**Contingency Learning Tracks with Stimulus-Response Proportion:**

**No Evidence of Misprediction Costs**

James R. Schmidt & Jan De Houwer

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

**Word Count:** 5845

**Abstract**

We investigate the processes involved in human contingency learning using the colour-word contingency learning paradigm. In this task, participants respond to the print colour of neutral words. Each word is frequently presented in one colour. Results show that participants respond faster and more accurately to words presented in their expected colour. In Experiment 1, we observed better performance for high relative to medium frequency word-colour pairs, and for medium relative to low frequency pairs. Within the medium frequency condition, it did not matter whether the word was predictive of a currently-unpresented colour, or the colour was predictive of a currently-unpresented word. We conclude that a given word facilitates each potential response proportional to how often they co-occurred. In contrast, there was no evidence for costs associated with violations of high-frequency expectancies. Experiment 2 further introduced a novel word baseline condition, which also provided no evidence for competition between retrieved responses.

**Keywords:** contingency learning; facilitation; interference; prediction; misprediction; frequency

**Introduction**

Understanding the mechanisms via which the human cognitive system is able to learn the regularities in its environment and, in turn, to use this information in a predictive way to maximize performance is a key area of interest in cognitive psychology (Allan, 2005; Beckers, De Houwer, & Matute, 2007; Shanks, 2010). One useful performance (i.e., response time) measure of learning is the colour-word contingency learning paradigm (Schmidt, Crump, Cheesman, & Besner, 2007; for a related study-test paradigm, see Musen & Squire, 1993). In this task, participants identify the print colour of neutral words (e.g., "find" printed in red; find_red). Unbeknownst to participants, each word is presented most often in one colour (e.g., "find" most often in red, "list" most often in green, etc.). Participants implicitly learn these contingencies very rapidly (Schmidt, De Houwer, & Besner, 2010), as indicated by faster responses to *high contingency* trials (where the word accurately predicts the response) relative to *low contingency* trials (where the word does not correctly predict the response). Interestingly, participants generally have very low levels of contingency awareness in this task (Schmidt & De Houwer, 2012a, 2012b, 2012c, 2012d), similar to other tasks (e.g., Destrebecqz & Cleeremans, 2001; Lewicki, 1985; McKelvie, 1987; Miller, 1987).

The present investigation aims to better understand the mechanisms underlying contingency learning in this sort of "implicit" performance paradigm by studying performance benefits (facilitation) and costs (interference). For instance, how much of a performance benefit is observed when the word accurately predicts the correct response (e.g., when "find" is presented in the high-contingency red colour)? Similarly, how much of a performance cost is observed when the word *incorrectly* predicts a response (e.g., when "find" is presented in a colour other than red)? As we will discuss, different mechanistic accounts of how learned contingencies are used to anticipate responses make different predictions about when benefits and costs should be observed.

Schmidt and Besner (2008) tested for facilitation and interference in a two-choice contingency learning experiment. One neutral word was presented most often in one colour, another neutral word most often in the other colour, and a third neutral word was presented equally often in both colours (medium contingency). They found that high contingency trials were responded to faster than medium contingency trials, but there was no difference between medium and low contingency trials. They took these results to suggest that contingency learning is inherently facilitative. More specifically, they propose that contingency knowledge is used to decrease the response threshold of the predicted response. This makes selecting the predicted (i.e., high contingency) response easier. Selecting an unpredicted (i.e., low contingency) response is not impaired, because the response threshold is not adjusted for unpredicted responses. Thus, low contingency trial performance should be equivalent to medium (i.e., chance) contingency trials, where no prediction is made at all. However, although power was high to detect a relatively small interference effect (6 ms), the two-choice task used by Schmidt and Besner did not generate very large contingency effects overall (e.g., the difference between high and low contingency trials was only 8 ms). Thus, it is possible that an interference effect would emerge in a task variant that produced larger contingency effects.

Related to this, data from a Stroop congruency experiment of Hazeltine and Mordkoff (2014) seem to suggest both facilitative effects of high contingency trials and interfering effects on low contingency trials. Unfortunately, explicit pairwise comparisons between high, low, and medium contingency trials were not reported. Similarly, Carlson and Flowers (1996) report a flanker contingency paradigm that included high, medium, and low contingency trials. The flanker contingency paradigm is similar to the colour-word contingency paradigm, except letters (rather than colours) serve as the target stimuli, and the predictive stimuli are "flanking" letters (or other stimuli). Pairwise comparisons were not reported for high versus

medium or for medium versus low contingency trials, but their figures seem suggestive of both facilitation and interference.

