

**False contingency beliefs reverse contingency learning effects in the valence contingency
learning task**

Carina G. Giesen¹, Matthäus Rudolph² & Klaus Rothermund²

¹Health and Medical University Erfurt, Germany

²Friedrich Schiller University Jena, Germany

Cognition & Emotion, in press

ORCID

Carina G. Giesen 0000-0002-2395-4435

Matthäus Rudolph 0009-0008-8305-5344

Klaus Rothermund 0000-0002-9350-5272

Corresponding author: Carina G. Giesen, HMU Health and Medical University Erfurt,
Faculty of Health, Department of Psychology, Anger 64-73 99084 Erfurt, Germany. E-Mail:
carina.giesen@hmu-erfurt.de. Hypotheses, design, and analyses were preregistered
(<https://aspredicted.org/6h9h-xbsn.pdf>). All materials, data, and analyses are available at the Open
Science Framework (OSF) at <https://osf.io/e73p5/>). Supplementary analyses of CL effects as a
function of block are accessible on the OSF, <https://osf.io/zv4bd>.

1 **Abstract**

2 In the valence contingency learning (VCL) task, participants evaluate target words which
3 are preceded by nonwords, which are predictive for positive/negative evaluation responses. This
4 produces robust contingency learning (CL) effects, reflected in faster and more accurate
5 performance for highly contingent nonword-valence pairings. Previous findings indicate that
6 controlling for episodic retrieval of transient stimulus-response episodes reduces CL effects but
7 does not eliminate them, as a residual CL effect remains. These residual CL effects are best
8 explained by propositional learning. To substantiate this, the present study manipulated
9 participants beliefs about contingencies in the VCL task. Participants received either true, false or
10 no instructions regarding the actual nonword-valence contingencies. Effects of contingency
11 learning and evaluative conditioning (EC) for nonwords were assessed. As expected, contingency
12 beliefs modulated contingency learning, as true instructions boosted residual CL effects; false
13 instructions reduced residual CL effects, relative to the no instruction condition. Exploratory
14 analyses revealed a modulatory influence of contingency beliefs on EC effects, which varied solely
15 as a function of (remembered) contingency instruction and were unaffected by experienced
16 contingencies. The present study conceptually replicates findings from color-word contingency
17 learning in the realm of evaluative learning. Implications for theories on processes underlying
18 contingency learning and evaluative conditioning are discussed.

19 *Keywords:* contingency learning, evaluative conditioning, stimulus-response episodes,
20 learning by instruction, contingency awareness, propositional learning.

21

1 In studies on human contingency learning, experimenters typically ask participants to
2 perform an arbitrary response like color or valence classification to stimuli (e.g., words) presented
3 on screen. The frequency of particular stimulus-response (SR) pairings is manipulated, meaning
4 that some SR pairings are presented more frequently than other SR combinations. Contingency
5 learning is defined as better performance for events that are consistent with a learnt regularity
6 compared with events that are inconsistent with the learnt regularity (Schmidt, 2021). In terms of
7 behavioral effects, contingency learning (CL) is reflected in better performance on trials with
8 *frequent* SR pairings (high contingency trials) than trials with *infrequent* SR pairings (low
9 contingency trials; Schmidt et al., 2007; Schmidt et al., 2012a, b; Schmidt, 2021; see also Miller,
10 1987).

11 There are many experimental procedures to study human contingency learning. In the
12 color-word contingency learning paradigm, participants classify the color of different words. Word
13 meaning is irrelevant for the task, and unbeknownst to participants, each word is presented most
14 often in one color and less frequently in the remaining colors. Yet, participants incidentally learn
15 these contingencies between words and color responses. This is indicated by better performance
16 on frequent compared with infrequent SR combinations (Schmidt et al., 2007). The valence
17 contingency learning task (Schmidt & De Houwer, 2012a) is another example to study contingency
18 learning in humans. Participants classify target words as positive or negative in a speeded response
19 task. Before each target word, a nonword is presented. Nonwords are irrelevant for the task, but
20 some occur more often with positive targets and others more often with negative targets.
21 Participants learn the contingencies between nonwords and valence responses. Effects of
22 contingency learning are reflected in faster and more accurate performance for frequent versus

infrequent nonword-valence pairings (Schmidt & De Houwer, 2012a; see also Gast et al., 2020; Giesen et al., 2025).

