Learning and Motivation 56 (2016) 15–30

Contents lists available at ScienceDirect

Learning and Motivation

journal homepage: www.elsevier.com/locate/l&m

# Time course of colour-word contingency learning: Practice curves, pre-exposure benefits, unlearning, and relearning

James R. Schmidt*, Jan De Houwer

*Department of Experimental Clinical and Health Psychology, Ghent University, Belgium*

ARTICLE INFO

*Article history:*
Received 12 July 2016
Received in revised form 5 September 2016
Accepted 6 September 2016
Available online 15 September 2016

*Keywords:*
Contingency learning
Acquisition
Practice
Power law
Unlearning
Relearning

ABSTRACT

In performance-based measures of implicit contingency learning, learning effects have been observed very early in the task (e.g., within a few trials) and remain stable throughout the experiment. This has been taken to suggest that the contingency knowledge underlying the learning effects is formed almost instantly and does not develop further across trials. One potential concern with the available evidence is that response times are overall much slower early on in an experiment and speed up over practice in a decelerating function. If learning effects scale with overall response time, then learning effects observed early on in an experiment might be artificially inflated. In the current report with the colour-word contingency learning paradigm, participants were given an extended practice phase before introducing predictive stimuli (words). Thus, learning could be assessed after the large practice speedup in performance had already occurred. In one experiment, the contingency learning effect was found to again be fairly stable, but with a hint of an increasing effect with time. In a second experiment, words were pre-exposed in a neutral hue before being coloured. This increased the magnitude of the learning effect, suggesting a preparation time benefit. More importantly, the contingency learning effect was observed to increase over time. In a third experiment, we assessed unlearning rates when the contingency was removed, and relearning when the contingencies were reintroduced. The results revealed a cumulative effect of contingencies acquired across multiple blocks. In sum, the evidence reported in this paper shows that, contrary to previous claims, implicit contingency learning is cumulative.



## 1. Introduction

In the study of implicit contingency learning, performance (i.e., response time) paradigms are very useful for assessing learning. For instance, in the colour-word contingency learning paradigm, participants respond to the print colour of neutral (colour-unrelated) words (Atalay & Misirlisoy, 2012; Hazeltine & Mordkoff, 2014; Levin & Tzelgov, 2016; Schmidt & Besner, 2008; Schmidt & De Houwer, 2012a, 2012d, 2016; Schmidt, Crump, Cheesman, & Besner, 2007; Schmidt, De Houwer, & Besner, 2010; see also, Musen & Squire, 1993). Each word is presented most often in one colour (e.g., "choose" most often in purple, "drive" most often in orange, etc.). Learning can be assessed by comparing *high contingency* trials, where the word is presented in the expected colour (e.g., "choose" in purple), to *low contingency* trials, where the word is presented in an unexpected colour (e.g., "choose" in orange). This produces a highly-robust learning effect: high contingency trials

* Corresponding author at: Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium.
*E-mail address:* james.schmidt@ugent.be (J.R. Schmidt).

http://dx.doi.org/10.1016/j.lmot.2016.09.002

are responded to faster (and more accurately) than low contingency trials. A similar paradigm is the flanker contingency paradigm, in which flanking letters are predictive of a centrally-presented target letter (Carlson & Flowers, 1996; Miller, 1987). Other performance paradigms include the serial response time task (Nissen & Bullemer, 1987) and hidden covariation detection (Lewicki, 1985, 1986; Lewicki, Hill, & Czyzewska, 1992).

One interesting finding with such performance paradigms is that the learning effect appears *very* early on in the experiment (i.e., within a few trials). In the most dramatic instance of this, a learning effect was observed after a single presentation of a stimulus by Lewicki (1985) in the hidden covariation paradigm. Similarly, contingency effects have been observed early on in sequence learning (Nissen & Bullemer, 1987). Also in the colour-word contingency learning paradigm, contingency effects emerge in the very first block of trials, with blocks as small as 18 trials. After this, the magnitude of the learning effect remains relatively stable (e.g., Schmidt & De Houwer, 2012b, 2012d, 2016; Schmidt et al., 2007, 2010). That is, the learning effect does not seem to increase in any notable way. Results such as this have been taken to indicate that the learning rate is extremely high. In other words, the contingencies are learned very rapidly and there is little more to learn thereafter.

First, it is important to note the distinction between the *learning effect* and the *underlying learning mechanism*. The learning effect (e.g., the difference in response speed to high vs. low contingency trials) is a behavioural observation. Of course, some underlying learning mechanism (i.e., acquisition of contingency knowledge) must be assumed to explain the learning effect. However, the learning effect is not a pure measure of the underlying knowledge. In addition to general error in estimates of high contingency RT, low contingency RT, and thus the difference between the two (Kaufman et al., 2010), the size of the contingency effect is also partially determined by the *expression of the underlying knowledge*. That is, the observed contingency effect is not a pure measure of how much is *known* about the contingency, but is also determined by how effectively this contingency knowledge is being retrieved from memory and by the processes via which it influences performance. In the present report, we assess the potential role of two other factors on the magnitude of the learning effect, which may also have major implications for inferences about the learning rate in performance paradigms: practice and stimulus pre-exposure.

It is well known that overall performance in any response time experiment improves with practice. Indeed, this occurs in a consistent enough manner that it is often described as a law of behaviour (e.g., Logan, 1988; Newell & Rosenbloom, 1981). Specifically, response times at the beginning of an experiment tend to be very slow. As the experiment progresses, response times rapidly improve early on. The improvements continue throughout the experiment, but at an ever-diminishing rate. That is, response times decrease in a decelerating function. In blocked analyses, this practice improvement can be represented with a power function: $RT = a + bN^{-c}$. In this formula, $a$ is the minimum RT that performance improves toward, $b$ is the difference between $a$ and Trial 1 performance, $N$ is the trial number, and $c$ is the learning rate (normally $\geq 0$). In more refined, trial-by-trial analyses on the data of an individual participant an exponential function is more appropriate (Heathcote, Brown, & Mewhort, 2000; Myung, Kim, & Pitt, 2000). In either case, performance approaches a theoretical asymptote over trials, with larger absolute changes in the earlier relative to later trials. The ever-decreasing rate of improvement with further practice is easily explained by the fact that the closer the current response speed is to asymptote (i.e., the fastest responding physically possible) the less room there is for further improvements. In an extreme example, if response time started out at 1400 ms per trial and has already improved to 400 ms per trial (1000 ms speedup), it is obviously impossible to improve another 1000 ms faster (i.e., to −600 ms) no matter how much one practices.

As we will shortly describe, these practice benefits might have implications for assessments of the acquisition of contingency knowledge. This is because response times do not merely decrease with practice; response time *effects* also seem, at least in some notable cases, to decrease with practice. For instance, in Stroop experiments the congruency effect is observed to decrease with practice (Dulaney & Rogers, 1994; Ellis & Dulaney, 1991; MacLeod, 1998; Simon, Craft, & Webster, 1973). This decrease in the congruency effect over time can be due to scaling with mean response time (Schmidt, 2016). That is, as mean response time decreases over practice, the congruency effect shrinks with it. Stated a different way, incongruent trials start out much slower than congruent trials, so they will gain more from practice. In yet other words, participants will get increasingly better at identifying the colour and executing the appropriate response over practice, giving the word less and less time to interfere.

Given these considerations, scaling of effects with practice can also be a concern for performance-based contingency effects. That is, overall learning effects might be larger when overall responding is slower (Stevens et al., 2002; Urry, Burns, & Baetu, 2015). In the initial blocks of learning, the contingency effect might be inflated simply because overall responding is slower early in the experiment. Thus, even if the amount of contingency *knowledge* acquired is relatively minimal, the contingency *effect* might nevertheless appear large due to response time scaling. Indeed, if we assume that: (a) contingency effects *do* scale with overall RT and (b) learning does reach peak very early on, then we should actually expect *much larger* contingency effects in the first (slower) blocks than in later (faster) blocks. This is illustrated in Panel A of Fig. 1. In particular, the response time contingency effect would be very large early on (due to the overall slow response times), and as participants become faster and faster at responding to the colour with practice, the absolute difference between high and low contingency trials would diminish. This is unlike what we have observed in the past.