Another experiment by Miller (1987, Experiment 1), also with a flanker contingency paradigm, provides clearer evidence for both facilitation and interference, albeit with a different sort of "neutral" condition. Specifically, flankers were either strongly predictive (92% in a two-choice task) or weakly predictive (58%). Responses on strongly-predictive high contingency trials were faster than those on weakly-predictive high contingency trials, suggesting facilitation. Similarly, responses on low contingency trials with strongly-predictive flankers were slower than those on low contingency trials with weakly-predictive flankers. Though the flanker contingency and colour-word contingency paradigms have some surface differences, this latter finding suggests an interference effect for low contingency trials might exist. Thus, one goal of the present manuscript is to assess whether low contingency trials in the colour-word contingency learning paradigm do produce interference relative to a medium contingency control.

In investigating facilitation and interference, however, one important consideration is what constitutes a "neutral" or medium contingency baseline. We consider three different types of medium contingency trials, which will allow us to differentiate various theoretical positions (see below). The frequency of word-colour combinations used are presented in Table 1. Two of the words (e.g., "give" and "hear") are presented most often (60% of the time) in one colour, very rarely (6.7%) in a second colour, and in an intermediate frequency (33.3%) in a third colour. The remaining word (e.g., "make") is presented equally often (33.3%) in all three colours.

**(Table 1)**

These manipulations create five unique trial types, as can be seen in Table 2. On *high contingency* trials, the word is presented in its most frequent colour (e.g., give$_{purple}$). On *low*

*contingency* trials, the word is presented in its *least* frequent colour (e.g., give$_{orange}$). Critically, the manipulations also allow for three types of *medium contingency* trials. On *biased-word* trials, a word that is usually predictive of one high contingency response is presented in a medium contingency colour. For instance, this would be the case for give$_{grey}$ if "give" is presented most often in purple and if grey is presented with all words at a medium ("chance") proportion. On *biased-colour* trials, the word is unpredictive of the correct response, but the colour is most often associated with a particular word. For instance, this would be the case for make$_{purple}$ if purple is most commonly presented with the word "give" and the word "make" is unpredictive of the colour. Finally, on *unbiased* trials, neither the word nor the colour are predictive of any other stimulus. For instance, make$_{grey}$ is unbiased because (a) "make" is unpredictive of what colour will be presented, and (b) grey is unpredictive of what word will be presented.

**(Table 2)**

We consider four possible accounts of how contingency knowledge is retrieved and impacts responding, each presented in a separate panel of Figure 1. Complimentary to this, Figure 2 presents the pattern of results that we should expect for each of these four accounts. The first we call the *prediction benefit* account (e.g., Schmidt & Besner, 2008), displayed in Figure 1a. According to this account, a response is anticipated if one of the potential responses is highly likely, and accurate response prediction benefits performance. As such, high contingency trials will be faster than all other trials. Critically, the prediction benefit account assumes that predicting a response does not impair the ability of the system to make any of the remaining responses (i.e., the predicted response does not compete with the non-predicted responses). For instance, there will be no performance cost on a low contingency trial resulting from incorrect response prediction. Thus, there will be no differences between the low, biased-word, biased-colour, and unbiased trial types, as illustrated in Figure 2a.

**(Figures 1 and 2)**

Another contender, displayed in Figure 1b, we call the *misprediction cost* account. According to this account, if the distracting word is strongly predictive of one response, then (a) making that response will be facilitated (i.e., as in the prediction benefit account), and (b) making any other response will be impaired via response competition. That is, contingency information is used to activate the anticipated (i.e., high contingency) response, and this activated response then competes with all other contending responses. For instance, if participants see give$_{orange}$ (low contingency), the word "give" activates the (high contingency) purple response. This activated purple response then competes with the correct orange response, for instance, via mutual inhibition between the various response representations. Both low contingency and biased-word trials should be slowed by this sort of interference, because the words in these conditions are predictive of a (high contingency) colour response that is *different* from the correct response. On biased-colour and unbiased trials, however, the words are unpredictive (e.g., "make" is presented equally often with all colour responses). Thus, the word will not activate a response that the colour needs to compete with and performance will be unimpaired. This predicted pattern of means is presented in Figure 2b.

A third possibility, displayed in Figure 1c, we call the *bidirectional cost* account. This account is identical to the misprediction cost account, except that it is additionally assumed that it is harder to make a colour response that is frequently associated with a specific word that is *not* present on the current trial. For instance, if "give" is presented most often in purple, then participants might come to expect both (a) "give" will be followed by a purple response, and (b) a purple response will tend to be *preceded* by "give." Thus, participants will be hesitant to make a purple response if they do not see the (expected) word "give," resulting in an impairment on biased-colour trials. For instance, the purple response might be inhibited when the word "give" is not detected, as illustrated in Figure 1c. The predicted pattern of

results for this account are presented in Figure 2c.

