP Team. (2023). *JASP (Version 0.18.0.0)* [Computer software]. https://jasp-stats.org/.

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kelly, S. P., Corbett, E. A., & O'Connell, R. G. (2021). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*, *5*(4), 467–481. https://doi.org/10.1038/s41562-020-00967-9

Ko, Y. T., Alsford, T., & Miller, J. (2012). Inhibitory effects on response force in the stop-signal paradigm. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(2), 465–477. https://doi.org/10.1037/a0027034

Lin, O. Y.-H., & MacLeod, C. M. (2018). The acquisition of simple associations as observed in color-word contingency learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(1), 99–106. https://doi.org/10.1037/xlm0000436

Logan, G. D., & Zbrodoff, J. N. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*(3), 166–174.

Luo, J., Yang, M., & Wang, L. (2022). Learned irrelevant stimulus-response associations and proportion congruency effect: A diffusion model account. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001158

MacLeod, C. M. (2019). Learning simple associations. *Canadian Psychology*, *60*(1), 3.

Mattes, S., Leuthold, H., & Ulrich, R. (2002). Stimulus-response compatibility in intensity-force relations. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *55*(4), 1175–1191. https://doi.org/10.1080/02724980244000152

Miller, J., & Alderton, M. (2006). Backward response-level crosstalk in the psychological refractory period paradigm. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(1), 149–165. https://doi.org/10.1037/0096-1523.32.1.149

Miller, J., & Rouast, N. M. (2016). Dissociations of spatial congruence effects across response measures: An examination of delta plots. *Psychological Research*, *80*(5), 805–820. https://doi.org/10.1007/s00426-015-0694-5

Moeller, B., & Frings, C. (2019). From simple to complex actions: Response-response bindings as a new approach to action sequences. *Journal of Experimental Psychology: General*, *148*(1), 174–183. https://doi.org/10.1037/xge0000483

Mulder, M. J., Boekel, W., Ratcliff, R., & Forstmann, B. U. (2014). Cortico-subthalamic connection predicts individual differences in value-driven choice bias. *Brain Structure & Function*, *219*(4), 1239–1249. https://doi.org/10.1007/s00429-013-0561-3

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience*, *32*(7), 2335–2343. https://doi.org/10.1523/JNEUROSCI.4156-11.2012

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Hochenberger, R., Sogo, H., Kastman, E., & Lindelov, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Rousseeuw, P. J., & Crouz, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schmidt, J. R. (2013). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-012-0373-0

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review*, *26*(3), 753–771. https://doi.org/10.3758/s13423-018-1520-z

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(3), 514–523. https://doi.org/10.1037/0278-7393.34.3.514

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*(2), 421–435. https://doi.org/10.1016/j.concog.2006.06.010

Schmidt, J. R., & De Houwer, J. (2016a). Contingency Learning Tracks With Stimulus-Response Proportion. *Experimental Psychology*, *63*(2), 79–88. https://doi.org/10.1027/1618-3169/a000313

Schmidt, J. R., & De Houwer, J. (2016b). Time course of colour-word contingency learning: Practice curves, pre-exposure benefits, unlearning, and relearning. *Learning and Motivation*, *56*, 15–30. https://doi.org/10.1016/j.lmot.2016.09.002

Schmidt, J. R., & De Houwer, J. (2019). Cue competition and incidental learning: No blocking or overshadowing in the colour-word contingency learning procedure without instructions to learn. *Collabra: Psychology*, *5*(1). https://doi.org/10.1525/collabra.236

Schmidt, J. R., De Houwer, J., & Rothermund, K. (2016). The Parallel Episodic Processing (PEP) model 2.0: A single computational model of stimulus-response binding, contingency learning, power curves, and mixing costs. *Cognitive Psychology*, *91*, 82–108. https://doi.org/10.1016/j.cogpsych.2016.10.004

Schmidt, J. R., Giesen, C. G., & Rothermund, K. (2020). Contingency learning as binding? Testing an exemplar view of the colour-word contingency learning effect. *Quarterly Journal of Experimental Psychology (2006)*, *73*(5), 739–761. https://doi.org/10.1177/1747021820906397

Schmidt, J. R., Houwer, J. D., & Besner, D. (2010). Contingency learning and unlearning in the blink of an eye: A resource dependent process. *Consciousness and Cognition*, *19*(1), 235–250. https://doi.org/10.1016/j.concog.2009.12.016

Spinelli, G., & Lupker, S. J. (2023). Robust evidence for proactive conflict adaptation in the proportion-congruent paradigm. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *49*(5), 675–700. https://doi.org/10.1037/xlm0001144

Urai, A. E., & Donner, T. H. (2022). Persistent activity in human parietal cortex mediates perceptual choice repetition bias. *Nature Communications*, *13*(1), 6015. https://doi.org/10.1038/s41467-022-33237-5

Weissman, D. H. (2019). Let your fingers do the walking: Finger force distinguishes competing accounts of the congruency sequence effect. *Psychonomic Bulletin & Review*, *26*(5), 1619–1626.

Weissman, D. H., Egner, T., Hawks, Z., & Link, J. (2015). The congruency sequence effect emerges when the distracter precedes the target. *Acta Psychologica*, *156*, 8–21. https://doi.org/10.1016/j.actpsy.2015.01.003

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology*. *Learning, Memory, and Cognition*, *40*(2), 385–398. https://doi.org/10.1037/a0034851