Two alternative possibilities are illustrated in Panels B and C of Fig. 1, both of which produce a seemingly flat acquisition slope. In Panel B, we see a situation where: (a) true acquisition of contingency knowledge is extremely rapid, but (b) the contingency effect does *not* scale with mean RT. In Panel C, we see the exactly opposite situation, where: (a) the true contingency knowledge is still developing early on, but (b) the contingency effect *does* scale with mean RT. As such, the absolute contingency effect in response times might nevertheless already be large in the initial blocks of learning given that overall response speed is slow. As practice progresses, the contingency effect both (a) increases due to better contingency

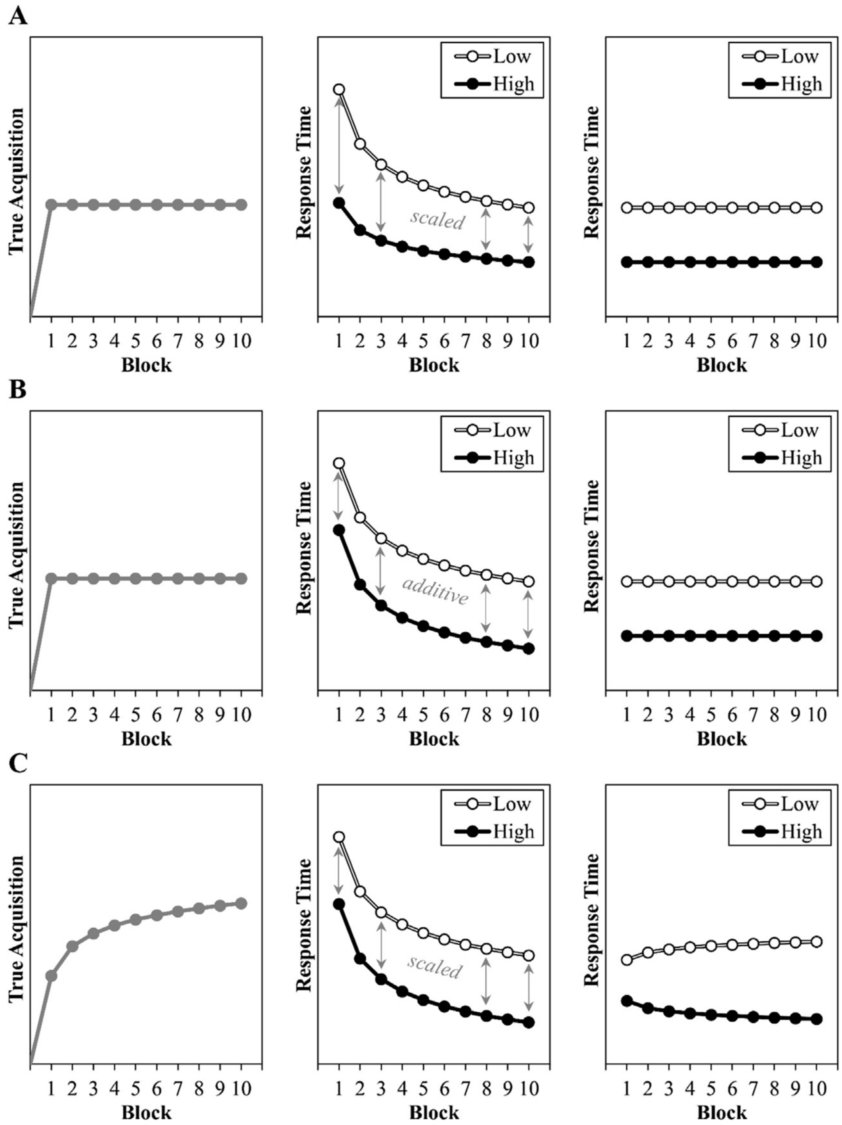


**Fig. 1.** Example true acquisition of contingency knowledge (left panels), practice speedups for high and low contingency trials (center panels), and predicted learning curves without practice speedups (right panels). A: near immediate acquisition that scales with mean RT. B: near immediate acquisition, but no scaling. C: gradual acquisition that scales with mean RT. Both B and C produce seemingly flat acquisition curves (center panels), but with very different underlying learning rates, leading to different predictions when practice speedups are eliminated (right panels).

knowledge, and (b) decreases due to scaling down with faster responses. The net effect is a seemingly-flat acquisition curve. That is, practice-based scaling of RT effects works against the ability of the experimenter to reveal any true effect of increasing contingency knowledge, especially if practice-based scaling is of a roughly comparable magnitude to the true growth in contingency knowledge. Given these considerations, it is possible that underlying contingency knowledge does accumulate over time, but this has been partially masked by practice-based scaling in prior reports. The only way to distinguish between the situation in Panel B and Panel C is to attempt to study contingency acquisition in an experiment with relatively stable mean block RTs. In this case, the two accounts make very different predictions, as indicating in the rightmost panels of Fig. 1.

Of course, it is never possible to separate practice from the acquisition of contingency knowledge entirely, because acquisition must be studied across time. That said, in the experiments to follow we control for practice partially. In particular, participants first practiced responding to a coloured control stimulus ("@@@@") for an extended period of time. This was to practice colour-to-key mappings prior to contingency learning. Predictive stimuli, presented in the same colours, were introduced only after this practice. Assuming that most of the practice benefit occurs relatively early in an experiment, little more of a practice benefit will be observed during the contingency learning phase (i.e., overall RT will be relatively stable). It is possible that under these conditions we will observe that the contingency effect increases over blocks. On the other hand, it might be the case that the contingency learning effect does not scale with response time. In other words, the contingency learning effect might be of similar magnitude both when overall responding is fast and slow. If this is the case, then we should again observe that the contingency learning effect does not increase across blocks.

In Experiment 1, we first aim to establish whether an increasing contingency effect across blocks is observable after an initial practice phase. In Experiment 2, we extend on these findings. In particular, another way in which to study the cumulative effect of acquired contingency knowledge is with *changes* in the contingency at different stages of the experiment. In particular, we will investigate how long the contingency effect persists during *unlearning*, where the contingency is removed (i.e., word switched from predictive to unpredictive of the colour). We will also study *relearning*, in order to investigate whether a prolonged unlearning phase slows acquisition of the contingency once it has been reintroduced.

## 2. Experiments 1a and 1b

Experiment 1 had three interrelated added aims. First and most importantly, the experiment aimed to see whether the contingency learning effect does or does not increase over time. Most critically and unlike past reports, we assessed acquisition *after* an initial practice phase, wherein participants were familiarized with the task and the colour-to-key mappings. This was to minimize the impact of scaling with mean RT on the contingency effect. In that vein, Experiment 1a was essentially a replication of a three-choice colour-word contingency learning paradigm, save for the extended practice phase. Second and relatedly, we determined to what extent mean RT correlates with the contingency effect, both in a between participant and within participant analysis. The latter tests can give us an indication to what extent the contingency effect scales with mean RT.

Third and more incidentally, we assess whether brief pre-exposure of the predictive word stimulus might boost contingency learning effects. We take our inspiration for this idea from paradigms that study the binding between distracting word stimuli and responses (see Frings, Rothermund, & Wentura, 2007; Rothermund, Wentura, & De Houwer, 2005). These paradigms also have participants respond to target stimuli while ignoring a distracting stimulus. Rather than manipulating the contingencies between words and colours in these paradigms (i.e., repeatedly-bound target-distracter pairs), these paradigms study the effect of *recently-encountered* target-distracter bindings on performance (for an extended discussion, see Schmidt & De Houwer, 2016). The distracting stimulus is typically briefly pre-exposed prior to target onset in these paradigms, and there is some indication that binding effects might be larger with this pre-exposure (Klaus Rothermund, personal communication). Because we propose that contingency learning and binding effects likely result from the same learning process (Schmidt & De Houwer, 2016), it seems possible that a larger and more robust contingency learning effect might be observable when the predictive word stimulus is pre-exposed (i.e., rather than being presented concurrently with the target colour). This could occur simply because the participant has more preparation time if the word comes earlier. In this vein, Experiment 1b was identical in all respects to Experiment 1a, except that the predictive word was presented first on the screen in a neutral (non-response set) colour before changing to one of the target colours 150 ms later. This manipulation is also relevant to our main question about acquisition. If the contingency learning effect is, indeed, more robust with pre-exposure, then we might have greater power to detect increases in the contingency learning effect over time (i.e., assuming that the effect does increase with time). That is, a larger contingency effect should produce an equally steeper learning curve.

### *2.1. Method*

#### *2.1.1. Participants*

Thirty-six Ghent University undergraduates participated in Experiment 1a and thirty-four participated in Experiment 1b in exchange for €5. No participants participated in both experiments.

#### *2.1.2. Apparatus*

Stimulus presentation and response timing were controlled by E-Prime 2 (Psychology Software Tools, Pittsburgh, PA). Participants responded with the “J” key for purple, “K” key for orange, and “L” key for grey with the first three fingers of their right hand.

#### *2.1.3. Design*

The display colours were purple (RGB: 128,0,128), orange (255,165,0), and grey (192,192,192), which are “purple,” “orange,” and “silver” in the standard E-Prime colour palate. The experiment began with 3 blocks of 30 *practice trials* (90 total). The length of this practice phase was determined by visually inspecting some of our old datasets to determine how many trials it took for overall mean RT to stabilize. The practice phase was then made slightly longer than this stabilization point. In each practice block, the stimulus “@@@@” was presented 10 times in each of the three colours. These frequencies exactly match the following learning blocks. After practice and a self-paced break, there were 10 blocks of 30 *learning trials* (300 total). Three four-letter, first-person Dutch verbs served as the predictive words (zoek [search], kies [choose], rijd [drive]). Each word was presented 80% (8 of 10 times per block) in one colour, and 10% (1 of 10) in each of the remaining two colours. One word was presented most often in purple, another most often in orange, and the third most often in grey.

Which word was presented most often in which colour was randomly determined on a participant-by-participant basis. In both parts of the experiment, trials were selected at random without replacement.