Finally, a fourth possibility, displayed in Figure 1d, we call the *pure proportion* account. Unlike the preceding three accounts, the pure proportion account suggests that response time will be determined by the proportion with which a given distracting word is presented with a given response. For instance, each response might be biased proportionally to the proportion episodes that are retrieved from memory in which the presented word co-occurred with said response. Thus, the high contingency response will be strongly biased (because most episodes will point to this response), so high contingency responses will be very quick. The low contingency response will be only very weakly activated (because very few episodes point to this response), so low contingency responses will be slow. Medium contingency responses will be activated at an intermediate level (because there are an intermediate proportion of episodes pointing to this response), resulting in intermediate speed medium contingency responses. Indeed, because the word-response contingencies for biased-colour, biased-word, and unbiased trials are all identical (i.e., the word is 33.3% predictive of the correct response in all cases), response time on these three trial types should be roughly identical. The predicted pattern of results is presented in Figure 2d. We tested these predictions in Experiment 1.

## Experiment 1

**Method**

**Participants.** Fifty-one Ghent University undergraduates participated in the study in exchange for €5.

**Apparatus.** Stimulus and response timing were controlled by E-Prime 2 software (Psychology Software Tools, Pittsburgh, PA). Responses were recorded on an AZERTY keyboard with the "J," "K," and "L" keys using the index, middle, and ring fingers of the right hand for the colours purple, orange, and grey, respectively.

**Design.** The stimuli of the experiment consisted of the Dutch words "geef" ("give"), "hoor" ("hear"), and "maak" ("make") presented in the purple (128,0,128), orange (255,165,0), and grey (192,192,192), corresponding to "purple," "orange," and "silver" in the standard E-Prime colour palate. Two of the words were presented 60% of the time (9 of every 15 presentations) in one colour, 6.7% of the time (1 of every 15 presentations) in a second colour, and 33.3% of the time (5 of every 15 presentations) in the third colour. The third word was presented equally often (i.e., 33.3%) in all colours. Which words were presented with which frequency in each colour was randomly counterbalanced across participants. Although colour-to-key mappings were identical for all participants, which colour was not associated with a high contingency word was counterbalanced. The resulting three counterbalancing orders were run in random orders in sets of three participants. In the main experiment, there were 400 trials, selected at random with replacement (see how these 400 trials divide across stimuli in Table 1). The main part of the experiment was preceded by a practice phase. This consisted of the stimulus "@@@@" presented in each of the 3 colours 8 times each, for a total of 24 trials, presented in a random order (without replacement).

**Procedure.** Stimuli were presented in bold, 18 pt. Courier New font in the center of a black screen (0,0,0). Each trial in the experiment consisted of three sequential events. First, a white (255,255,255) fixation cross ("+") was presented for 150 ms. Second, a black screen was presented for another 150 ms. Third, the stimulus was presented until either a response was made or 1500 ms had elapsed. The next trial began immediately if the response was correct. If participants indicated the wrong response or failed to respond in 1500 ms, the stimulus "XXX" was presented in white for 1000 ms before the next trial began.

**Data Analysis.** Correct response times and percentage errors were analysed. Practice trials were not analysed. For the main part of the experiment, all trials on which participants failed to respond were excluded from analyses (0.4% of the data). The counterbalancing

factor was included as a factor in all analyses. This was done because average response speed varies from finger to finger (e.g., Hayes & Halpin, 1978) and which response key was the non-contingent response key varied according to counterbalancing (also true of display colour). This represents noise, which can be controlled for by adding the counterbalancing factor to the ANOVA. We do not discussed the results of this factor below (i.e., as it is orthogonal to the contrasts that we are actually interested in).[1]

**Results**

    **Response Times.** The response time data are presented in Figure 3. The five contingency conditions (high, low, biased-word, biased-colour, unbiased) were submitted to an ANOVA. Most importantly, this produced a significant main effect of condition, $F(4,192) = 6.636$, $MSE = 844$, $p < .001$, $\eta_p^2 = .12$. As several of our contrasts of interest concern the three medium contingency trial types, a second ANOVA was conducted with only these three conditions retained (biased-word, biased-colour, unbiased). Interestingly, this did *not* produce a significant effect of condition, $F(2,96) = .011$, $MSE = 672$, $p = .989$, $\eta_p^2 < .01$. More specifically, there was no difference between biased-word and unbiased trials, $F(1,48) < .001$, $MSE = 602$, $p = .984$, $\eta_p^2 < .01$, contrary to what the misprediction cost and bidirectional cost accounts would predict. There was also no difference between biased-colour and unbiased trials, $F(1,48) = .011$, $MSE = 846$, $p = .917$, $\eta_p^2 < .01$, contrary to what the bidirectional cost account would predict. Indeed, mean RT in these three conditions all differed by less than 1 ms, consistent only with the prediction benefit and pure proportion accounts. Note that power was high (.8) to detect an effect as small as 14 ms and 16 ms for the last two comparisons.

