# Two sources of color-word contingency learning:

# Episodic retrieval of SR bindings and propositional knowledge

Matthäus Rudolph & Klaus Rothermund

Friedrich Schiller University Jena, Jena, Germany

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, in press

## Abstract

There is an ongoing debate about the cognitive mechanisms behind human contingency learning (CL). While, in some studies, episodic retrieval of previous responses fully explained the observed CL effects (Giesen et al., 2020; Schmidt et al., 2020), other findings suggest that global contingencies have an additional effect on behavior (Xu & Mordkoff, 2020). In a high-powered (*N* = 500), pre-registered study, we investigated CL effects after controlling for episodic retrieval of distractor-target (S-S) and distractor-response (S-R) bindings. Retrieval explained a large part of the CL effect. However, we still found a reliable residual CL effect even after controlling for retrieval. Notably, the residual CL effect depended on contingency awareness: The residual CL effect only occurred for trials for which participants correctly detected the respective color-word contingency, whereas for trials without contingency awareness there was no residual CL effect. Collectively, our findings suggest that human contingency learning is driven by two independent sources: (1) episodic retrieval of S-S and S-R bindings, and (2) propositional knowledge of the contingencies.


Keywords*: Contingency learning, contingency awareness, episodic retrieval, propositional learning*

Intelligent organisms must acquire knowledge about the regularities (contingencies) between events, actions, and their consequences to understand and interact proficiently with their environment (Allan, 2005; De Houwer & Hughes, 2017; Garner, 2014; Shanks, 2010). One useful paradigm to study human contingency learning is the color-word contingency learning paradigm (Schmidt et al., 2007): Participants have to classify the color of neutral adjectives, which are presented more often in some colors (high contingency combinations) and less often in other colors (low contingency combinations). Irrespective of the fact that the words are irrelevant for the task, the contingencies between words and colors influence performance, resulting in a global contingency learning (CL) effect, that is, faster and more accurate responses in high vs. low contingency trials.

The original explanation for CL effects is that participants learn the relationship between the irrelevant attribute and the correct response (Miller, 1987; Mordkoff, 1996; Schmidt et al., 2007; Schmidt & De Houwer, 2016a, 2016b). Following the principles of classical learning theory, people learn this correlation through associative learning (Pavlov, 1927; Rescorla, 1972). Essentially, the associative learning account postulates the automatic formation of associative links between mental representations of stimuli and responses through repeated exposure. This process is automatic in a sense that it operates unconsciously and requires no cognitive engagement (Bargh, 1994; De Houwer, 2009; McLaren et al., 1994). Once an associative link is formed, the presentation of a stimulus automatically activates the corresponding stimulus or response representation (i.e., "warm" *activates* the color "red" or the left-hand response), which then leads to faster responding in high compared to low contingency trials.

There is, however, another account that explains the CL effect in terms of recency-based episodic retrieval (Giesen et al., 2020). From this perspective, the CL effect results from (a) storing stimulus-response episodes in memory, and (b) retrieving the most-recent of these episodes when the stimulus is encountered again. By retrieval, the response which was bound to the stimulus during the last occasion becomes reactivated (i.e., "warm" *retrieves* the color "red" or the left-hand response).

Consequently, the CL effect is a result of automatically retrieving previously formed episodic stimulus-response bindings.

In CL studies, global contingencies are confounded with the color that was presented and the response that was performed during the most recent matching episode (Giesen et al., 2020; Schmidt et al., 2020): Since the last episode in which a stimulus occurred is most likely a high contingency trial, high contingency trials have a high probability that the same color/response is retrieved from the last episode, which facilitates responding, whereas for low contingency trials, the probability of retrieving a different color/response from the last episode in which the stimulus occurred is high, which interferes with responding. Although the confounding between contingency and retrieving the same color/response from the last matching episode is strong, it is not perfect. There are also some trials in which a different color/response is retrieved in a high frequency trial, and there are also trials in which the same response is retrieved in a low frequency trial.

To investigate whether contingencies have a unique effect on responding that is independent of episodic retrieval, both predictors (contingency, and response relation between the current and the last matching episode) have to be implemented within the same analysis to predict speed and accuracy of responding. In several studies, controlling for episodic retrieval fully eliminated the effect of global contingencies (Giesen et al., 2020; Schmidt et al., 2020). Based on these results, the authors concluded that recency rather than frequency is the guiding principle of contingency learning (the "law of recency", Giesen et al., 2012; see also Jiménez et al., 2022; Rothermund et al., 2022).

In a study by Xu and Mordkoff (2020), however, a significant CL effect was found even after controlling for episodic retrieval. This result seems to indicate that episodic retrieval alone is not sufficient to explain the overall CL effect. Yet, although a many-to-one mapping was used to assign six colors to two responses in their task, the authors did not control for effects of episodic S-S retrieval, which may have led to a biased estimation of the residual contingency effect. The difference in findings between these studies thus may be due to differences in design, however, it may also reflect a lack of sufficient power to detect a residual CL effect, that is, an effect of contingencies that influences responding over and above episodic retrieval[1].

Another relevant question regarding a possible residual CL effect is the nature of such an effect. If there is an effect of CL after controlling for recency-based episodic retrieval of S-S and S-R bindings, such an effect could reflect a genuine effect of associative learning resulting from the accumulated influence of all previous episodes. Alternatively, a residual CL effect could reflect propositional knowledge, that is, it could be based on awareness of the color-word contingencies, which is then used to govern responding. Previous research with the CL paradigm showed that awareness of the color-word contingencies amplified the CL effect (Schmidt & De Houwer, 2012), but CL effects were present also if awareness was absent (Miller, 1987; Schmidt et al., 2007). Similar findings of an influence of contingency awareness on CL effects have recently been reported also in related paradigms (Arunkumar, Rothermund, & Giesen, 2023; Arunkumar et al., 2022; Arunkumar, Rothermund, Kunde, et al., 2023; Rothermund et al., 2023).