*2.1.4. Procedure*

All stimuli were presented in bold, 18 pt. Courier New font on a black screen. In Experiment 1a, each trial began with a white (255,255,255) fixation "+" for 150 ms, followed by a blank screen for 150 ms. The stimulus was then presented until either a response was made or 1500 ms had elapsed. After correct responses, the next trial immediately began. If participants responded incorrectly or failed to respond in 1500 ms, then "XXX" in white was presented for 1000 ms. The procedure of Experiment 1b was identical in all respects to Experiment 1a, with one exception. Instead of a 150 ms blank screen between the fixation and target stimulus the word was pre-exposed for 150 ms in a light brown (255,183,113). This colour was selected to both have high enough contrast with the black background and to be noticeably different from the three target colours (i.e., purple, orange, and grey).

*2.1.5. Data analysis*

Correct response time and error percentages were calculated. All participants had acceptable error rates (<20%), so no participants were excluded. For the scaling analyses, mean RT was computed as the experiment-wide (between subject analysis) or block-wide (within participant analysis) average response time, including all trials in which a response (correct or incorrect) was made. The contingency effect was computed as low minus high contingency mean RT for each participant, also experiment- or block-wide. For the between participant analysis, both parametric *r* and nonparametric Spearman's $\rho$ were calculated. For the within participant analysis, a linear mixed effect (LME) analysis was conducted using restricted maximum likelihood estimation and diagonal variance structure.

## 2.2. Results

*2.2.1. Response times*

First, the practice blocks were analysed with a block (1–3) ANOVA for each experiment. The linear contrast of block was significant in Experiment 1a, $F(1,35) = 35.997$, $MSE = 936$, $p < 0.001$, $\eta_p^2 = 0.51$, and in Experiment 1b, $F(1,33) = 13.957$, $MSE = 3603$, $p < 0.001$, $\eta_p^2 = 0.30$, indicating that performance improved with practice. More precisely, in Experiment 1a Block 1 performance (593 ms) was significantly slower than Block 2 (554 ms), $t(35) = 5.261$, $SE_{diff} = 7$, $p < 0.001$, $\eta_p^2 = 0.44$, and Block 3 (550 ms), $t(35) = 6.000$, $SE_{diff} = 7$, $p < 0.001$, $\eta_p^2 = 0.51$, but there was no difference between Blocks 2 and 3, $t(35) = 0.618$, $SE_{diff} = 7$, $p = 0.540$, $\eta_p^2 = 0.01$. Similarly in Experiment 1b, Block 1 performance (636 ms) was significantly slower than Block 2 (580 ms), $t(33) = 5.721$, $SE_{diff} = 10$, $p < 0.001$, $\eta_p^2 = 0.50$, and Block 3 (581 ms), $t(33) = 3.736$, $SE_{diff} = 15$, $p < 0.001$, $\eta_p^2 = 0.30$, but there was no difference between Blocks 2 and 3, $t(33) = 0.124$, $SE_{diff} = 11$, $p = 0.902$, $\eta_p^2 < 0.01$. Thus, there was clear evidence for practice benefits, but this was most noticeable in the shift from the first to second block.

Next, a contingency (high vs. low) by block (1–10) ANOVA was conducted on the learning blocks for each experiment. The data are presented in Fig. 2. Unsurprisingly, the main effect of contingency was significant in both Experiment 1a, $F(1,35) = 50.326$, $MSE = 6915$, $p < 0.001$, $\eta_p^2 = 0.59$, and Experiment 1b, $F(1,33) = 133.932$, $MSE = 5835$, $p < 0.001$, $\eta_p^2 = 0.80$, because responses were faster to high contingency trials. The linear main effect of block was not significant in Experiment 1a, $F(1,35) = 0.092$, $MSE = 8588$, $p = 0.764$, $\eta_p^2 < 0.01$, but was significant in Experiment 1b, $F(1,33) = 7.054$, $MSE = 7111$, $p = 0.012$, $\eta_p^2 = 0.18$. This main effect of block might seem to suggest that responses slowed over the experiment, but this is mostly illusory: low contingency trials contribute 50% to the main effect calculation, despite their low frequency. Overall RT (i.e., ignoring the distinction between conditions) did not slow significantly across blocks, $F(1,33) = 1.992$, $MSE = 2249$, $p = 0.167$, $\eta_p^2 = 0.06$.

Most critical is the interaction between contingency and block. This interaction was significant in Experiment 1b (with pre-exposure of the word), $F(1,33) = 7.685$, $MSE = 4888$, $p = 0.009$, $\eta_p^2 = 0.19$, indicating that the contingency increased over time. In Experiment 1a, however, the interaction was not significant, $F(1,35) = 0.059$, $MSE = 5363$, $p = 0.810$, $\eta_p^2 < 0.01$, suggesting a relatively stable contingency effect. However, it is noteworthy that the effect in Block 1 was abnormally large in Experiment 1a, both relative to the immediately following blocks and to what we have observed in past reports (and Experiment 1b). With Block 1 eliminated, results are more suggestive of an increasing effect of contingency with block, but this is still not significant, $F(1,35) = 1.975$, $MSE = 3863$, $p = 0.169$, $\eta_p^2 = 0.05$. Indeed, assume that we (a) retain Block 1, and (b) assume that the true slope varied anywhere from 0 ms per block (minimum bound) to 5 ms (maximum bound), meaning a 45 ms increase in the contingency effect from Blocks 1–10 (which exceeds the overall observed contingency effect). Despite the fact that these two assumptions favor the null, the resulting Bayes factor (using the calculator of Dienes, 2014) using a uniform distribution and the observed slope of 0.460 ms/block ($SE = 2.236$) was 0.64, which is not conclusive evidence for a null slope (i.e., because it is greater than 1/3 or 0.33).

Given that Experiment 1b observed some significant effects that were only (non-significant) trends in Experiment 1a, it is interesting to consider what the overall trends were across experiments. We are also able to test whether the pre-presentation variant of the paradigm is, in fact, effective in boosting colour-word contingency learning effects and, in turn, acquisition curves. Thus, the learning blocks for RTs were reanalyzed in a contingency (high vs. low) by block (1–10) by experiment (1a vs. 1b) ANOVA. Interestingly, the contingency effect was significantly larger in Experiment 1b than in Exper-

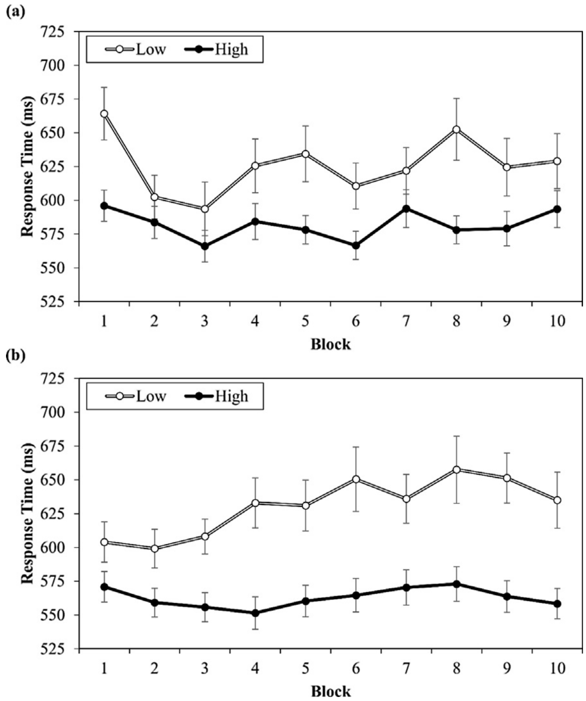


**Fig. 2.** High and low contingency response times with standard errors for (a) Experiment 1a (no pre-exposure) and (b) Experiment 1b (pre-exposed words).

iment 1a, $F(1,68) = 7.770$, $MSE = 6391$, $p = 0.007$, $\eta_p^2 = 0.10$, indicating that word pre-exposure boosts the learning effect. The analysis also replicated the interaction between contingency and block, $F(1,68) = 4.463$, $MSE = 5133$, $p = 0.038$, $\eta_p^2 = 0.06$. The three-way interaction was only marginal, $F(1,68) = 3.125$, $MSE = 5133$, $p = 0.082$, $\eta_p^2 = 0.04$, suggesting that the change in the contingency effect over time in Experiment 2 was not only more robust, but slightly bigger.

*2.2.2. Error percentages*

First, the practice blocks were analysed with a block (1–3) ANOVA for each experiment. In Experiment 1a, the linear contrast of block was not significant, $F(1,35) = 0.279$, $MSE = 9.8$, $p = 0.601$, $\eta_p^2 < 0.01$. As such, the practice blocks were not analysed further. However, mean error rates for the three blocks were 6.9%, 6.4%, and 6.5%, respectively. In Experiment 1b, the main effect of block was marginal, $F(1,33) = 3.421$, $MSE = 49.2$, $p = 0.073$, $\eta_p^2 = 0.09$. More precisely, there were significantly more errors in Block 1 (12.0%) than in Block 2 (9.5%), $t(33) = 2.315$, $SE_{diff} = 1.1$, $p = 0.027$, $\eta_p^2 = 0.14$, and marginally more errors in Block 1 than in Block 3 (8.9%), $t(33) = 1.850$, $SE_{diff} = 1.7$, $p = 0.073$, $\eta_p^2 = 0.09$. However, there was no difference between Blocks 2 and 3, $t(33) = 0.367$, $SE_{diff} = 1.6$, $p = 0.716$, $\eta_p^2 < 0.01$. Thus, there was also some limited evidence for practice benefits in errors.