<div align="center">

**(Figure 3)**

</div>

    For the remaining analyses, the biased-word, biased-colour, unbiased conditions were collapsed into one medium contingency condition. We then compared high, medium, and low contingency trials. Unsurprisingly, high contingency trials were responded to significantly

faster (529 ms) than low contingency trials (558 ms), $F(1,48) = 24.139$, $MSE = 856$, $p < .001$,

$\eta_p^2 = .33$. Interestingly, medium contingency trials (538 ms) were responded to significantly

slower than high contingency trials, $F(1,48) = 8.136$, $MSE = 226$, $p = .006$, $\eta_p^2 = .14$, and

significantly faster than low contingency trials, $F(1,48) = 8.964$, $MSE = 1134$, $p = .004$, $\eta_p^2 =$

.16. This final comparison is compatible with the pure proportion account, but inconsistent

with the prediction benefit account.

As a supplementary analysis, we further divided the data into two blocks (200 trials

each) to test for any changes in contingency effects over time. The two blocks (Block 1, Block

2) by five contingency conditions (high, low, biased-word, biased-colour, unbiased) were

added to an ANOVA. Critically, there was no interaction between contingency and block,

$F(4,192) = .780$, $MSE = 1770$, $p = .539$, $\eta_p^2 = .02$. As with our previous reports, these results

suggest that contingencies are learned quickly and remain relatively stable throughout the task

(Schmidt et al., 2007, 2010; Schmidt & De Houwer, 2012b, 2012d).

**Percentage Errors.** The percentage error data are presented in Figure 4. The five

contingency conditions (high, low, biased-word, biased-colour, unbiased) were submitted to

an ANOVA. Most importantly, this produced a significant main effect of condition, $F(4,192)$

$= 4.643$, $MSE = 18.2$, $p < .001$, $\eta_p^2 = .09$. As in the analyses of the RT data, a second ANOVA

with only the three medium contingency conditions (biased-word, biased-colour, unbiased)

did *not* produce a significant effect of condition, $F(2,96) = .321$, $MSE = 9.1$, $p = .726$, $\eta_p^2 <$

.01. More specifically, there was no difference between biased-word and unbiased trials,

$F(1,48) = .721$, $MSE = 7.8$, $p = .400$, $\eta_p^2 = .01$, contrary to what the misprediction cost and

bidirectional cost accounts would predict. There was also no difference between biased-colour

and unbiased trials, $F(1,48) = .058$, $MSE = 10.8$, $p = .811$, $\eta_p^2 < .01$, contrary to what the

bidirectional cost account would predict. Indeed, the numerical differences between

conditions were small (<0.5%) and in the opposite direction to those predicted by the

misprediction cost and bidirectional cost accounts. Note that power was high (.8) to detect an effect as small as 1.6% and 1.9% for the last two comparisons.

**(Figure 4)**

For the remaining analyses, the biased-word, biased-colour, unbiased conditions were collapsed into one medium contingency condition. We then compared high, medium, and low contingency trials. Unsurprisingly, there were significantly less errors on high contingency trials (5.0%) than low contingency trials (8.4%), $F(1,48) = 9.391$, $MSE = 30.5$, $p = .004$, $\eta_p^2 = .16$. Interestingly, medium contingency trials (5.8%) produced significantly more errors than high contingency trials, $F(1,48) = 5.634$, $MSE = 3.1$, $p = .022$, $\eta_p^2 = .11$, and significantly less errors than low contingency trials, $F(1,48) = 5.410$, $MSE = 30.2$, $p = .024$, $\eta_p^2 = .10$.

We further analysed the error data for block effects. The two blocks (Block 1, Block 2) by five contingency conditions (high, low, biased-word, biased-colour, unbiased) were added to an ANOVA. Critically, there was no interaction between contingency and block, $F(4,192) = .299$, $MSE = 25.9$, $p = .878$, $\eta_p^2 < .01$. Again, this suggests that contingencies are learned quickly and remain relatively stable throughout the task.

**Discussion**

The results of Experiment 1 were inconsistent with the prediction benefit, misprediction cost, and bidirectional cost accounts. In particular, we found (a) slower and more error-prone responses to low relative to medium contingency trials, inconsistent with the former account, and (b) no differences between the three medium contingency conditions, inconsistent with the latter two accounts. The results were most consistent with the pure proportion account, which proposes that presentation of a word leads to retrieval activation of all responses, each proportional to the proportion of co-occurrences of that word with that response.