These initial findings suggest that although contingency awareness is not a necessary precondition for CL to occur, it may still modulate the CL effect. Within these early studies investigating

---

[1] As the design of Xu and Mordkoff (2020) and their report of a residual, retrieval-independent CL effect inspired the present work, we hereby mention some features of their study which in our view limit the interpretability of their findings, and how we addressed them in the current experiment. Careful reanalysis of their data revealed that not all relevant combinations were realized in their design. For example, there were no low contingency trials that required the same response during the last occurrence. This is because, within each experimental block, every low contingency combination was presented exactly once, making it impossible that a low probability combination repeats. This renders an appropriate analysis of episodic retrieval effects difficult. To address this issue, we increased the number of trials per block and used random sets of stimuli that follow a given probability pattern (instead of fixed lists). While this can lead to small deviations from the given contingencies, it resulted in a sufficiently large number of trials for each contingency-response relation combination. Further, Xu and Mordkoff (2020) used letters (i.e., H and X) as distractors which are perceptually similar, which makes a control of episodic retrieval effects difficult and imperfect. To render retrieval unambiguous, we decided to use words with a unique identity as distractor stimuli. Further, we added a target category to the low contingency trials (i.e., low contingency trials appear in two colors instead of one color only), which allows us to separate S-S and S-R retrieval effects also for these trials. We thank Xu and Mordkoff for providing us with their data and thereby giving us the opportunity to adapt and further improve their experimental design.

the relationship of CL and contingency awareness (Miller, 1987; Schmidt et al., 2007; Schmidt & De Houwer, 2012), however, the role of episodic retrieval was not considered. Hence, in the current investigation, we aim to clarify whether the residual CL effect, that is, the effect that remains after controlling for episodic retrieval, is related to contingency awareness.

**The present study**

The goal of the present study is twofold: First, we wanted to clarify whether there is a CL effect after controlling for episodic retrieval of the most recent matching episode, given sufficient power, and after controlling for both, S-S and S-R retrieval. Second, if it exists, our aim was to examine the source of such a residual CL effect. Specifically, we investigated whether the strength of a residual CL effect is modulated by contingency awareness, and whether such an effect can be found even in the absence of contingency awareness, which would provide strong evidence for associative learning.

In our study, we adopted some of the design features of Xu and Mordkoff (2020): First, we used a many-to-one mapping, assigning two colors to each response key, which allows us to separately control for the effects of both, episodic S-S and S-R retrieval. Second, we induced a contingency of medium strength between the irrelevant attribute and the correct color (5:1:1:1), which should produce some variance regarding the awareness of these contingencies while at the same time providing a sufficiently strong contingency effect. We also used four distractor words and corresponding color-word contingencies (rather than just three as it was often done in previous studies; e.g., Schmidt et al., 2007), rendering the contingencies harder to detect, thus preventing ceiling effects in objective contingency awareness (Schmidt & De Houwer, 2012). To assess contingency awareness, we used a forced choice recognition task at the end of the experiment to determine awareness of color-word contingencies separately for each word within each participant (see Hirshman, 1995; Shanks & John, 1994; Stahl & Bading, 2020; Stahl et al., 2016).

### Hypotheses

First, we expect a robust CL effect, which is characterized through faster responding in high compared to low contingency trials. Second, we investigate whether the CL effect can be fully explained by the most recent occurrence of the distractor, from which the color or the response could be retrieved, or whether there is a residual effect of contingencies after controlling for retrieval. Third, we investigated whether such a residual CL effect depends on contingency awareness, and whether there is still a CL effect for those trials for which the color-word contingency was not detected.

## Method

### Ethics vote, pre-registration, and open access

Ethical approval was granted by the Ethics Committee of the FSU Jena (FSV 22/052). We conducted one Experiment in two waves of data collection. Prior to each data collection, the exact method, design, hypotheses, data preparation, and planned analyses were pre-registered online (wave 1: https://aspredicted.org/6qn6n.pdf; wave 2: https://aspredicted.org/3f9tg.pdf). All data and analyses scripts are available at https://osf.io/ubqfx/.

### Required sample size and a priori power calculations

Based on the results of the first wave of data collection ($N$ = 120), we conducted a power analysis using the *simr* package in *R*. The results indicated that we need a total of $N$ = 380 participants to identify a small interaction effect between contingency learning and contingency awareness (similar to the one that we observed in the first wave) with a power of *ß* = .80 in a hierarchical multilevel analysis (while controlling for episodic retrieval effects). We then collected data from an additional $N$ = 380 participants, but pre-registered the option to jointly analyze the data of the two waves to further increase power if there were no interactions with wave of data collection.

## Participants

A total of *N* = 122 participants took part in the first wave of data collection. One participant failed the attention check (more than 20% errors in the training block), the data of another participant could not be processed due to technical difficulties. This resulted in a sample size of 120 participants (48 male, 72 female) with a mean age of 27.29 years (*SD* = 5.26). Another *N* = 381 participants took part in the second wave of data collection. One participant did not complete the whole experiment and was excluded from the study. This resulted in a sample size of 380 participants (208 male, 172 female) with a mean age of 28.42 years (*SD* = 5.06). Taken together, *N* = 500 participants completed the experiment. All of them were recruited online via Prolific and received a reward of approximately 4 Pounds. All participants were pre-screened to be native English speakers, between 18 and 35 years old, running the experiment on a notebook or desktop computer and could only take part in the study once. All participants gave their informed consent prior to participation in the study. On average, participants spent 30 minutes on the study (*SD* = 9.60).

## Design

The experiment comprised a 2 (contingency: high vs. low frequency combination of color and word) x 3 (relation to the last occurrence of the same word: same target color/same response vs. different target color/same response vs. different target color/different response) repeated measures design, with both factors being manipulated within participants (see Figure 1 for an illustration of the factors of our experimental design). Reaction times (RT) functioned as the dependent variable of interest.