Next, a contingency (high vs. low) by block (1–10) ANOVA was conducted on the learning blocks for each experiment. The data are presented in Fig. 3. The main effect of contingency was significant in both Experiment 1a, $F(1,35) = 27.331$, $MSE = 96.2$, $p < 0.001$, $\eta_p^2 = 0.44$, and Experiment 1b, $F(1,33) = 33.782$, $MSE = 94.6$, $p < 0.001$, $\eta_p^2 = 0.51$, because errors were less frequent to high contingency trials. The linear main effect of block was also significant in both Experiment 1a, $F(1,35) = 8.409$, $MSE = 100.6$, $p = 0.006$, $\eta_p^2 = 0.19$, and Experiment 1b, $F(1,33) = 13.534$, $MSE = 48.7$, $p < 0.001$, $\eta_p^2 = 0.29$, indicating an increase of errors with block.

Most critical is the interaction between block and contingency. This interaction was significant in Experiment 1b, $F(1,33) = 10.517$, $MSE = 81.4$, $p = 0.003$, $\eta_p^2 = 0.24$, indicating that the contingency effect increased over time. The interaction was not significant in Experiment 1a, $F(1,35) = 2.445$, $MSE = 154.3$, $p = 0.127$, $\eta_p^2 = 0.07$. It is noteworthy, however, that error rates were relatively noisy. Visual inspection of the data hints at an increasing contingency effect, much like the RT data. Indeed, Bayes analysis again revealed no conclusive evidence for the null: with a minimum bound of 0.0% per block, maximum bound of 0.5% (or 4.5% from Block 1–10, again exceeding the overall observed contingency effect), and the observed slope of 0.504%/block ($SE = 0.379$), the Bayes factor was 1.86, which hints at a true effect but inconclusively (i.e., because it is <3).

We again conducted a between-experiment comparison. Unlike the response time data, the contingency effect did not differ significantly between experiments, $F(1,68) = 0.241$, $MSE = 95.4$, $p = 0.625$, $\eta_p^2 < 0.01$. However, the overall analysis replicated the interaction between contingency and block, $F(1,68) = 10.020$, $MSE = 118.9$, $p = 0.002$, $\eta_p^2 = 0.13$, and this was not

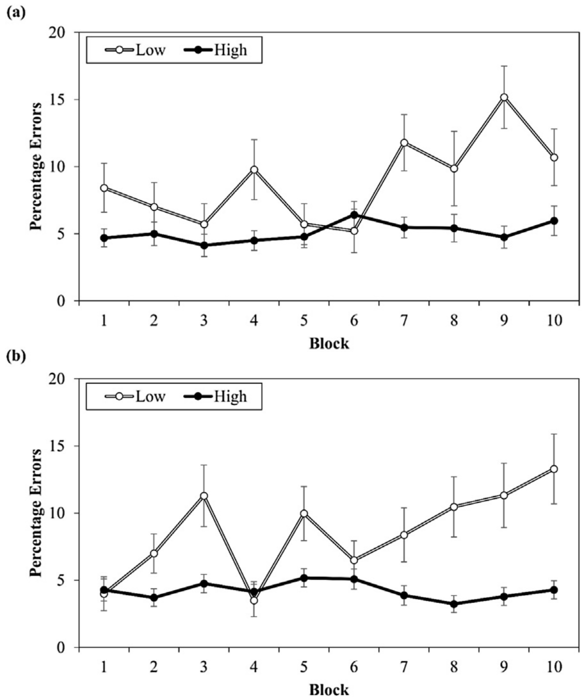


**Fig. 3.** High and low contingency percentage errors with standard errors for (a) Experiment 1a (no pre-exposure) and (b) Experiment 1b (pre-exposed words).

modified by a three-way interaction, $F(1,68) = 0.466$, $MSE = 118.9$, $p = 0.497$, $\eta_p^2 < 0.01$. Thus, also in errors there was evidence for increasing contingency effects, but the pre-exposure manipulation proved the more robust measure.

### 2.2.3. Scaling analysis

Finally, we assessed to what extent contingency learning effects scaled with practice. The data are presented in Fig. 4. In a between participant analysis, the RT contingency learning effect correlated marginally with mean RT in both the parametric, $r(34) = 0.308$, $p = 0.067$, and nonparametric test, $\rho(34) = 0.322$, $p = 0.055$, in Experiment 1a. Thus, there was some evidence for scaling. In Experiment 1b, the correlation was significant with both tests, $r(32) = 0.497$, $p = 0.003$ and $\rho(32) = 0.473$, $p = 0.005$, respectively. Thus, there was clear evidence for scaling. The error contingency learning effect did not correlate with mean RT in the parametric, $r(34) = 0.031$, $p = 0.858$, or nonparametric test, $\rho(34) < 0.001$, $p = 0.998$ in Experiment 1a. This was also true of Experiment 1b, $r(32) = 0.018$, $p = 0.919$ and $\rho(32) = -0.075$, $p = 0.672$, respectively.

In a within participant analysis, an LME was conducted with mean block RT as a scale predictor and the block contingency effect as the dependent measure, with ten blocks as a repeated measure and a subject intercept for each experiment. Thus, we test to what extent, within individual participants, the contingency effect changes with mean block RT. Mean block RT was significantly correlated with the contingency effect in RT both in Experiment 1a (parameter estimate: 0.304), $t(79) = 4.414$, $SE = 0.069$, $p < 0.001$, and in Experiment 1b (parameter estimate: 0.213), $t(63) = 3.120$, $SE = 0.068$, $p = 0.003$, again consistent with scaling. The parameter estimates indicate, respectively, that the contingency effect increased by 0.304 ms and 0.213 ms for every 1 ms change in mean block RT. Mean block RT did not correlate with the contingency effect in errors both in Experiment 1a (parameter estimate: 0.017), $t(76) = 1.180$, $SE = 0.010$, $p = 0.241$, and Experiment 1b (parameter estimate: $-0.007$), $t(57) = -0.749$, $SE = 0.009$, $p = 0.457$.

## 2.3. Discussion

Experiment 1 achieved several interesting and novel things. First, the overall contingency learning effect was markedly increased with pre-exposed words. This suggests a useful adjustment for all future work with the colour-word contingency learning paradigm (or even other learning paradigms). Though the effect is already extraordinarily robust without this change, anything to boost the magnitude of the effect is certain welcome. This is particularly the case when aiming to test the influence of other factors on the contingency learning effect (e.g., acquisition), which may be less robust (e.g., because trials must be broken down into smaller blocks, each of which contains very few low contingency trials).

Related to this, the second interesting observation of Experiment 1 is that the contingency learning effect *was* observed to increase across blocks, but this was clearer when the predictive words were pre-exposed (Experiment 1b) rather than presented concurrently with the target colour (Experiment 1a). The same patterns were observed in Experiment 1a, albeit

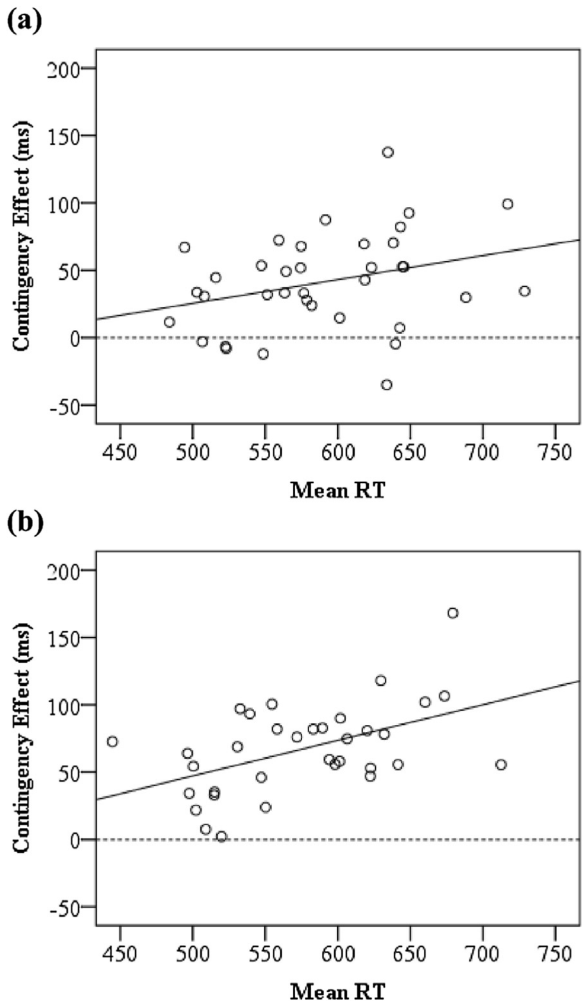


**Fig. 4.** Scatterplots relating mean response time (x-axis) to the contingency effect (y-axis) for (a) Experiment 1a (no pre-exposure) and (b) Experiment 1b (pre-exposed words).

much less robustly. This is easily explained by the added potency of pre-exposed words: a larger contingency effect produces a concomitantly larger slope in the block by contingency interaction. In other words, the interaction *scales* with the mean contingency effect (i.e., in the same way and for the same reason that the contingency effect scales with mean RT).