Paradigms of contingency learning are of high interest in cognitive research. They allow to investigate the underlying processes which enable organisms to learn about contingencies and adapt their behavior to regularities in their environment (De Houwer & Hughes, 2020; Shanks, 2010). Recent findings indicate that CL effects can be the outcome of two independent processes¹, namely (1) stimulus-based *episodic retrieval of the previously executed response* and (2) the application of *contingency knowledge* (Giesen et al., 2020, 2025; Rudolph & Rothermund, 2025; Rudolph et al., 2025a). The former process represents an automatic process that operates independently of conscious awareness (Rudolph et al., 2025b) or goals (Foerster et al., 2021; Parmar et al., 2022; Parmar & Rothermund, 2024). The latter process represents the influence of propositional knowledge or beliefs (Schmidt & De Houwer, 2012b; Mitchell et al., 2009). Specifically, when encountering a specific cue, beliefs about regularities and co-occurrences can be used to predict and prepare a response for the respective trial. This will speed up responding on trials with frequent SR pairings but will interfere with responding on trials with infrequent SR pairings, because the correct or an incorrect response has been prepared, respectively.

¹ We would like to emphasize here that we are specifically interested in distinguishing the influence of two different operating principles on the VCL effect: Episodic retrieval and expectancy-based response preparation based on propositional beliefs about abstract regularities. Although it is plausible to assume that these processes also differ with regard to their operating conditions (Gawronski & Bodenhausen, 2009, 2014; see also Corneille & Stahl, 2019), with episodic retrieval operating automatically and independently of awareness (e.g., Rudolph et al., 2025b) while contingency-related beliefs can be assumed to operate in a controlled mode, also requiring awareness, this difference in operating conditions is not the focus of our study.

How does episodic retrieval explain effects of contingency learning, namely improved performance on frequent compared to infrequent SR pairings? As soon as a stimulus is repeatedly presented in a task, stimulus-based episodic retrieval will reactivate the response that was executed on the last (i.e., most recent) previous occurrence of the same stimulus (*law of recency*; Giesen et al., 2020). Hence, episodic retrieval can explain the performance pattern that is otherwise attributed to contingency learning, due to a confounding between contingency and retrieval: Whereas high contingency trials retrieve mostly matching responses (namely: responses that conform with the more frequent SR combination), which will improve performance, low contingency trials will retrieve mostly mismatching responses from the preceding episodes (again: responses that conform with the more frequent SR combination), which will impede performance. Initial studies indicated that controlling for episodic retrieval of previous responses eliminates CL effects completely (Giesen et al., 2020; Schmidt et al., 2020), which attests to the strong influence that episodic retrieval processes have on CL effects.

Controlling for influences of episodic retrieval processes in contingency learning

Recent studies indicate that stimulus-based episodic retrieval of the last previously executed response explains a large proportion of systematic variance in the CL effect (Giesen et al., 2020; Rudolph & Rothermund, 2025; Schmidt et al., 2020; Xu & Mordkoff, 2020). The basis of this reasoning is a hierarchical multi-level analysis. In Model 1, trial RT is predicted by trial contingency (high vs. low). Typically, this reveals the typical CL effect. In Model 2, episodic response retrieval is added as a second predictor. Effectively, Model 2 tests for systematic variance due to stimulus-specific episodic retrieval of previous responses. To achieve this, each trial is referenced back to the last previous occurrence of the same stimulus (see Figure 1B). Then, the relation between the response required in the current trial and the trial in which the last previous

stimulus occurrence took place is determined. The previous response can either match (i.e., previous response = same response as current trial) or mismatch (i.e., previous response $\neq$ same response as current trial, Figure 1B) with the current correct response.