**Experiment 2**

In Experiment 1, performance on biased word trials such as give$_{grey}$, did not differ from performance on unbiased trials, such as make$_{grey}$. Because response competition should have been stronger on biased word trials (i.e., competition between the correct response and the high contingency response) than on unbiased trials (see Table 1 and Figure 1), this finding argues against the idea that response competition underlies performance in our task. On the other hand, it might be argued that on unbiased trials, a correct response (e.g., grey) has to compete with two medium strength responses, which could slow down performance as much as a competition with one strongly activated response (as on biased word trials). Hence, the results of Experiment 1 are not conclusive regarding the contribution of response competition.

Experiment 2 aims to provide a clearer test of whether interference between retrieved responses occurs. For this, we used a much simpler design during an initial *learning phase*, presented in Table 3. Specifically, we dropped the medium contingency conditions from the design. This was to strengthen the contingency manipulation for the remaining high and low contingency trials. In a second *test phase*, the design remained the same except for the addition of novel *neutral* word trials. These neutral words were never presented before the neutral word trial on which they appeared. Because the neutral words were not previously paired with colors, response activation and thus response competition should be minimal on neutral word trials. If interference between retrieved responses *does* occur, then we should expect that low contingency trials will be responded to more slowly that the neutral trials. That is, on a low contingency trial, such as give$_{orange}$, the "high contingency" purple response should interfere with making the correct orange response. On neutral trials, this interference will not occur.

**(Table 3)**

On the other hand, if interference does *not* occur, then low contingency trials might

actually be responded to *faster* than neutral trials. For instance, on a trial such as give$_{orange}$

participants might be particularly biased toward a purple response, but they will also be

partially prepared for an orange response. The word "give" and the colour orange already co-

occurred on several trials, and this will produce a benefit (albeit small) according to the pure

proportion account. In contrast, the word and colour are not previously experienced on a

neutral trial, so none of the responses will be activated by the word. Errors, on the other hand,

should be more likely in the low contingency condition than the neutral condition, because an

incorrect high contingency response will be activated by the word, occasionally strong enough

to result in selection of the incorrect response (see Schmidt & Besner, 2008).

**Method**

      **Participants.** Fifty-one Ghent University undergraduates participated in the study in

exchange for €5.

      **Design.** Experiment 2 had three phases. The first was the same *practice phase* used in

Experiment 1. This was followed by a *learning phase*. Out of a list of 33 four letter, first

person Dutch verbs presented in Table 4, 3 were randomly chosen for the learning phase for

each participant. Each of these three words was presented 40 times in one colour, and 5 times

in each of the remaining two colours (80% contingency). Thus, there were 150 trials total,

presented randomly without replacement. The final *test phase* was identical to the learning

phase, except that 30 neutral trials were intermixed in the procedure (180 trials total). These

neutral trials consisted of the remaining 30 verbs, each presented only once (10 in each

colour). Which neutral words were presented in which colour was also randomly determined

on a participant-by-participant basis. The total experiment contained 354 trials (24 practice +

150 learning + 180 test).

<div align="center">**(Table 4)**</div>

      **Apparatus, Procedure, and Data Analysis.** The apparatus and procedure of

Experiment 2 were identical to Experiment 1 in all respects. The data analysis was also

identical with the exception that there was no counterbalancing factor to add to the ANOVA.

**Results**

   **Response times.** The response time data are presented in Figure 5. In the initial

learning phase of the experiment, responding was significantly faster on high contingency

(520 ms) relative to low contingency trials (550 ms), $t(50) = 5.851$, $SE_{diff} = 5$, $p < .001$, $\eta_p^2 =$

.41. For the following test block, we first tested the one-way ANOVA for the condition factor

(high vs. low vs. neutral), which was significant, $F(2,100) = 20.553$, $MSE = 890$, $p < .001$, $\eta_p^2$

$= .29$. Decomposing this effect, we observed significantly faster responses to high

contingency trials (534 ms) relative to both low contingency (560 ms), $t(50) = 4.874$, $SE_{diff} =$

5, $p < .001$, $\eta_p^2 = .32$, and neutral (571 ms), $t(50) = 6.253$, $SE_{diff} = 6$, $p < .001$, $\eta_p^2 = .44$. Most

critically, low contingency responses were numerically *faster* than neutral, albeit not

significantly so, $t(50) = 1.604$, $SE_{diff} = 6$, $p = .115$, $\eta_p^2 = .05$.