Third, participants performed an extended colour identification practice phase to reinforce colour-to-key mappings before the learning phase began. This procedure eliminated the typical practice-based speedups normally seen at the start of an experiment (i.e., overall mean RT was not substantially slower in the first few blocks than in later blocks). Unlike our previous studies on acquisition, this manipulation produced an increase in the contingency effect across blocks, visually in Experiment 1a and significantly in Experiment 1b in both response times and errors (and also significantly when averaged across both experiments). It is noteworthy that, while we did observe increasing contingency effects, robust learning effects were already evident as early as the first block of 30 trials. Thus, learning is certainly rapid, but not as instantaneously asymptotic as previously supposed.

## 3. Experiment 2

Experiment 2 had three primary aims. First, we aimed to replicate the finding of an increasing contingency effect over blocks with pre-presentation of the predictive word stimulus. This seems particularly important given that Experiment 1 was the first such demonstration. Second, we aimed to test to what extent a cumulative learning effect might impact rates of *unlearning*. In Schmidt et al. (2010), participants first learned contingencies between words and colours, and then the contingencies were removed from the task (i.e., the same words were now presented equally often in all colours). The aim was to determine how quickly the learning effect would disappear (i.e., be unlearned). After three short, 18-trial learning blocks, the contingency effect was already cut in half in the very first 18-trial block of unlearning. In the following nine blocks, the effect was eliminated. This was taken to suggest that unlearning, like learning, is extremely rapid and only based on a small "window" of immediately-preceding trials. Given the results of the current investigation, however, it might be supposed that unlearning will occur at a less rapid rate if the initial learning phase is longer. In particular, if there is a cumulative effect of contingency knowledge over time, then it might take longer to unlearn a contingency that has been extensively reinforced than one that was encountered only briefly. As such, half of the participants in the present experiment had a short learning phase before unlearning, and the other half had a longer learning phase. Additionally, given that pre-presentation of the word seems to boost learning, we might additionally expect that the contingency effect will persist longer during the unlearning phase.

The third goal of the present study is to investigate *relearning* for the first time in the colour-word contingency learning task. After either a short learning phase and long unlearning phase or a long learning phase and a short unlearning phase, contingencies were reintroduced to the task. The reintroduced contingencies in this relearning phase were identical to the initial contingencies in the learning phase (e.g., if "search" was presented most often in purple in the learning phase, then it was again presented most often in purple in the relearning phase). The first more general question is how rapidly relearning progresses. After a period of null (unpredictive) contingencies, will relearning appear rapidly? That is, will the contingency effect reappear in the first block or two of relearning (similar to the fast acquisition rate during initial learning at the start of an experiment)? Or will recovery from the unlearning phase take an extended period of time? The second more specific question is whether relearning rates will differ between the two groups of participants. That is, will relearning be faster or more pronounced in participants experiencing an extended learning phase and brief unlearning phase relative to participants experiencing a brief learning phase and an extended unlearning phase? Thus, Experiment 2 provides us with two added ways (i.e., unlearning and relearning) to overcome problems with response speeding with practice (i.e., by studying adaptations to changes in contingencies at different phases of the experiment) and to investigate potentially longer-term cumulative effects of contingencies.

### 3.1. Methods

#### 3.1.1. Participants

Ninety-six Ghent University undergraduates participated in the experiment in exchange for €5. None of the participants had participated in Experiment 1.

#### 3.1.2. Apparatus and data analysis

The apparatus and data analysis for Experiment 2 were identical in all respects to Experiments 1. Two participants (one from each group) were excluded for having error rates in excess of 20% in the main phases of the experiment.

#### 3.1.3. Design and procedure

The design and procedure of Experiment 2 was identical in all respects to Experiment 1b with the following exceptions. The initial practice phase consisted of 3 shorter blocks of 18 trials each (54 total), consisting again of the stimulus "@@@@" equally often in each of the three colours (six times each per block). The learning phase consisted of either 3 or 12 blocks of 18 trials each. The words and colours were identical to the previous experiments, only with slightly different contingencies: each word was presented 67% (4 of 6 times per block) in one colour, and 17% (1 of 6) in each of the remaining two colours. This was followed by an unlearning phase consisting of either 12 or 3 blocks of 18 trials each. Participants who received the short learning phase received the long unlearning phase, and vice versa. In unlearning blocks, each word was now presented equally often in all colours (2 of 6, or 33%). Finally, this was followed by a *relearning phase*, in which words were again presented with the learning phase contingencies. This comprised of three blocks of 18 trials, with identical contingencies to the initial learning phase. Thus, each participant received a total of 18 blocks of 18 experimental trials each, in addition to 54 practice trials (378 total).

### 3.2. Results

#### 3.2.1. Response times

First, the practice blocks were analysed with a block (1–3) by condition (long vs. short learning) ANOVA. Reassuringly, there was no main effect of condition, $F(1,92) = 1.628$, $MSE = 20413$, $p = 0.205$, $\eta_p^2 = 0.02$, and also no interaction between condition and block, $F(1,92) = 0.091$, $MSE = 4169$, $p = 0.764$, $\eta_p^2 < 0.01$, indicating no pre-existing differences between groups. The linear contrast of block was significant, $F(1,92) = 105.424$, $MSE = 4169$, $p < 0.001$, $\eta_p^2 = 0.53$, indicating that performance improved with practice. More precisely, Block 1 performance (687 ms) was significantly slower than Block 2 (610 ms), $t(93) = 6.935$, $SE_{diff} = 11$, $p < 0.001$, $\eta_p^2 = 0.34$, and Block 3 (590 ms), $t(93) = 10.318$, $SE_{diff} = 9$, $p < 0.001$, $\eta_p^2 = 0.53$, and Block 2 was significantly slower than Block 3, $t(93) = 2.859$, $SE_{diff} = 7$, $p = 0.005$, $\eta_p^2 = 0.08$. As in the previous two experiments, there was clear evidence for practice benefits, particularly early on.

Next, a contingency (high vs. low) by block (1–12) ANOVA was conducted on the learning blocks of the long learning phase group. The data are presented in Fig. 5a. The main effect of contingency was significant, $F(1,46) = 78.832$, $MSE = 3948$, $p < 0.001$, $\eta_p^2 = 0.63$, because responses were faster on high contingency trials. The linear main effect of block was not significant,

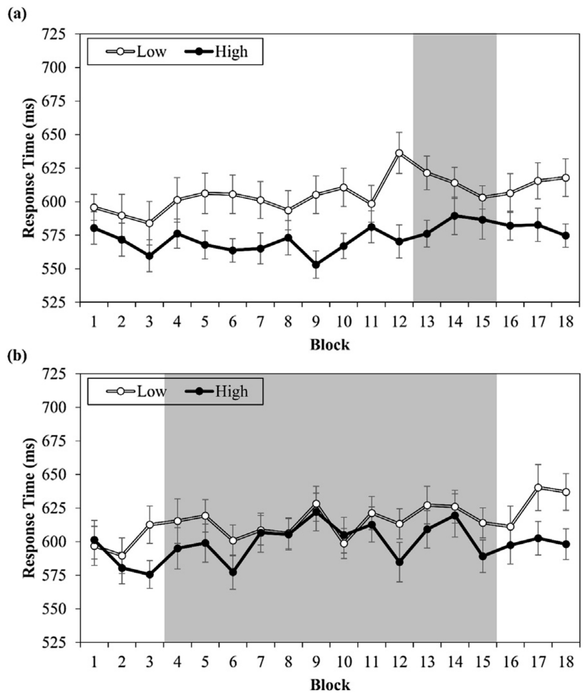


**Fig. 5.** Experiment 2 high and low contingency response times with standard errors for (a) long learning phase (short unlearning phase) and (b) short learning phase (long unlearning phase). Unlearning phase marked in grey.

$F(1,46) = 1.702$, $MSE = 9080$, $p = 0.199$, $\eta_p^2 = 0.04$. As before, contingency and block significantly interacted, $F(1,46) = 8.086$, $MSE = 2937$, $p = 0.007$, $\eta_p^2 = 0.15$, again indicating that the contingency effect increased over time.

Next, a similar contingency (high vs. low) by block (1–3) ANOVA was conducted on the learning blocks of the short learning phase group. The data are presented in Fig. 5b. The main effect of contingency was significant, $F(1,46) = 4.911$, $MSE = 2780$, $p = 0.032$, $\eta_p^2 = 0.10$, because responses were faster on high contingency trials. The linear main effect of block was not significant, $F(1,46) = 0.204$, $MSE = 5775$, $p = 0.654$, $\eta_p^2 < 0.01$. Critically, contingency and block significantly interacted, $F(1,46) = 4.083$, $MSE = 4966$, $p = 0.049$, $\eta_p^2 = 0.08$. Thus, evidence for an increasing contingency effect was already observable with just three short blocks of 18 trials. Reassuringly, an additional contingency (high vs. low) by block (1–3) by condition (short vs. long learning phase) ANOVA on the first three blocks revealed no effects of condition on any of the main effects or interactions (all $F\text{s} \leq 1.607$, all $p\text{s} \geq 0.208$).