The most relevant question for Model 2 is whether the contingency predictor will remain significant, even when episodic retrieval processes are statistically controlled for by the previous response predictor. If it does, this is considered as a *residual CL effect*. In our view, only these residual CL effects reflect pure or proper learning that is not confounded by episodic retrieval processes (Rudolph & Rothermund, 2025; Rudolph et al., 2025a), since episodic retrieval processes always refer to a single, most recent episode. Single episodes do not reflect regularities or contingencies. Thus, they will not produce any lasting effect on behavior, as encountering a single episode that deviates from a series of previously encountered episodes will completely reverse the retrieval effect, leaving no trace of all previous episodes. If the contingency predictor is not significant in the Model 2 analysis (i.e., if episodic retrieval has been statistically controlled), this implies that the systematic variance in RT data that was previously attributed to the contingency predictor is in fact fully explained by episodic retrieval processes (e.g., Giesen et al., 2020; Schmidt et al., 2020).

Contingency knowledge as an (additional) source of contingency learning

Recent studies indicate that a significant *residual CL* effect often remains after controlling for effects of stimulus-based episodic retrieval of previous responses (Giesen et al., 2025; Rothermund et al., 2025; Rudolph & Rothermund, 2025; Xu & Mordkoff, 2020), which reflects the impact of global contingencies (rather than retrieval of single episodes). Given that residual, pure CL effects emerged only if participants were aware of the experienced contingencies gave rise to the idea that contingency knowledge is a source of genuine CL effects above and beyond

any systematic effect of episodic retrieval processes. Contingency knowledge can be understood as an outcome of propositional reasoning. It follows from creating and testing hypotheses about the relation between events, such as stimuli and responses in a task (Mitchell et al., 2009). Contingency knowledge in this regard is declarative, meaning that participants can verbalize their assumptions about the underlying stimulus-stimulus and/or stimulus-response relationships. Furthermore, it is propositional in nature, meaning that it specifies how two events are related, e.g., whether a stimulus predicts another stimulus and/or response (Mitchell et al., 2009).

Targeted manipulations allow testing the distinct influence of contingency knowledge and episodic retrieval on contingency learning effects: For instance, Giesen et al. (2025) compared CL effects in the valence contingency learning task under conditions of incidental vs. instructed learning. Participants in the incidental group were not informed about the presence of nonword-valence contingencies. Participants in the instructed group were informed that contingencies are present in the experiment. Nevertheless, both groups showed effects of valence contingency learning, although the effect was stronger in the instructed group. Intriguingly, the CL effect was *fully* explained by episodic retrieval processes for participants in the incidental group. In turn, a significant, residual CL effect remained for participants in the instructed group. This suggests that being instructed about the presence of contingencies helps participants to detect the underlying relationships, thereby facilitating the formation of genuine CL effects that go beyond episodic retrieval of previous responses.

Note that in this study, participants in the instructed learning group received only general instructions that contingencies were present in the experiment. Rudolph et al. (2025a) went one step further and manipulated word-specific contingency beliefs in a color-word contingency task: Participants either received true instructions (e.g., “the word ‘short’ will mostly appear in red font”),

1 false instructions (e.g., “the word ‘fast’ will mostly appear in blue font”, although ‘fast’ most often
2 occurred in yellow font) or no instructions before the color categorization task. The findings
3 indicated larger residual CL effects in the true instruction condition. In turn, residual CL effects
4 were smaller in the false instruction condition. Intriguingly, residual CL effects were *reversed* for
5 participants who maintained a belief in the false contingency instruction over the entire task (i.e.,
6 reflecting *faster performance in low contingency trials* compared to high contingency trials). The
7 latter finding is particularly interesting: It demonstrates that some participants held a sustained
8 belief in the instruction although their experience (i.e., working through more than 500 trials on
9 task) was in exact opposition to this belief. Together, the findings by Giesen et al. (2025) and
10 Rudolph et al. (2025a) support the conclusion that propositional inferences or beliefs causally
11 contribute to human contingency learning.