**(Figure 5)**

   **Error percentages.** The percentage error data are presented in Figure 6. In the

learning phase, there were significantly less errors on high contingency (2.8%) relative to low

contingency trials (5.5%), $t(50) = 2.767$, $SE_{diff} = 0.9$, $p = .008$, $\eta_p^2 = .13$. For the following test

block, the ANOVA for condition (high vs. low vs. neutral) was significant, $F(2,100) = 12.715$,

$MSE = 20.4$, $p < .001$, $\eta_p^2 = .20$. Decomposing this effect, we observed significantly less errors

to high contingency trials (5.0%) relative to both low contingency (9.5%), $t(50) = 4.933$, $SE_{diff}$

$= 0.9$, $p < .001$, $\eta_p^2 = .33$, and neutral (6.9%), $t(50) = 2.372$, $SE_{diff} = 0.8$, $p = .022$, $\eta_p^2 = .10$.

Importantly, there were more errors to low contingency relative to neutral trials, $t(50) = 2.688$,

$SE_{diff} = 1.0$, $p = .010$, $\eta_p^2 = .13$. As a final note, 62.5% of errors on low contingency trials were

the high contingency response (i.e., rather than the other low contingency response), and this

is a rate greater than chance (i.e., 50%), $t(43) = 2.474$, $p = .017$. This supports the notion that

increased low contingency trial errors are due to accidental selection of a high contingency response (see Schmidt & Besner, 2008).

**(Figure 6)**

**Discussion**

Contrary to the idea that response competition underlies performance, responses tended to be faster on low contingency trials relative to neutral trials. Note that while the pure proportion account does predict a facilitation effect, it implies that the effect should be very small. With proportional retrieval, low contingency responses will not receive much activation. For each eight presentations, a word will only be presented once (10%) with a given low contingency response. As such, if the contingency effect scales (roughly) with proportion, then the difference between low (10%) and neutral (0%) should be around seven times smaller than the difference between high (80%) and low (10%). Finally, there were more errors to low contingency relative to neutral trials. This finding is consistent with a proportion account irrespective of whether response competition operates also.

**General Discussion**

We aimed to better understand the processes involved in human contingency learning within the context of the colour-word contingency learning paradigm. Interestingly, in Experiment 1 we found that both speed of responding and accuracy directly tracked with the proportion that the distracting word was presented with the correct response. That is, high contingency trials were faster and more accurate than medium contingency trials, which were in turn faster and more accurate than low contingency trials. This pattern of results was only consistent with one of the four accounts discussed in the Introduction, namely, the pure proportion account.

The design of Experiment 1 further allowed for an investigation of differing types of medium contingency trials. However, we found that these distinctions did not matter. It did

not matter whether the word was strongly predictive of another (incorrect) response. In other words, there was no cost for making a response prediction that turned out to be incorrect, inconsistent with the misprediction cost account. It similarly did not matter if the correct colour response was strongly linked to a specific currently-unpresented stimulus word, inconsistent with the bidirectional cost account. That is, the current results do not support the idea that it is harder to make a given colour response in the absence of an expected distracting word. Together, these results suggest that the contingency mechanism responsible for producing this contingency effect biases each response roughly proportional to the extent that it co-occurs with the predictive cue (i.e., word), as proposed by the pure proportion account.

Experiment 2 further failed to find any evidence for response competition between the predicted and correct response. Responses to low contingency trials were actually numerically *faster* than responses to once-presented neutral words (though not significantly). Errors were increased in the low contingency condition relative to neutral, but this does not necessarily indicate response competition. Instead, this can be due to accidental selection of a highly activated (but incorrect) high contingency response on low contingency trials. An analysis of high versus low contingency errors supported this proposition.

**Episodic Learning**

These results are partially consistent with the episodic learning account of Logan (1988), who argued that responding to a stimulus will be roughly negatively proportional to the frequency that the stimulus is presented. The more specific claim is that the processing and/or encoding of a stimulus into episodic memory will become easier with repeated presentation. However, it is known that effects in the colour-word contingency learning paradigm are driven by word-response contingencies and not word-colour (i.e., stimulus) frequencies (e.g., Schmidt et al., 2007; Schmidt & De Houwer, 2012b; see also Miller, 1987). Thus, we suggest that it is not the *encoding* or *stimulus processing* of frequent events that aids

performance, but instead the *retrieval* of frequent word-response pairings.

The current results can also be explained via a simple episodic retrieval mechanism. Specifically, participants store episodic memories of each trial that they have experienced, then retrieve these episodes on subsequent trials. For instance, if the word "give" is presented, then episodes representing trials in which "give" was presented will be retrieved and used to anticipate the correct response. What the current results do *not* suggest is that participants generate a single prediction corresponding to the high contingency response. Instead, the results suggest that each potential response is biased proportionally to the proportion of episodes pointing to that response (e.g., see Figure 1d). For instance, the stimulus "give" in Table 1 will strongly bias a purple response, because this is the most frequent response. The grey response will also be moderately activated, because roughly a third of the episodes will point to this response. Finally, the orange response will receive very little activation at all, given that very few "give" nodes will point to the orange response.