Results from the unlearning and relearning phases can also be observed in Fig. 5. There were no main effects for block in either phase for either group of participants (all $F\text{s} \leq 1.107$, all $p\text{s} \geq 0.298$). For participants with the long learning phase (and therefore short unlearning phase), there was a robust overall contingency effect during unlearning, $F(1,46) = 14.525$, $MSE = 4020$, $p < 0.001$, $\eta_p^2 = 0.24$. The linear decrease in the contingency effect across unlearning blocks was marginal, $F(1,46) = 3.002$, $MSE = 3235$, $p = 0.090$, $\eta_p^2 = 0.06$. Including the last learning block in the ANOVA (i.e., Blocks 12–15), this decreasing contingency effect across blocks was significant, $F(1,46) = 11.747$, $MSE = 2863$, $p = 0.001$, $\eta_p^2 = 0.20$. Thus, some carryover from the learning phase to the unlearning phase was apparent, but even with a lengthier learning phase, the contingency effect did diminish. This is consistent with the notion that the learning rate is high (i.e., given that the effect does diminish), but there is some cumulative effect of contingencies learned across the experiment (i.e., given that the effect does persist partially after the contingency is removed). For the same group of participants, the overall contingency effect was highly robust during relearning, $F(1,46) = 28.287$, $MSE = 2782$, $p < 0.001$, $\eta_p^2 = 0.38$. Though there was some numerical trend for an increasing contingency effect across blocks, this was not significant, $F(1,46) = 1.630$, $MSE = 2589$, $p = 0.208$, $\eta_p^2 = 0.03$, indicating that the contingency was relearned very quickly.

For participants with the short learning phase (and therefore long unlearning phase), the contingency effect seemed to hover slightly above zero across unlearning blocks. The overall contingency effect was significant, $F(1,46) = 10.820$, $MSE = 4240$, $p = 0.002$, $\eta_p^2 = 0.19$. There was also no evidence for a decrease in the contingency effect across blocks, $F(1,46) = 0.012$, $MSE = 3238$, $p = 0.912$, $\eta_p^2 < 0.01$, primarily because the contingency effect was small right from the start of the unlearning phase. Indeed, if the last block of learning and first three blocks of unlearning are included with the condition factor in an ANOVA, the rate of unlearning was marginally steeper in the long learning phase group, $F(1,92) = 3.058$, $MSE = 3178$, $p = 0.084$, $\eta_p^2 = 0.03$. Thus, carryover from the learning phase to the unlearning phase was again apparent, albeit less dramatically as in the long learning phase group. Interestingly, however, if the unlearning phase was divided into the

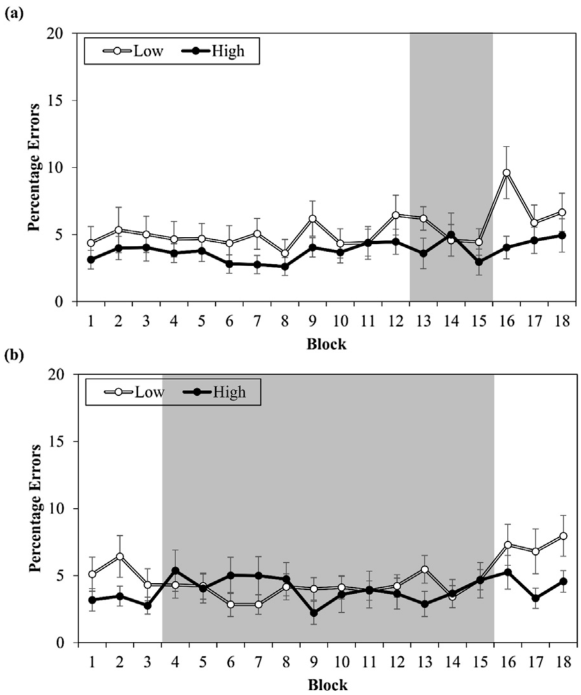


**Fig. 6.** Experiment 2 high and low contingency percentage errors with standard errors for (a) long learning phase (short unlearning phase) and (b) short learning phase (long unlearning phase). Unlearning phase marked in grey.

first and last halves of unlearning (i.e., Blocks 4–9 and Blocks 10–15), the overall contingency effect was significant in both the first half of unlearning, $F(1,46) = 5.612$, $MSE = 3713$, $p = 0.022$, $\eta_p^2 = 0.11$, and the second half, $F(1,46) = 6.939$, $MSE = 3623$, $p = 0.011$, $\eta_p^2 = 0.13$. This indicates that, though reduced, a cumulative effect of contingencies persists even after a relatively short learning phase and long unlearning phase. In the same group of participants, there was a significant contingency effect during relearning, $F(1,46) = 13.943$, $MSE = 4590$, $p < 0.001$, $\eta_p^2 = 0.23$. As with the long learning phase group, there was some numerical trend for an increasing contingency effect across blocks, but this was again not significant, $F(1,46) = 2.225$, $MSE = 3314$, $p = 0.143$, $\eta_p^2 = 0.05$, indicating again that the contingency was relearned very quickly. There was no apparent difference in the rate of relearning across groups, $F(1,92) = 0.074$, $MSE = 2952$, $p = 0.786$, $\eta_p^2 < 0.01$, probably because of how quickly the contingency was relearned in both groups.

### 3.2.2. Error percentages

First, the practice blocks were analysed with a block (1–3) by condition (long vs. short learning) ANOVA. Again, there was no main effect of condition, $F(1,92) = 0.063$, $MSE = 388$, $p = 0.802$, $\eta_p^2 < 0.01$, and also no interaction between condition and block, $F(1,92) = 0.009$, $MSE = 75$, $p = 0.962$, $\eta_p^2 < 0.01$, indicating no pre-existing differences between groups. The linear contrast of block was significant, $F(1,92) = 8.915$, $MSE = 75$, $p = 0.004$, $\eta_p^2 = 0.09$, indicating that errors decreased with practice. More precisely, Block 1 errors (9.2%) were significantly higher than in Block 2 (6.2%), $t(93) = 3.203$, $SE_{diff} = 0.9$, $p = 0.002$, $\eta_p^2 = 0.10$, and Block 3 (5.4%), $t(93) = 3.002$, $SE_{diff} = 1.3$, $p = 0.003$, $\eta_p^2 = 0.09$. Blocks 2 and 3 did not significantly differ, $t(93) = 0.789$, $SE_{diff} = 1.0$, $p = 0.432$, $\eta_p^2 < 0.01$. Again, there was clear evidence for practice benefits, particularly early on.

The data for the main blocks of the experiment are presented in Fig. 6. Given the binary nature of errors, the small block sizes, and the uneven distribution of high and low contingency trials per block (i.e., relatively few low contingency trials in the learning and relearning blocks, and relatively few "high contingency" trials in the unlearning blocks), it is probably not surprising that the errors were generally less informative than response times for the main blocks of the experiment. During the learning phase, there was a significant overall contingency effect in errors for the long learning phase group, $F(1,46) = 6.624$, $MSE = 68$, $p = 0.013$, $\eta_p^2 = 0.13$, and the short learning phase group, $F(1,46) = 5.783$, $MSE = 56$, $p = 0.020$, $\eta_p^2 = 0.11$. However, there were no main effects of block or interactions between block and contingency for either group (all $F$s $\leq 0.716$, all $p$s $\geq 0.402$). Thus, acquisition curves were not observed in the errors, likely due to noisier error data.

During the unlearning phase, the contingency effect was not significant for the long learning phase group, $F(1,46) = 1.758$, $MSE = 60$, $p = 0.191$, $\eta_p^2 = 0.04$, or the short learning phase group, $F(1,46) = 0.010$, $MSE = 70$, $p = 0.921$, $\eta_p^2 < 0.01$. During the relearning phase, the contingency effect was again significant for the long learning phase group, $F(1,46) = 5.569$, $MSE = 104$, $p = 0.023$, $\eta_p^2 = 0.11$, and the short learning phase group, $F(1,46) = 9.935$, $MSE = 63$, $p = 0.003$, $\eta_p^2 = 0.18$. There were no main effects of block or interactions between block and contingency for either group (all $Fs \leq 2.336$, all $ps \geq 0.133$). Thus, the error data captures only the larger trends in the data, but is otherwise generally consistent with the response times.

### 3.2.3. Scaling analysis

As in Experiment 1, we then performed scaling analyses on the learning phase data with the identical data analysis strategy. Of course, including the unlearning and relearning data would confound the analysis, given the change in contingencies, so these data were excluded. The learning phase in the long learning phase group did include more blocks (12) than in Experiment 1 (10), but overall less observations (216 and 300 trials, respectively), given the smaller block sizes. For the short learning phase group, scaling analyses are less meaningful given the small number of blocks (3) and trials (54), and understandably revealed no significant effects (all $ps > 0.05$). For the participant-level analysis of the long learning phase group, the correlation between mean RT and the response time contingency effect was marginal with the parametric test, $r(47) = 0.266$, $p = 0.071$, but not significant with the nonparametric test, $\rho(47) = 0.208$, $p = 0.161$. As before, mean RT did not correlated with the error contingency effect in either test, $r(47) = -0.094$, $p = 0.531$ and $\rho(34) = -0.130$, $p = 0.382$. For the block level analysis, mean block RT correlated significantly with the block response time contingency effect (parameter estimate: 0.131), $t(114) = 2.711$, $SE = 0.048$, $p = 0.008$. Again, mean block RT did not correlate with the block error contingency effect (parameter estimate: −0.001), $t(133) = -0.179$, $SE = 0.006$, $p = 0.858$. Thus, evidence for scaling was again observed in the response time data, but less robustly in the participant-level analysis.