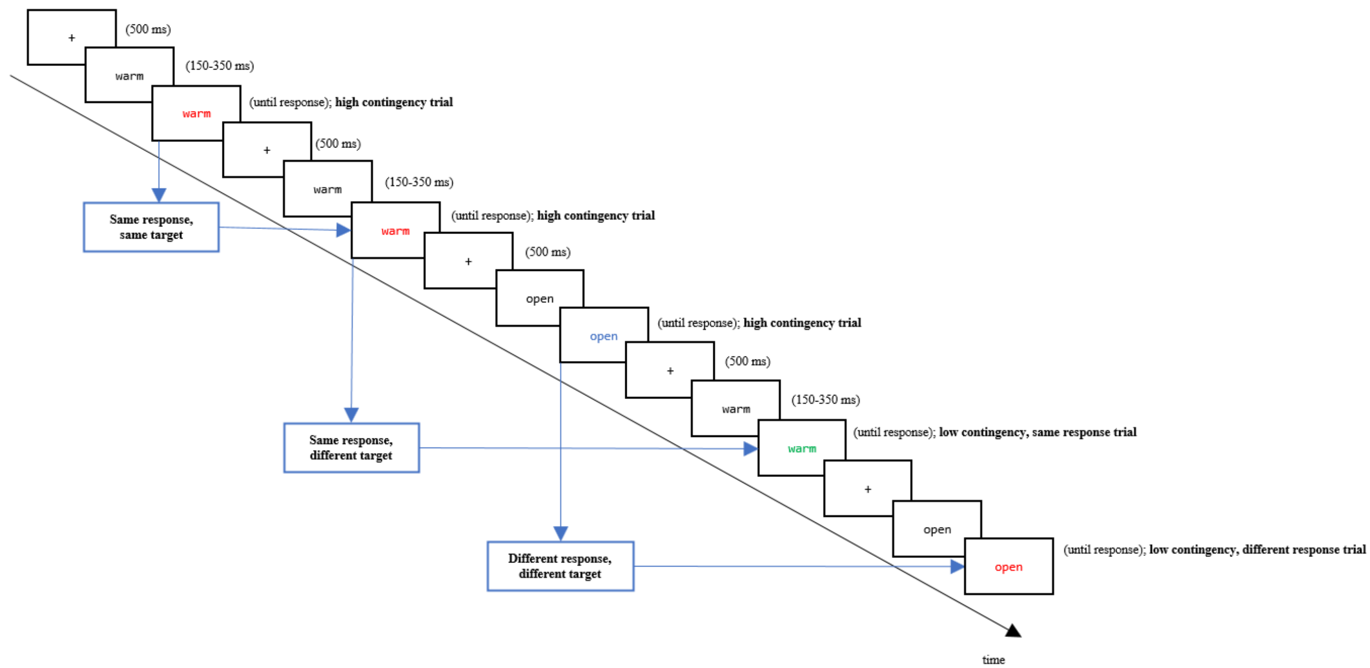


***Figure 1.*** Schematic trial procedure in the color classification task. Trials are classified as high contingency, low contingency/same response and low contingency/different response (baseline) trials. Arrows illustrate different types of response and target (color) relations between the current trial and the last occurrence of the same distractor word (same target color/same response, different target color/same response, different target color/different response).

To manipulate the contingency, we created three conditions depending on the frequency of the color-word combinations: high contingency trials (hc) and two types of low contingency trials (low contingency/same response [lc/sr] and low contingency/different response trials [lc/dr]). In the lc/sr condition, distractors are presented with target attributes which are assigned to the same response key as the high contingency color, whereas in the lc/dr condition the distractor word appears in a color that is assigned to the opposite key (baseline condition; see Figure 1, Table 1).

**Table 1**

*Color-word contingency manipulations in the color classification task*

| Target (color) | Correct response | Distractor (word) | | | |
|---|---|---|---|---|---|
| | | warm | quiet | open | clear |
| red | left | 5 (hc) | 1 (lc/sr) | 1 (lc/dr) | 1 (lc/dr) |
| green | left | 1 (lc/sr) | 5 (hc) | 1 (lc/dr) | 1 (lc/dr) |
| blue | right | 1 (lc/dr) | 1 (lc/dr) | 5 (hc) | 1 (lc/sr) |
| yellow | right | 1 (lc/dr) | 1 (lc/dr) | 1 (lc/sr) | 5 (hc) |

*Note.* hc: high contingency color-word combinations; lc/sr: low contingency color-word combinations sharing the response key with the hc trials; lc/dr: low contingency color-word combinations mapped to a different response key (baseline condition).

## Apparatus and Stimuli

The Experiment was programmed with *E-Prime 3.0*. To take part in the Experiment, participants were redirected from Prolific to *E-Prime Go* and had to download a small task file on their computer. Stimuli were four neutral English adjectives (warm, quiet, open, clear) which appeared in one of four colors (red, green, blue, yellow), resulting in 4 x 4 = 16 color-word combinations. Stimuli were presented in Consolas font (18 pts.) on a black background. We informed participants beforehand via Prolific that the Experiment requires a screen size about 15 inches and Windows as operating system. The keyboard served to collect responses. Two target (color) stimuli were assigned to one response key. Participants used the index fingers of both hands to press the "s" (left) and "k" (right) keys to classify the colors of the words.

**Procedure**

Participants first read a primer of the study content and requirements on Prolific. Demographic information (gender, age, handedness) was gathered at the beginning of each experiment, followed by the consent page. If participants consented to participate, instructions were given; otherwise, the study was terminated. Participants were informed that they will see words which appear in different colors and that their task is to categorize the color of the words. After providing an example for the task, participants were asked to keep the index fingers of both hands above the response keys and to react as fast as possible. Further, they were informed that the experiment consists of one training block and eight experimental blocks.

After reading the instructions, participants worked through the training block that was identical to the experimental blocks (regarding the color-word contingencies) and consisted of 80 trials. Participants had to repeat the training block if they committed more than 20% errors. If the error rates still exceeded 20% after the repetition, the experiment was aborted. At the end of the experiment, we asked participants if they were aware of the color-word contingencies. After finishing the experiment, participants were given a code that served as a proof for successful completion. When they entered the code in the Prolific portal, participants were rewarded accordingly.