**Implications and Future Research**

The present results suggest further caveats for scenarios in which contingency learning biases represent an unintended confound. For instance, this is often the case in Stroop experiments in which colour words (e.g., "blue") are sometimes presented more often in the congruent colour (i.e., blue) than in any of the individual incongruent colours (e.g., red, green, and yellow; see Schmidt, 2013, 2014; Schmidt & Besner, 2008; see also, Mordkoff, 1996). If one does *not* intend to study contingency learning and wishes to eliminate contingency biases from the design, then it is not only important to ensure that distracting stimuli (e.g., words) do not predict a single (high contingency) response, but also that all distracting stimuli are presented equally often with all responses. A low-frequency word-response combination will be responded to slower than a medium-frequency word-response combination, even if words are not predictive of a specific response in both cases.

Future research may also aim to distinguish between learning based on (a) the *proportion* (or contingency) with which a given distracter is presented with a given response and (b) the absolute *frequency* (i.e., number of occurrences) with which a given distracter and response co-occurred. That is, we propose that participants use the word to anticipate the most probable response based on the contingencies between a word and the responses (e.g., "give" predicts purple, because "give" is presented 60% of the time in purple). However, participants might alternatively/additionally improve performance each time they see a given word with a given response. For instance, responding is faster to give$_{purple}$ than give$_{orange}$ because a participant has made the purple response to "give," say, 63 times, but have only made the orange response to "give" 7 times. In the present paradigm (and in all our past work with this task), the absolute frequency of word-response pairings and the contingencies between words and responses were completely confounded. This could be changed in future research to distinguish between our pure proportion account and a "pure frequency" account. For instance, the number of occurrences of sets of words can be manipulated independently from the contingencies between words and responses.

Future research might also investigate the relation between the present results and findings in the S-R binding literature. Participants are faster to make the same rather than different response to a repeated target stimulus, and are *slower* to make a different response to repeated rather than non-repeated target stimulus (Hommel, 1998). This is argued to occur because on the second presentation of the target, the response previously made to that stimulus is automatically retrieved. Critically, these benefits and costs are also observed for *distracting* stimulus repetitions (Frings, Rothermund, & Wentura, 2007; see also, Giesen & Rothermund, in press; Rothermund, Wentura, & De Houwer, 2005). That is, responses are faster when making the same rather than a different response to a repeated distracter, and are *slower* when making a different response to repeated rather than non-repeated distracter. The

latter of these two findings might suggest that there *is* competition between a retrieved response and other potential responses, contrary to what we observe in our Experiment 2.

A response retrieved on the basis of recent S-R binding is typically argued to be due to the same process as a response retrieved on the basis of a repeatedly-reinforced contingency, so any inconsistencies in the two literatures are problematic. We would argue, however, that our pure proportion account is consistent with S-R binding results. We propose that each response is biased proportionally to the proportion of retrieved episodes that point to it. It is also known that the most recently occurring episodes have the largest impact on retrieval (e.g., Schmidt et al., 2010). Thus, a recent presentation of a given distracter-response pairing will shift retrieval in favor of the just-made response. Because retrieval is proportional, an increased retrieval of the just-made response entails a relative *decrease* in retrieval of the remaining responses. That is to say, even without response competition between the retrieved and remaining responses, our account still predicts slower (impaired) performance when making a different response to a repeated rather than non-repeated distracter. This might be described as *retrieval interference*, different from interference via *response competition*. Future work in the S-R binding literature might investigate this notion more directly.

**References**

Allan, L. G. (2005). Learning of contingent relationships. *Learning & Behavior, 33,* 127–129.

Beckers, T., De Houwer, J., & Matute, H. (2007). Editorial: Human contingency learning. *Quarterly Journal of Experimental Psychology*, 60, 289–290.

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

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review, 8,* 343–350.

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

Giesen, C., & Rothermund, K. (in press). Adapting to stimulus-response contingencies without noticing them. *Journal of Experimental Psychology: Human Perception and Performance*.

Hayes, V., & Halpin, G. (1978). Reaction time of the fingers with responses measured on a typewriter keyboard. *Perceptual and Motor Skills, 47,* 863–867.

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

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition, 5,* 183–216.

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

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

McKelvie, S. J. (1987). Learning and awareness in the Hebb digits task. *Journal of General Psychology, 114,* 75–88.