### 3.3. Discussion

In Experiment 2, we replicated the observation of increasing contingency effects over blocks, in addition to scaling with mean response times. We additionally observed some persistence of the contingency effect from the learning phase to the unlearning phase, particularly in the group of participants with a longer learning phase. However, even in the short learning phase group, a contingency effect was still observable in the second half of the unlearning blocks. These findings are consistent with the notion that recent events have the largest effect on behaviour (i.e., high learning rate), but that more distant events do have some cumulative effect on learning. Relearning also seems to occur very quickly. Indeed, relearning was quick enough that we did not observe a significant increase in the contingency effect across relearning blocks (i.e., because the effect was already present from the start of the relearning phase). There were some suggestive (but non-significant) trends, but the change in the contingency effect across blocks was not particularly drastic.

## 4. General discussion

In performance paradigms, acquisition of contingency knowledge appears to occur almost instantaneously (e.g., Lewicki, 1985, 1986; Nissen & Bullemer, 1987; Schmidt et al., 2010). The present results further demonstrate that learning does, however, continue to grow over the course of an experiment. Most likely, this is due to learning following a power or exponential function with a high learning rate. That is, responding is heavily determined by learning from recent trials, and previous trials have ever decreasing effects on the current trial the longer ago it occurred. Thus, like the overall performance improvements with practice, contingency learning effects will increase toward a theoretical asymptote over time, with less gains to be achieved the more that has already been learned.

### 4.1. Mechanisms of contingency learning

Our results suggest that learning occurs very rapidly, but does not reach asymptote immediately. We consider the implications of this for mechanistic accounts of contingency learning. First, it is clearly not the case that participants are responding in a fixed way to the conditional probabilities, or $\Delta$P (Allan, 1980; Jenkins & Ward, 1965; Ward & Jenkins, 1965), between stimuli and responses. Because the contingencies do not change during acquisition, it seems unlikely that participants are responding on the basis of the computed probability of a given response given the stimuli presented. Given that changes in task contingencies (i.e., during unlearning and relearning) are similarly extremely rapid it also seems clear that learning is not simply cumulative across the task as a whole. Rapid adaptation to a new contingency requires heavier weighting of recent events (discussed in more detail later).

Another possibility is that learning in the colour-word contingency paradigm occurs via the storage and retrieval of trial episodes (Logan, 1988). On each trial, an event is encoded into memory, which records the stimuli that were presented and the response that was made. On subsequent trials, presentation of a stimulus leads to automatic retrieval of similar episodes, thereby biasing the most frequent response. For instance, if the word "choose" is presented most often in purple, then experiencing the word "choose" again will lead to the retrieval of episodes of trials in which "choose" was also presented. Due to the contingency, most of these will point to a purple response. A computational model of this sort of learning, the Parallel Episodic Processing (PEP) model, has already been demonstrated to produce such contingency learning effects (Schmidt, 2013a; see also, Schmidt & Weissman, 2016; Schmidt, 2013b, 2016) and a forthcoming paper successfully models the acquisition curves observed in the current report (Schmidt, De Houwer, & Rothermund, 2016).

In order for an episodic model to explain all of our results, however, it must be assumed that recently-encoded episodes are more strongly retrieved than more distant events. This is the case in the PEP model, but not in the model of Logan (where each episode "races" at an equal speed for retrieval). If all episodes were equally weighted (i.e., without decay of older episodes), then adaptation to a change in the contingency later on in an experiment would be slow. For instance, consider participants experiencing the long unlearning phase in Experiment 3. Short learning phase aside, most of their trial memories will point to a null contingency. Thus, the relearning phase would have to be quite long before a contingency effect would begin to re-emerge (i.e., until the experiment-wide contingency increased sufficiently). If recently-encoded memories are retrieved more strongly than older ones, however, then adaptation to a new contingency will unfold rapidly, as we observed. The rapid unlearning rate, even after a relatively long learning phase, also adds credence to this notion.

The implications are similar for associative models of learning, such as the Rescorla-Wagner model (Rescorla & Wagner, 1972). The learning rate must be assumed to be high enough to produce rapid learning and equally rapid adjustment to changes in contingencies (e.g., during unlearning or relearning), but not so high that associative strengths approach asymptote too rapidly. In general, our results suggest that a viable account of learning in this sort of performance task must be rapid enough to acquire the contingency quickly and to be highly adaptive to changes in stimulus-response contingencies.

### *4.2. Performance measures*

As we mentioned at the outset of the manuscript, performance measures of learning are highly useful in that they provide highly robust measures of learning that can be studied on-line (i.e., while learning is occurring). However, the present manuscript also highlights the potential complications with practice-based performance improvements when attempting to study acquisition. Because contingency effects seem to scale with response time and response times tend to be slower early on in an experiment, contingency effects will tend to be increased early on. Though obviously still a genuine learning effect regardless of how the effect is scaled, this does make it difficult to directly relate the magnitude of a contingency effect to the amount of underlying contingency knowledge. Similar points have been raised outside the acquisition domain (Kaufman et al., 2010; Stevens et al., 2002; Urry et al., 2015). The extended training phase used prior to contingency learning in the present experiments partially circumvents this problem. Similarly, unlearning and/or relearning can be used to study transfer of contingency knowledge after varying amounts of learning and at different points during practice.

Another way of dealing with practice-induced scaling would be to convert the response time contingency effect to a proportion of mean RT (e.g., Kaufman et al., 2010). In the case of the current paradigm, this would entail the following transform to each block: (low contingency – high contingency)/mean RT. The mean RT could alternatively be replaced by either the high or low contingency RT. With such a transform, the contingency effect is rescaled to mean RT, thereby controlling for practice statistically. The reason that we did not opt for this approach in the current study of acquisition is the potentially circular logic this would entail for our purposes. That is, we set out to answer the questions: (a) does the contingency effect scale with mean RT? and (b) if so, does the contingency effect increase over practice after accounting for scaling? Controlling for overall RT statistically presumes a priori that the contingency effect *does* scale with mean RT (Question a) in order to assess acquisition (Question b). If the contingency effect did not scale with mean RT and was additive with the practice effect (as in Panel B of Fig. 1), then a statistical control for mean RT could produce false evidence for a positive acquisition curve (as in Panel C of Fig. 1). Thus, an experimental control seemed more appropriate. On the other hand, our results indicate that the contingency effect does scale with mean RT, which might suggest that some form of statistical control could be appropriate in future research. Incidentally, unreported analyses confirmed that this statistical control has little impact on the present datasets, due to the relatively stable mean block RT.

As a more general point, it is also noteworthy that practice-based improvements can be conceived as resulting from the same (or similar) processes that produce the contingency learning effect. Contingency learning effects seem to be driven by the learning of word-response correspondences (Schmidt & De Houwer, 2012b; Schmidt et al., 2007), which we have previously suggested might be due to episodic storage and retrieval. Practice benefits might also be due to episodic learning (Logan, 1988), but the learned regularity is between the colour and the response. For instance, each time a participant presses the "j" key to a purple stimulus a new episode is created that includes the colour purple and the response "j" (in addition to whichever word was presented). Over time, these episodes accumulate, such that there are more and more "purple-'j' key" episodes to retrieve to assist in responding. Thus, while we have treated practice effects as a "confound" in the current investigation of contingency learning acquisition, both effects can parsimoniously be explained by the same memory mechanisms.

### *4.3. Relation to past results and future directions*

The present results might seem to conflict with the findings of Schmidt and De Houwer (2012c). In that work, the authors varied the stimulus onset asynchrony (SOA) or inter-stimulus interval (ISI) between predictive nonwords (e.g., "yalan," "zarif," etc.) and colour word targets. Nonwords were predictive of colour word targets, much in the same way that neutral words were predictive of print colour targets in the present investigation. Over a wide range of lags (from 1200 ms to 50 ms pre-exposure) no noticeable changes in the contingency learning effect were observed (for related work, see Elsner & Hommel, 2004). They interpreted this as indicating that temporal contiguity (i.e., closeness in time) between predictive stimuli and target stimuli has minimal effect on learning. In the present report, however, we observed that pre-exposing the word led to very noticeable increases in the contingency learning effect. Though it is certainly possible that more incidental differences in the procedure account for this difference (e.g., the use of nonwords vs. real words, or the use of colour word vs. print colour targets), one explanation seems more sensible to us. As discussed in the aforementioned paper, there were two potentially counteracting factors in the SOA/ISI studies. On the one hand, extra pre-exposure might allow for more preparation time (as we propose), thereby boosting learning effects (i.e., the expression of learning). On the other hand, non-integrated stimuli that are presented closer together in time might be more strongly bound together, thereby boosting the strength of the underlying learned contingencies. Thus, longer delays between (non)words and colour targets provides extra preparation time, but comes at the cost of weaker binding. By colourizing the words in the present Experiments 1b and 2, however, we might both: (a) maintain a strong binding between words and colours by integrating them, and (b) allow extra preparation time. Future research might thus aim to clarify these issues further (e.g., by manipulating SOA with our pre-exposure procedure).