Every trial started with a fixation cross (500 ms) followed by a short presentation of the distractor (word in white). The duration was randomly selected from one out of five possible values (150, 200, 250, 300 or 350 ms) after which the final color-word combination appeared until key press. Incorrect responses elicited the feedback “Error – be more accurate! Continue with space-bar...”. Responses slower than 1,250 ms elicited the feedback “Too slow – respond faster! Continue with space-bar...”, but only in the training block. In the experimental blocks, participants received error feedback only. Feedback was displayed in white font on red background until key press. Afterwards, the next trial started immediately. Throughout the experiment, the assignment of colors to corresponding responses (left or right) was displayed on the upper left and upper right side of the screen (see Figure 2). By constantly showing the response mapping on screen, participants do not need to rely on their memory while performing the task, whereby the task becomes easier to learn and less demanding, resulting in faster and more correct responding.

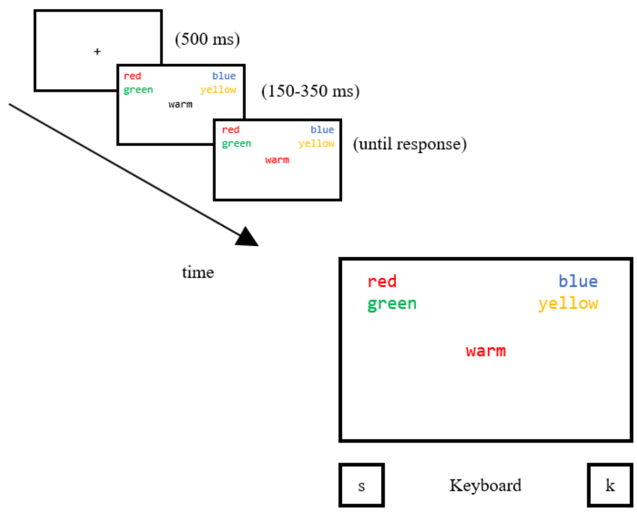


***Figure 2***. Exemplary trial sequence in the color classification task.

After completing eight experimental blocks, participants responded to four contingency awareness questions that were presented on screen (see Figure 3). They had to indicate via key press in which color the displayed distractor word was presented most often (forced choice). The assignment of the four possible colors to the response keys (1-4) was randomized for each word. Responses were coded separately for each word within each participant as being either correct (exact color in which the word had been presented most often = contingency aware) or incorrect[2] (different color that was assigned to a

[2] Responses in which the color was chosen that shared the response key with the high frequency color (i.e., open in yellow) were dropped from the analyses, since although incorrect, using this contingency information may either facilitate or interfere with responding: While predicting the wrong color might impede responding, predicting the correct response might facilitate responding. Since this form of "intermediate knowledge" my either produce or prevent the emergence of a residual CL effect, we omitted these trials from the analysis.

different key = contingency unaware). Overall, 54% of the color-word combinations were recognized correctly, which significantly exceeds the chance level of 25%, $t[500] = 20.63$, $p < 0.001$, but is still far from being perfect.

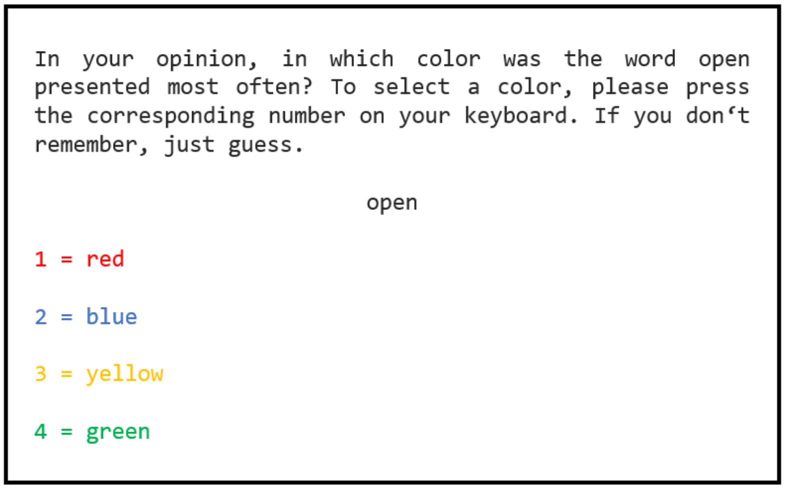


***Figure 3***. Exemplary question in the contingency awareness assessment.

## Data preparation

Prior to analysis, the first trials of each block (1.3%) and trials with erroneous responses (3.2%) were excluded. Next, we excluded all trials that were preceded by trials with erroneous responses during the last occurrence of the same distractor (6.6%). Finally, we eliminated outliers by removing all trials with RTs that were below 250ms or that were more than 3 interquartile ranges above the third quartile of the individual RT distribution (1.2%; "far out values" according to Tukey, 1977). Practice blocks were omitted from the analysis.

## Results

We conducted a series of hierarchical multi-level analyses based on individual trials, treating trials as nested within subjects, while allowing for random intercepts to control for differences in response speed between participants. RT was the dependent variable of interest. Predictors (i.e., CL [contingency learning]: high vs. low contingency trials, TR [target retrieval]: same color vs. different color during the last occurrence of the distractor word, RR [response retrieval]: same response key vs. different response key during the last occurrence of the distractor word, CA [contingency awareness]: choosing the correct color vs. a color that was assigned to a different response key) were successively added to the models.