Miller, J. (1987). Priming is not necessary for selective-attention failures: Semantic effects of

unattended, unprimed letters. *Perception & Psychophysics, 41,* 419–434.

Mordkoff, J. T. (1996). Selective attention and internal constraints: There is more to the

flanker effect than biased contingencies. In A. Kramer, M. G. H. Coles, & G. Logan

(Eds.), *Converging operations in the study of visual selective attention* (pp. 483–502).

Washington, DC: APA.

Musen, G., & Squire, L. R. (1993). Implicit learning of color-word associations using a

Stroop paradigm. *Journal of Experimental Psychology: Learning, Memory, &

Cognition, 19,* 789–798.

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. (2013). 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. (2014). Contingencies and attentional capture: The importance of matching

stimulus informativeness in the item-specific proportion congruent task. *Frontiers in

Psychology, 5,* Article 540.

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

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology*, *61,* 273–301.

**Footnote**

[1]     In all our previous work with the colour-word contingency learning paradigm it was always the case, for every participant, that every word, colour, and response contributed in equal proportions to the high, low, and (if present) medium contingency conditions. Thus, any differences in response speed with different keys or different stimuli were balanced within each participant. In the novel design of the current experiment this was no longer the case. Note that while the main effect of counterbalancing was not significant for any test, counterbalancing order *did* interact with contingency for several tests. This is particularly the case in the response time data, where differences in response speed with different fingers is a concern. Counterbalancing had minimal impact on errors, even though the contingency effects in the errors were just as robust as those in the response times.

*Table 1.* Example Experiment 1 stimulus pairings (of 400), with word-colour co-occurrence fractions in brackets.

| Colours | Responses | Words | | |
|---|---|---|---|---|
| | | give | hear | make |
| purple | J | 80 | 8.9̇ | 44.4̇ |
| | | (9/15) | (1/15) | (5/15) |
| orange | K | 8.9̇ | 80 | 44.4̇ |
| | | (1/15) | (9/15) | (5/15) |
| grey | L | 44.4̇ | 44.4̇ | 44.4̇ |
| | | (5/15) | (5/15) | (5/15) |

*Table 2.* Experiment 1 conditions.

| Condition Name | Example Stimulus | Correct Response | Word Predicts | Colour Predicts | Stimulus Frequency |
|---|---|---|---|---|---|
| high | give$_{purple}$ | J key | purple (J key) | give | 80/400 |
| low | give$_{orange}$ | K key | purple (J key) | hear | 8.9/400 |
| biased-word | give$_{grey}$ | L key | purple (J key) | - | 44.4/400 |
| biased-colour | make$_{purple}$ | J key | - | give | 44.4/400 |
| unbiased | make$_{grey}$ | L key | - | - | 44.4/400 |

*Table 3.* Example Experiment 2 learning phase pairings (of 150), with word-colour co-occurrence fractions in brackets.

| Colours | Responses | Words | | |
|---------|-----------|-------|------|--------|
|         |           | think | find | search |
| purple  | J         | 40 (8/10) | 5 (1/10) | 5 (1/10) |
| orange  | K         | 5 (1/10) | 40 (8/10) | 5 (1/10) |
| grey    | L         | 5 (1/10) | 5 (1/10) | 40 (8/10) |

*Table 4.* Experiment 2 stimuli with English translations.

| Dutch | English | Dutch | English | Dutch | English |
|-------|---------|-------|---------|-------|---------|
| bied | offer | jaag | hunt | trek | pull |
| bijt | bite | kies | choose | vang | catch |
| bind | tie | lach | laugh | vind | find |
| buig | bow | lieg | lie | weeg | weigh |
| denk | think | loop | run | werp | throw |
| duik | dive | maak | make | wijt | blame |
| geef | give | neem | take | word | become |
| giet | pour | raad | guess | zend | send |
| hang | hang | rijd | drive | zoek | search |
| help | help | roep | call | zuig | suck |
| hoor | hear | ruik | smell | zwem | swim |

**Figures**

*Figure 1.* Four potential accounts of contingency learning: (a) prediction benefit, (b) misprediction cost, (c) bidirectional cost, and (d) pure proportion. *Note:* The darkness of nodes and thickness of lines indicates strength of activation. Lines ending with arrows represent activation, and lines ending with dots indicate interference.

*Figure 2.* Predicted outcomes of the four accounts in Figure 1: (a) prediction benefit, (b) misprediction cost, (c) bidirectional cost, and (d) pure proportion.

*Figure 3.* Experiment 1 response time data (in milliseconds).

*Figure 4.* Experiment 1 percentage error data.

*Figure 5.* Experiment 2 response time data (in milliseconds).

*Figure 6.* Experiment 2 percentage error data.

**(a)**



**(b)**



**(c)**



**(d)**