Also interesting, Schmidt et al. (2010) argued that the "window" of previous trials that impact the contingency effect is quite small. In other words, they suggested that only a small number of immediately-preceding trials influenced the magnitude of the effect. However, in the present experiments we observed that the contingency effect continued to rise across the learning blocks. Similarly, in Experiment 2 we observed that an initially learned contingency can persist for quite some time after the contingencies are removed, albeit in a much diminished form. This might suggest that more distant trial memories, though weaker than more recent ones, also influence the observed effect. Future research might investigate these issues further by, for instance, seeing to what extent manipulations of the strength of contingencies early on in an experiment influence the size of contingency effects later on in the procedure.

In addition to research on acquisition rates, unlearning and extinction have proved important in studying various types of learning (De Houwer, 2009; De Houwer, Thomas, & Baeyens, 2001; Shanks, 2007). In addition to the current Experiment 2, only one study so far has investigated unlearning in the colour-word contingency learning paradigm (Schmidt et al., 2010). As with the prior report, we observed that the contingency effect diminishes when contingencies are removed (i.e., during unlearning). Different from the prior report, we did observe some persistence of the contingency effect during unlearning. We also investigated, for the first time, to what extent a longer learning phase might impact the unlearning rate. Interestingly, we did *not* observe that the contingency effect remained large longer for the group of participants with a longer learning phase. Indeed, the contingency effect decreased very rapidly in this group. However, future research might investigate to what extent an even longer learning phase does or does not lead to a more strongly persistent contingency effect. For instance, if participants are given *days* of practice with contingencies before unlearning, will the contingency effect again decrease quickly during unlearning or will the heavily-reinforced contingency knowledge take longer to unlearn?

### *4.4. Limitations*

One limitation of the present work is that we did not directly compare learning in participants with versus without an initial practice phase. That is, while the present report provides clear evidence that contingency effects increase over time, we did not directly test the reason *why* the current procedure would reveal this acquisition curve, whereas past reports failed to. We reasoned that eliminating initial performance-based speedups with practice would allow us to observe a positive acquisition curve. Consistent with this, we observed an increasing contingency effect across blocks, which is unlike what we observed in our previous reports with this paradigm. More ideally, it would be useful to show directly that the slope of the acquisition curve is statistically steeper with versus without an initial practice phase. Practically speaking, however, the sample size required to detect such a between-participant interaction might be prohibitive. Furthermore, the differences in the acquisition curves observed between the current three experiments and our previous work are striking enough already. In addition, our scaling analyses provide clear converging evidence that failures to control for initial practice in the task is problematic.

As another limitation, the increasing effects of contingencies across blocks might be reinterpreted in another manner. It might be argued that the contingency effect grows over time because retrieval of contingency knowledge from memory becomes more efficient over time. That is, even if contingency knowledge *does* reach asymptote near instantly (e.g., Panels A and B in Fig. 1), the contingency effect might nevertheless continue to grow over the course of the experiment as a result of increasingly more effective *use* of this contingency information to anticipate the likely response. The persistence of the contingency effect across the unlearning blocks might seem to partially argue against this possibility (i.e., some carryover of contingency knowledge must be assumed), but it is certainly possible that changes in the expression of contingency knowledge over time do play some role in the acquisition curves observed in the present report. Future research might therefore aim to study true acquisition and the expression of contingency knowledge separately.

## 5. Conclusions

Though it should be clear that contingency knowledge is extracted quite rapidly in the colour-word contingency learning paradigm, the present results further suggest that the effect does continue to grow over time. The present report also novelly illustrates how performance improvements with practice can play a potentially confounding role when investigating the development of an effect over time. With regard to the contingency effect in particular, slower early block performance might inflate estimates of contingency acquisition. Our scaling analyses lend credence to this notion. The present results also demonstrate for the first time that performance indices of contingency learning can be magnified by pre-exposing predictive words, likely because of the added preparation time.

## Acknowledgements

This research was supported by Grant BOF09/01M00209 of Ghent University to Jan De Houwer and by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33). James R. Schmidt is supported by a postdoctoral researcher grant (121184N) of the Research Foundation – Flanders (FWO – Vlaanderen).

## References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*, 147–149.

Atalay, N. B., & Misirlisoy, M. (2012). Can contingency learning alone account for item-specific control? Evidence from within- and between-language ISPC effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1578–1590.

Carlson, K. A., & Flowers, J. H. (1996). Intentional versus unintentional use of contingencies between perceptual events. *Perception & Psychophysics*, *58*, 460–470.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853–869.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*, 1–20.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20*, 470–484 [Article 781].

Ellis, N. R., & Dulaney, C. L. (1991). Further evidence for cognitive inertia of persons with mental retardation. *American Journal on Mental Retardation*, *95*, 613–621.

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research*, *68*, 138–154.

Frings, C., Rothermund, K., & Wentura, D. (2007). Distractor repetitions retrieve previous responses to targets. *Quarterly Journal of Experimental Psychology*, *60*, 1367–1377.

Hazeltine, E., & Mordkoff, J. T. (2014). Resolved but not forgotten: Stroop conflict dredges up the past. *Frontiers in Psychology*, *5* [Article 1327].

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185–207.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, *79*, 1–17.

Kaufman, S. B., De Young, C. G., Gray, J. R., Jimenez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Levin, Y., & Tzelgov, J. (2016). Contingency learning is not affected by conflict experience: Evidence from a task conflict-free, item-specific Stroop paradigm. *Acta Psychologica*, *164*, 39–45.

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *American Psychologist*, *47*, 796–801.

Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology*, *48*, 563–574.

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 135–146.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, *26*, 201–211.

Miller, J. (1987). Priming is not necessary for selective-attention failures: Semantic effects of unattended, unprimed letters. *Perception & Psychophysics*, *41*, 419–434.

Musen, G., & Squire, L. R. (1993). Implicit learning of color-word associations using a Stroop paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *19*, 789–798.

Myung, I. J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory & Cognition*, *28*, 832–840.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.

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

Rothermund, K., Wentura, D., & De Houwer, J. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 482–495.

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*, 514–523.

Schmidt, J. R., & De Houwer, J. (2012a). Adding the goal to learn strengthens learning in an unintentional learning task. *Psychonomic Bulletin & Review*, *19*, 723–728.

Schmidt, J. R., & De Houwer, J. (2012b). Contingency learning with evaluative stimuli: Testing the generality of contingency learning in a performance paradigm. *Experimental Psychology*, *59*, 175–182.

Schmidt, J. R., & De Houwer, J. (2012c). Does temporal contiguity moderate contingency learning in a speeded performance task? *Quarterly Journal of Experimental Psychology*, *65*, 408–425.

Schmidt, J. R., & De Houwer, J. (2012d). Learning, awareness, and instruction: Subjective contingency awareness does matter in the colour-word contingency learning paradigm. *Consciousness and Cognition*, *21*, 1754–1768.

Schmidt, J. R., & De Houwer, J. (2016). Contingency learning tracks with stimulus-response proportion: No evidence of misprediction costs. *Experimental Psychology*, *63*, 79–88.

Schmidt, J. R., & Weissman, D. H. (2016). Congruency sequence effects and previous response times: Conflict adaptation or temporal learning? *Psychological Research*, *80*, 590–607.

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*, 421–435.

Schmidt, J. R., De Houwer, J., & Besner, D. (2010). Contingency learning and unlearning in the blink of an eye: A resource dependent process. *Consciousness and Cognition*, *19*, 235–250.

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*. Manuscript submitted for publication.

Schmidt, J. R. (2013a). The Parallel Episodic Processing (PEP) model: Dissociating contingency and conflict adaptation in the item-specific proportion congruent paradigm. *Acta Psychologica*, *142*, 119–126.

Schmidt, J. R. (2013b). Temporal learning and list-level proportion congruency: Conflict adaptation or learning when to respond? *PLoS One*, *8*, e0082320.

Schmidt, J. R. (2016). Proportion congruency and practice: A contingency learning account of asymmetric list shifting effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1496–1505.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, *60*, 291–309.

Simon, J. R., Craft, J. L., & Webster, J. B. (1973). Reactions toward stimulus source: Analysis of correct responses and errors over a five-day period. *Journal of Experimental Psychology*, *101*, 175–178.

Stevens, A., Schwarz, J., Schwarz, B., Ruf, I., Kolter, T., & Czekalla, J. (2002). Implicit and explicit learning in schizophrenics treated with olanzapine and with classic neuroleptics. *Psychopharmacology*, *160*, 299–306.

Urry, K., Burns, N. R., & Baetu, I. (2015). Accuracy-based measures provide a better measure of sequence learning than reaction time-based measures. *Frontiers in Psychology*, *6*.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, *19*, 231–241 [Article 1158].