All predictors indicating a contrast between two conditions were coded to have (1) a mean of zero across all trials within the analysis, and (2) a difference of 1 between the two weights. The general formulas that will satisfy these standards are:

(1) $w_1 = \frac{f_2}{(f_1+f_2)}$ and

(2) $w_2 = \frac{-f_1}{(f_1+f_2)}$

where $w_1$ and $w_2$ are the regression weights that define a contrast, and $f_1$ and $f_2$ are the number of trials per condition. Equal trial frequencies in each condition will result in a contrast coding of $w_1$ = .5 and $w_2$ = -.5. If the frequencies are unequal, however, this will result in unequal weights, so that the condition with the higher frequency receives a lower (absolute) weight (and vice versa). Thus, the resulting regression coefficient reflects the difference between the two conditions (in milliseconds), and the main effects and interactions of the predictors can be interpreted simultaneously.

The results of our analyses are presented in Table 2. As there were no significant interactions with wave, we only report the results of the comprehensive analysis here, including the full data set of $N$ = 500 participants. To assess contingency effects, we focus on the contrast between high contingency (hc, coded as +.28) vs. low contingency/different response trials (lc/dr, baseline condition; coded as -.72), as this corresponds to conventional CL effects. Results of an additional analysis (lc/sr vs. lc/dr) can be found in the Appendix A (Table A1)[3].

**Simple contingency learning effects (Model 1)**

In a first analysis, we tested the effects of contingencies without controlling for retrieval. Including contingency as the only level 1 predictor yields a highly significant CL effect: On average, participants responded 20 ms faster in high contingency compared to low contingency trials, $t(245756) = -37.43$, $p < .001$.

**CL effects after controlling for recency-based episodic S-S and S-R retrieval (Model 2)**

Next, we entered episodic retrieval of target (S-S) and response (S-R) information from the last matching episode as additional predictors to the regression model. Target retrieval was coded as +.20 if the probe word was presented in the *same color* during its last occurrence, and as -.80 if the word was presented in a *different color* during its last occurrence (within the same response condition). Response retrieval was coded as +.76 if the probe word was presented in a color that shared the *same response* key for this word during its last occurrence, and as -.24 if the word was presented in a color that was assigned to the *opposite response* key during its last occurrence. As retrieval effects decrease with higher distance to the last occurrence of the same distractor, we also included distance (number of trials between the current trial and the trial in which the word occurred the last time it had been shown; logarithmized and

---

[3] This type of analysis has been used by Xu and Mordkoff (2020) in their study to assess transfer effects of color-word contingencies to other colors that share the same response with the high contingency trials. The results are comparable to those reported in Table 2, but we refrained from reporting them in the main text, since (a) this type of indirect CL effect does not allow for a straightforward analysis of modulating effects of contingency awareness, and (b) there is a confound of the lc/sr vs. lc/dr contrast with color repetitions that artificially reduces the effect. Specifically, within the "same response trials", the ratio of "same color" trials to "different color" trials is 1:5 in the lc/sr condition, whereas the ratio is balanced (1:1) for the "same response" trials in the lc/dr condition. This confound leads to slower responses in the same response trials of the lc/sr trials, and thus undermines the CL effect.

centered) and the interactions of distance with target and response retrieval in this step of our regression model.

**Table 2**

*Results of a stepwise multi-level regression analysis predicting RT based on contingencies (high vs. low frequency, step 1), episodic retrieval (retrieval of a matching/mismatching color [TR] or response [RR], as a function of the distance between the current trial and the last occurrence of the distractor word), and contingency awareness and its interaction with CL*

| Predictor | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Intercept | 512.55***<br>[505.85, 519.25] | 512.67***<br>[505.98, 519.36] | 512.69***<br>[506.03, 519.36] |
| CL (HC vs. LC/DR) | -19.73***<br>[-20.76, -18.69] | -6.46***<br>[-7.62, -5.29] | -6.49***<br>[-7.65, -5.32] |
| Target retrieval (TR) | | -65.60***<br>[-68.21, -62.99] | -65.50***<br>[-68.11, -62.89] |
| Response retrieval (RR) | | -52.55***<br>[-55.24, -49.85] | -52.45***<br>[-55.14, -49.76] |
| Distance (D) | | 12.03***<br>[11.45, 12.61] | 12.04***<br>[11.46, 12.62] |
| TR x D | | 86.36***<br>[83.43, 89.29] | 86.36***<br>[83.43, 89.29] |
| RR x D | | 67.31***<br>[64.30, 70.31] | 67.30***<br>[64.30, 70.30] |
| CA | | | -5.59***<br>[-6.91, -4.27] |
| CL x CA | | | -11.94***<br>[-14.29, -9.59] |
| BIC | 3049921 | 3042634 | 3042488 |
| Δ BIC | – | -7287 | -146 |

*Note*. *p < 0.05, **p < 0.01, ***p < 0.001. CL, Contingency learning. HC, high contingency trials. LC/DR, low contingency trials requiring a different response than the HC trials (baseline). CA, contingency awareness. BIC, Bayesian information criterion. We implemented a person specific intercept (*df* = 498) to control for individual differences in RTs. All other variables were implemented on a trial level. Values in brackets indicate the 95% confidence interval (lower and upper limit) for each regression weight. Regression weights (*ß*) reflect the difference in milliseconds between the conditions that define a contrast.

Adding target and response retrieval as additional level 1 predictors yields large and highly significant effects for these variables, indicating faster response times if the color of the last matching episode was the same as in the current trial ($M$ = -65.60 ms, $t$[245751] = -49.23, $p < .001$), and also if the response of the last matching episode was the same as the response required in the current trial ($M$ = -52.55 ms, $t$[245751] = -38.24, $p < .001$). Interaction terms of distance and the two retrieval effects were also significant, indicating that the strength of retrieval was larger the shorter the distance between the current and the most recent matching episode. Controlling for both types of retrieval reduced the CL effect substantially (from 20 ms to 6 ms), indicating that episodic retrieval explains a large part of the CL effect. Importantly, however, the CL effect was still significant after controlling for recency-based retrieval, $t$(245749) = -10.85, $p < .001$, indicating a robust residual effect of CL.