12 As the finding of a reversed residual CL effect (Rudolph et al., 2025a) is based on a single
13 study, the present study had a straightforward motivation: We aimed to replicate the effects of the
14 manipulation of contingency beliefs, but in the context of a valence contingency learning (VCL)
15 task. This way, our study aims at substantiating the evidence for the involvement of propositional
16 beliefs in human contingency learning. Furthermore, our study allows to test the generalizability
17 of findings from color-word contingency learning to other forms of human contingency learning.
18 Eventually, this will further our understanding and provide deeper insight into the processes
19 underlying contingency learning effects.

20 **The present study**

21 In the present study, participants received either true, false, or no instructions regarding the
22 actual nonword-valence contingencies in a VCL task (see Figure 1). In the VCL task, participants
23 had to evaluate target words, which were preceded by nonwords. Each nonword was predictive for

1 a specific valence response. Three nonwords more often preceded positive responses and only
2 rarely preceded negative responses. Three other nonwords more often preceded negative responses
3 and only rarely preceded positive responses (reflecting a nonword-valence contingency ratio of
4 3:1). We then assessed performance (RT) in the VCL task as a function of actual contingency and
5 instruction condition.

6 After the VCL, a rating task followed, in which we asked participants to evaluate nonwords
7 to assess evaluative conditioning (EC) effects. EC effects are defined as a change in liking of the
8 (formerly neutral) nonwords due to their previous pairing or co-occurrence with valenced words
9 (De Houwer, 2007). That is, after frequent pairings with positive stimuli, nonwords are evaluated
10 more positively. In turn, after frequent pairings with negative stimuli, nonwords are typically
11 evaluated more negatively. Collecting EC ratings allows us to connect research on contingency
12 learning and evaluative conditioning. Specifically, we wanted to investigate whether contingency
13 beliefs are also relevant for resulting changes in valence of the nonwords, or whether EC effects
14 reflect more automatic effects of actual pairings with positive or negative words.

15 After evaluating the nonwords, we collected contingency awareness ratings for all (i.e.,
16 instructed as well as experienced) nonword-valence pairings.

17 *Planned analyses and expectations*

18 We analyzed performance in the VCL task in similar fashion as Rudolph et al. (2025a) with
19 the hierarchical multilevel modeling approach described earlier. Trial RT served as dependent
20 variable. Model 1 only considered contingency (high contingency [hc] vs. low contingency [lc])
21 as predictor. In Model 2, we added a factor that statistically controlled for episodic retrieval
22 processes (retrieval of a matching vs. mismatching response). We expected to obtain residual CL
23 effects even after controlling for episodic retrieval processes. In Model 3, we added two dummy-

1 coded predictors to assess the influence of instructions on trial RT, along with interactions with the
2 contingency factor. We expected that residual CL effects should vary as a function of instruction
3 condition: (i) Residual CL effects should be boosted under true instruction, reflected in an
4 interaction of contingency (hc vs. lc) $\times$ true instruction (true vs. no instruction). (ii) Residual CL
5 effects should be reduced under false instruction, reflected in an interaction of the contingency (hc
6 vs. lc) $\times$ false instruction (false vs. no instruction). As we aimed to replicate findings from Rudolph
7 et al. (2025a), residual CL effects should be *reversed* for the false instruction condition, if
8 participants display belief in the falsely instructed contingencies in the awareness assessment.

9 Analyses of EC effects were explorative in nature (see preregistration for details). EC were
10 collected at the end of the experiment, by having participants evaluate the nonwords, treating the
11 previous VCL task as an EC training procedure. Differences in evaluations for nonwords that had
12 been paired with positive and negative target words during the VCL task, respectively, thus served
13 as a standard rating-based indicator of evaluative learning for the nonword-valence pairings during
14 the VCL task that was completely independent of the RT-based VCL effects that were obtained for
15 these non-words during the task.