**Contingency awareness (Model 3)**

In a final step, contingency awareness (coded as +.35 if the correct color was detected as the high frequency color for the respective word by the participant; coded as -.65 if an incorrect color had been indicated for this word; trials in which the low frequency color was chosen that shared the response with the high frequency color were eliminated by coding them as 0 in the predictor, see fn. 2) and its interaction with CL were entered as additional predictors to the model. Importantly, the interaction of contingency awareness with CL yields a highly significant effect ($M$ = -11.94 ms, $t$[245748] = -9.97, $p < .001$), indicating that the residual CL effect is stronger for participants with contingency awareness. To follow up on this finding, we analyzed the strength of the residual CL effect (i.e., after controlling for both types of episodic retrieval), separately for each awareness condition. The results are summarized in

Table 3. Whereas a significant residual CL effect emerged when participants identified the correct color linked to the distractor ($M$ = -10.62 ms, $t$[131949] = -13.23, $p$ < .001), there was no residual effect of CL for those trials for which the respective participant was unable to recognize the contingency ($M$ = 1.77 ms, $t$[70331] = 1.51, $p$ = .131)[4].

**Table 3**

*Contingency learning effects for trials for which the correct color-word contingency was (CA+) vs. was not (CA-) correctly identified*

| Predictor | CA+ | CA- |
|---|---|---|
| Intercept | 507.70*** [500.75, 514.65] | 523.33*** [515.18, 531.49] |
| CL (HC vs. LC/DR) | -10.62*** [-12.20, -9.05] | 1.77 [-0.52, 4.06] |
| Target retrieval (TR) | -66.00*** [-69.51, -62.49] | -68.37*** [-73.52, -63.21] |
| Response retrieval (RR) | -52.13*** [-55.76, -48.51] | -54.54*** [-59.85, -49.23] |

*Note.* *p < 0.05, **p < 0.01, ***p < 0.001. CL, Contingency learning. HC, high contingency trials. LC/DR, low contingency trials requiring a different response than the hc trials (baseline). Values in brackets indicate the 95% confidence interval (lower and upper limit) for each regression weight. Regression weights ($\beta$) are reported in milliseconds.

[4] For trials in which participants reported the color that shared the response key with the high frequency color, the residual CL effect was also significant ($M$ = -7.75 ms, $t$[42931] = -5.52, $p$ < .001).

## Discussion

The primary goal of the present study was to investigate whether there is a residual CL effect after controlling for episodic retrieval effects. While former studies reported that the effect of global contingencies is fully explained by retrieval of the last matching episode (Giesen et al., 2020; Schmidt et al., 2020), other findings suggested that a CL effect persists after controlling for episodic retrieval (Xu & Mordkoff, 2020). First, within the present study, we obtained a robust CL effect, indicating that participants learned the distractor-response associations over the course of the experiment. In a second step, controlling for episodic retrieval explained a large chunk of the CL effect, but not all of it, suggesting that the effect of contingency learning is not fully accounted for by recency-based retrieval. Third, we found that the residual CL effect depends on contingency awareness, as no residual CL effect occurred if participants were not aware of the color-word contingencies.

### Episodic retrieval as the primary source of contingency learning

In line with previous findings, the main part of the CL effect can be explained by retrieval of the most recent matching episode, which adds further evidence to the "law of recency" (Giesen et al., 2020; Schmidt et al., 2020). We found robust retrieval effects of distractor-response (S-R) and distractor-target (S-S) bindings, indicating that the retrieval mechanism is not limited to actions that have been previously executed, but extends to relevant stimulus attributes as well (see also Giesen & Rothermund, 2014). Further, the strength of retrieval effects obtained in this study decreased with higher distance to the last occurrence of the same distractor, replicating findings reported by Giesen et al. (2020; see also Schmidt et al., 2020), which reflects a crucial feature of episodic retrieval effects (Frings, 2010; Frings et al., 2020). In sum, then, our findings add further evidence to the existing literature that the color-word contingency learning to a large extent reflects recency-based episodic retrieval.

In contrast to our findings, some former studies report that controlling for episodic retrieval of previous responses fully eliminated the CL effect (Giesen et al., 2020; Schmidt et al., 2020). This result may be attributed to the use of weaker (i.e., 2:1) and more complex contingencies, where one distractor word was predictive for two colors instead of one. This renders contingencies harder to detect and to memorize, which hinders propositional knowledge to unfold. Another reason could be the insufficient power in previous studies to detect a residual CL effect.

**The propositional nature of human contingency learning**

While controlling for episodic retrieval of previous targets and responses substantially reduced the CL effect, a numerically small but reliable residual CL effect remained, indicating that not all of the CL effect is due to episodic binding and retrieval (see also Xu & Mordkoff, 2020). To further investigate the nature of the underlying processes of this residual CL effect, we investigated its modulation by contingency awareness. Our findings demonstrated a robust residual CL effect when contingency awareness of the related color-word contingency was present.

This finding extends prior research demonstrating stronger CL effects if participants were aware of the contingencies (i.e., acknowledged noticing the contingencies, Schmidt & De Houwer, 2012). Second, we found no consistent evidence for a residual CL effect in the absence of contingency awareness, despite the fact that our study had a high power to demonstrate such an effect if it existed. This finding supports an explanation of the residual CL effect in terms of propositional knowledge (Mitchell et al., 2009), which is based on a reasoning process that produces conscious declarative knowledge about the contingencies between words and colors. Consequently, such a propositional learning account predicts that a residual CL effect depends on participants becoming aware of and applying the relevant contingencies in their behavioral decisions, which is exactly in line with our results.

In contrast, an associative learning account would predict that residual CL effects should reflect differences in the amount of exposure to certain color-word combinations and thus should occur even in the absence of contingency awareness. Our current findings do not support these predictions, as we did not find any evidence for CL effects in the absence of awareness. Exposure to distractor-target contingencies was exactly the same for all distractor words. Still, we obtained a residual CL effect only for those trials in which participants correctly identified the respective color-word contingency. This influence of contingency awareness on the residual CL effect lines up with the results by Schmidt and De Houwer (2012) who found that giving correct/false instructions about the contingencies in a task led to stronger/weaker CL effects, which attests to the causal influence of contingency awareness on CL effects.