16 Influential findings in evaluative learning research suggest that EC effects are driven by
17 propositional processes; for instance, they hinge on contingency awareness for CS-US (here:
18 nonword-valence) pairings (e.g., Baeyens et al., 1990; Pleyers et al., 2007; Stahl & Unkelbach,
19 2009) and may be induced solely by instructing participants about CS-US pairings (Benedict et al.,
20 2019; Gast & De Houwer, 2012, 2013; Hütter & De Houwer, 2017). This would imply that EC
21 effects should also vary as a function of instruction condition in the present study, that is, one
22 would expect typical EC effects for the true instruction condition and reversed EC effects for the
23 false instruction condition.

Method

Open Science and preregistration

Study aims, hypothesis, design, planned analyses and exploratory analyses were preregistered at www.aspredicted.org (<https://aspredicted.org/6h9h-xbsn.pdf>). All materials, data, and analyses (Markdown scripts) will be available at the Open Science Foundation pending acceptance of the study (view-only link for review process: https://osf.io/e73p5/?view_only=5072381c4d7d44cc89b64be3b20783cf).

A-priori power analyses, required sample size, and ethical approval

We performed a-priori power calculations (G*Power 3.1, Faul et al., 2007) for matched-pairs *t*-test (one-tailed) with a power of $1-\beta = .80$ and $\alpha = 0.05$. The critical effects of interest correspond to two-way interactions between true instruction vs. no instruction $\times$ contingency, and false instruction vs. no instruction $\times$ contingency, respectively. Effect size estimates were referenced from three existing studies (see preregistration for details) and ranged between $d = 0.23$ to $d = 0.31$, which would require a sample size ranging from $N = 66$ to $N = 119$. Based on these estimates, we planned to recruit a minimum of $N = 66$ participants. Data collection continued as long as the labs were available (five weeks in total) or until a maximum of $N = 119$ was reached. Ethical approval for the study was granted by the Ethics Committee of the FSU Jena (FSV23/071).

Participants

Overall, 81 participants ($M_{\text{age}} = 21$ years, $SD_{\text{age}} = 6.3$ years; 67 female, 13 male, 1 diverse) took part in the study, all reported to be native German speakers. Participants were university students (psychology and medicine) as well as visitors of the Open Campus Day, HMU Erfurt. The study took part in the lab, using desktop computers. Study duration was 25 minutes. Psychology students received partial course credit for taking part, all other participants received cake/waffles

or sweets as compensation. Explicit consent to participate was collected electronically at the beginning of the study from all participants. No participant was excluded from data analyses.

Stimulus material and task

The experiment was programmed in E-Prime 3.0. We used different stimuli during practice and in the main experiment to prevent transfer effects. We used four nonwords during practice (*bodange, hilrube, ureltis, intolek*) and six nonwords in the main experiment (*nijaron, fevkani, kadirga, lokanta, gowange, marenes*). For target words, we used 20 valenced words (10 positive, 10 negative) during practice and 24 valenced words (12 positive, 12 negative) in the main experiment (Appendix A). Words/nonwords were presented in white font on a black screen.

During practice and in the main experiment, participants classified the valence of target words via key press (D or L on a QWERTZ-keyboard). Assignment of positive/negative valence to left/right keys was balanced across participants.

Design

The study had a 3×2 repeated measures design with the within-subject factors contingency instructions (true vs. false vs. no instruction) and contingency (high vs. low). Contingency instructions were manipulated at the level of individual nonwords via specific instructions that were given after the practice block before the main experiment started. For two nonwords, participants were instructed about the correct nonword-valence contingency in the main task (*true instruction*). For two other nonwords, participants were falsely instructed, as the instructed contingency was exactly opposite to the experienced nonword-valence contingency in the task (*false instruction*). For the remaining two nonwords, no information regarding nonword-valence contingencies was given to participants (*no instruction*). Assignment of nonwords to instruction condition was random for each participant. In each instruction condition, one nonword

was predictive for a positive response, the other was predictive for a negative response. Contingency was manipulated by presenting each of the six nonwords 18 times with the frequent valence and 6 times with the infrequent valence category per block, which corresponded to a 3:1 contingency ratio (see Table 1). The main task comprised of four blocks in total (576 trials). Assignment of nonwords to each valence was balanced across participants.