**Limitations and suggestions for future research**

Responding in the forced-choice discrimination task can be attributed to different kinds of processes (i.e., associative learning and propositional knowledge), and above-chance levels in contingency awareness may reflect a mixture of explicit knowledge and implicit learning. Importantly, however, our focus was not on the global levels of contingency awareness. Instead, we analyzed the differences in contingency awareness for different distractors within participants and used these differences to predict differences in CL. Assuming that exposure was identical for each distractor (i.e., for each distractor word, the same number of high and low contingency color-word combinations was presented), possible associations should not differ systematically between distractors. Still, we found systematic differences in contingency awareness that were accompanied by differences in the magnitude of the CL effect. Hence, we conclude that systematic effects of contingency awareness on CL reflect propositional knowledge, as they cannot easily be attributed to associative learning. Importantly, we do not claim that the answers in the forced-choice discrimination task reflect propositional knowledge only. Still, we think that our results provide, at least, indirect evidence for the importance of propositional knowledge for CL.

Since we did not experimentally manipulate contingency awareness in our study, our findings do not allow us to infer an unambiguous cause-effect relationship between contingency awareness and CL. Although our findings imply that CL depends on awareness, it is still possible that this relation does not reflect a causal effect of contingency awareness on CL. Instead, some background variable might influence both contingency awareness and CL in similar ways, rendering the relation spurious. For example, associative learning might be necessary for contingency awareness to emerge, and might also be the underlying source of a residual CL effect. As argued above, however, it remains unclear why there should be systematic differences in contingency awareness based on associative learning, since exposure to the contingencies was identical for all distractor words. Association formation thus seems an unlikely candidate for explaining these differences in contingency awareness and CL (De Houwer, 2009, 2014a, 2014b; Mitchell et al., 2009; Stahl et al., 2016). Still, even a propositional account does not offer any explanation where these differences in contingency awareness come from (e.g., one might speculate that contingencies that were detected first might have a higher chance to be remembered and applied, but we have no evidence for this conjecture). Further, the assumption that associative learning is solely influenced by the frequency of color-word pairings might be too simplistic, as other factors (i.e., attention, salience, and presentation sequence) might contribute to the emergence of the CL effect. These factors might also explain differences in contingency awareness (i.e., why some contingencies are noticed while others are not).

Future investigations could implement experimental manipulations of awareness and contingency learning (e.g., by increasing the number of rules to be learned). If the residual CL effect reflects propositional reasoning, it should be sensitive to the amount of available cognitive resources (i.e., high cognitive load should reduce the residual CL effect). Similarly, subjective awareness of contingencies could be manipulated by correctly vs. incorrectly informing participants about the contingencies before the experiment, which should affect the resulting residual CL effect (cf. Schmidt & De Houwer, 2012). Another promising approach for future research is to investigate the potential sources of CL and contingency awareness (i.e., differences in learning and awareness between conditions showing the same frequency of pairings) by measuring and/or manipulating other factors than the frequency of pairings (i.e., attention, salience, and presentation sequence of color-word combinations). Finally, future studies could employ different measures of contingency awareness (i.e., by evaluating participants confidence in each answer), which would facilitate drawing conclusions about the origins of contingency awareness.

## Conclusion

The present study provides new insights about the role of episodic retrieval and contingency awareness in human contingency learning. First, we observed a robust CL effect. Second, this effect was, for the most part, eliminated after controlling for episodic retrieval of previous targets and responses. Still, we found a residual CL effect which was amplified if contingency awareness was high. Notably, no CL effect remained if participants lacked awareness of the color-word contingencies. Our findings (1) support episodic retrieval accounts of human contingency learning and (2) underline the importance of propositional mechanisms which operate independently of episodic binding and retrieval.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

This research was supported by grants from the HONOURS program for science-oriented students of the FSU Jena and the German Research Foundation (RO 1272/13-1, 13-2).

## Notes

Data and R scripts are publicly available on the Open Science Framework: https://osf.io/ubqfx/. More information is available from the lead author on request.

**References**


Allan, L. G. (2005). Learning of contingent relationships. *Learning & Behavior, 33*, 127-129. https://doi.org/10.3758/BF03196057

Arunkumar, M., Rothermund, K., & Giesen, C. G. (2023). One link to link them all: Indirect response activation through stimulus-stimulus associations in contingency learning. *Manuscript submitted for publication.*

Arunkumar, M., Rothermund, K., Kunde, W., & Giesen, C. G. (2022). Being in the know: The role of awareness and retrieval of transient stimulus-response bindings in selective contingency learning. *Journal of Cognition*, *5*(1), 1-21. https://doi.org/10.5334/joc.227

Arunkumar, M., Rothermund, K., Kunde, W., Mocke, V., & Giesen, C. G. (2023). Cling together, swing together? Assessing indirect retrieval of stimulus-response bindings for associated stimuli. *Manuscript submitted for publication*.

Bargh, J. A. (1994). The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues. In R. S. J. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1-40). Hillsdale, NJ: Erlbaum.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*(1), 1-20. https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2014a). A propositional perspective on context effects in human associative learning. *Behavioural processes*, *104*(2014), 2-12. https://doi.org/10.1016/j.beproc.2014.02.002

De Houwer, J. (2014b). Why a propositional single-process model of associative learning deserves to be defended. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 530–541). The Guilford Press.

De Houwer, J., & Hughes, S. (2017). Environmental regularities as a concept for carving up the realm of learning research: Implications for Relational Frame Theory. *Journal of Contextual Behavioral Science, 6*(3), 343-346. https://doi.org/10.1016/j.jcbs.2016.07.002

Frings, C. (2010). On the decay of distractor-response episodes. *Experimental Psychology, 58*(2), 125-131. https://doi.org/10.1027/1618-3169/a000077

Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020). Binding and retrieval in action control (BRAC). *Trends in cognitive sciences, 24*(5), 375-387. https://doi.org/10.1016/j.tics.2020.02.004

Garner, W. R. (2014). *The processing of information and structure*. New York, NY: Psychology Press.

Giesen, C., Frings, C., & Rothermund, K. (2012). Differences in the strength of distractor inhibition do not affect distractor–response bindings. *Memory & Cognition, 40*(3), 373-387.

Giesen, C., & Rothermund, K. (2014). Distractor repetitions retrieve previous responses and previous targets: Experimental dissociations of distractor–response and distractor–target bindings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(3), 645-659. https://doi.org/10.1037/a0035278

Giesen, C. G., Schmidt, J. R., & Rothermund, K. (2020). The law of recency: An episodic stimulus-response retrieval account of habit acquisition. *Frontiers in Psychology, 10*, 2927. https://doi.org/10.3389/fpsyg.2019.02927

Hirshman, E. (1995). Decision processes in recognition memory: criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 302-313.

Jiménez, L., Gallego, D., Agra, O., Lorda, M. J., & Méndez, C. (2022). Proportion of conflict, contingency learning, and recency effects in a Stroop task. *Quarterly Journal of Experimental Psychology*, *75*(8), 1528-1540. https://doi.org/10.1177/17470218211056813

McLaren, I. P. L., Green, R. E. A., & Mackintosh, N. J. (1994). Animal learning and the implicit/explicit distinction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 313-332). Academic Press.

Miller, J. (1987). Priming is not necessary for selective-attention failures: Semantic effects of unattended, unprimed letters. *Perception & Psychophysics*, *41*(5), 419-434. https://doi.org/10.3758/BF03203035

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183-198. https://doi.org/10.1017/S0140525X09000855

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: American Psychological Association. https://doi.org/10.1037/10187-018

Pavlov, I. P. (1927). *Conditioned reflexes*. London, England: Oxford University Press.

Rescorla, R. A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64-69). New York, NY: Appleton-Century-Crofts.

Rothermund, K., Giesen, C. G., & Gollnick, N. (2022). Proportion congruency effects in the Stroop task: Retrieval of stimulus-response episodes explains it all. *Journal of Cognition, 5*(1), 1-16. https://doi.org/10.5334/joc.232

Rothermund, K., Kapinos, L., De Houwer, J., & Schmidt, J. R. (2023). Long-term contingency learning: The role of episodic retrieval and contingency awareness. *Manuscript submitted for publication*.

Rudolph, M. (2023, August 24). Two sources of color-word contingency learning: Episodic retrieval of SR bindings and propositional knowledge. Retrieved from osf.io/ubqfx

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

Schmidt, J. R., & De Houwer, J. (2012). Learning, awareness, and instruction: Subjective contingency awareness does matter in the colour-word contingency learning paradigm. *Consciousness and cognition, 21*(4), 1754-1768. https://doi.org/10.1016/j.concog.2012.10.006

Schmidt, J. R., & De Houwer, J. (2016a). Contingency learning tracks with stimulus-response proportion: No evidence of misprediction costs. *Experimental Psychology*, *63*(2), 79-88. https://doi.org/10.1027/1618-3169/a000313

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

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

Shanks, D. R. (2010). Learning: From association to cognition. *Annual review of psychology, 61*, 273-301. https://doi.org/10.1146/annurev.psych.093008.100519

Shanks, D. R., & John, M. F. S. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367-395. https://doi.org/10.1017/S0140525X00035032

Stahl, C., & Bading, K. C. (2020). Evaluative conditioning of pattern-masked nonwords requires perceptual awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(5), 822-850. https://doi.org/10.1037/xlm0000757

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal evaluative conditioning? Above-chance CS identification may be necessary and insufficient for attitude learning. *Journal of Experimental Psychology: General*, *145*(9), 1107-1131. https://doi.org/10.1037/xge0000191

Xu, G., & Mordkoff, J. T. (2020). Reliable correlational cuing while controlling for most-recent-pairing effects. *Frontiers in Psychology*, *11:592377*. https://doi.org/10.3389/fpsyg.2020.592377

## Appendix A

**Table A1**

*Results of a stepwise multi-level regression analysis predicting RT based on contingencies (high vs. low frequency, step 1), and episodic retrieval (retrieval of a matching/mismatching color [TR] or response [RR], as a function of the distance between the current trial and the last occurrence of the distractor word; step 2)*

| Predictor | Model 1 | Model 2 |
|---|---|---|
| Intercept | 524.92***<br>[518.07, 531.77] | 524.97***<br>[518.12, 531.82] |
| CL (LC/SR vs. LC/DR) | -5.33***<br>[-6.87, -3.80] | -4.31***<br>[-6.11, -2.51] |
| Target retrieval (TR) | | -35.08***<br>[-37.79, -32.36] |
| Response retrieval (RR) | | -12.33***<br>[-14.44, -10.23] |
| Distance (D) | | -2.02***<br>[-2.93, -1.11] |
| TR x D | | 44.89***<br>[41.45, 48.33] |
| RR x D | | 9.36***<br>[7.04, 11.69] |
| BIC | 1301014 | 1299728 |
| Δ BIC | – | -1286 |

*Note*. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. CL, Contingency learning. LC/SR, low contingency trials requiring the same response as the HC trials. LC/DR, low contingency trials requiring a different response than the HC trials. CA, contingency awareness. BIC, Bayesian information criterion. We implemented a person specific intercept ($df$ = 498) to control for individual differences in RTs. All other variables were implemented on a trial level ($df_{\text{Trial}}$ = 104222). Values in brackets indicate the 95% confidence interval (lower and upper limit) for each regression weight. Regression weights (*ß*) reflect the difference in milliseconds between the conditions that define a contrast.

The predictors were coded as follows to guarantee a mean of zero and a difference of 1 between the weights of the conditions: contingency (high contingency = 0.67, low contingency = -0.33), target retrieval (same target = 0.70, different target = -0.30), response retrieval (same response = 0.67, different response = -0.33).