Response times served as dependent variable of interest in the main task, as well as contingency awareness ratings collected at the end of the study. For exploratory purposes, we also collected valence ratings for nonwords to assess evaluative conditioning effects.

Procedure

At the beginning of the study, participants provided demographic information and gave explicit consent to participate. They were informed that they would first see a fantasy word (i.e., the nonword) on screen, followed by a positive or negative target word, presented either above or below the nonword (target word position was determined randomly per trial). Participants were instructed to respond to the target word and classify its valence by pressing a left (D) or right (L) key as fast and correctly as possible. Then, participants had to pass an task instruction check (two questions: “Which key do you have to press for positive (negative) words?”). If they did not respond 100% accurately in this task instruction check, participants were redirected back to reading the instructions. Otherwise, participants continued to the practice block (40 trials). After practice, the instruction manipulation followed. Each nonword was randomly assigned to one of six variables (nonword1 to nonword6). Nonwords with odd numbers were predictive of positive responses, whereas nonwords with even numbers were predictive of negative responses. Of these, nonword1 and nonword2 represented nonwords that were used in the *no instruction* condition; nonword3 and nonword4 were used in the *true instruction* condition, and nonword5 and nonword6

1 were used in the *false instruction* condition. Participants first received instructions regarding
2 nonword3 and nonword6, followed by instructions regarding nonword4 and nonword5.
3 Instructions were as follows: First, participants received general information: “*Important! Some*
4 *fantasy words appear more often with positive target words. Keeping this in mind will help you*
5 *during the task. Please take your time to memorize the following information, because a memory*
6 *test will follow.*”. On the second page, participants received specific instructions for nonword3 to
7 nonword6: “*The fantasy word [nonword3/nonword6] will mostly precede positive target words.*
8 *The fantasy word [nonword4/nonword5] will mostly precede negative target words.*” (for true and
9 false instruction conditions, respectively, cf. Table 1). After these instructions, participants were
10 prompted with a memory test for each nonword instruction (“*Did you memorize the instructions?*
11 *Please report whether the fantasy word presented below will mostly precede positive or negative*
12 *target words by pressing p (positive) or n (negative)*”). To continue to the main task, participants
13 had to respond with 100% accuracy in the memory test. If accuracy was smaller, they were
14 redirected back to the instruction manipulation. Overall, participants could take the memory test
15 three times. The experiment was aborted for participants who did not pass the accuracy criterion
16 after the third round (which never happened).

17 The main task comprised of 576 trials that were constructed with respect to the factorial
18 design. Every nonword was presented 96 (72 high contingency, 24 low contingency) times in total,
19 resulting from 24 (18 high contingency, 6 low contingency) presentations across four blocks.
20 Within each block, trials were presented randomly. After each block, participants were given a
21 brief, self-paced break (until response to continue).

22 The trial procedure in practice and main blocks was identical (Figure 1): Each trial started
23 with a fixation (“+”, 250 ms), followed by a blank black screen (50 ms). Then, a nonword appeared

1 alone. After 250 ms, a target word appeared either above or below the nonword (until response or
2 until 2000 ms had passed). Participants classified the valence of the target word by pressing D or
3 L. In case participants committed an error, a feedback message (“Wrong response! Continue with
4 correct response...” in red font) was displayed (until response). Then, the next trial started.

5 After the main task, a rating task followed, to assess evaluative conditioning effects for
6 nonwords. All six nonwords as well as two positive (flower, hug) and two negative (guns, crime)
7 words were presented in randomized order. Participants were asked to indicate how pleasant they
8 perceived each word on a 8-point Likert scale ranging from 1 (negative) to 8 (positive) by pressing
9 the corresponding number on the keyboard.

10 After the rating task, we assessed participants contingency awareness for the experienced
11 nonword-valence contingencies for each nonword. Participants were asked to indicate whether
12 they had the impression that a given fantasy word was presented more often with positive or
13 negative target words. Instructions explicitly mentioned that the correspondence did not have to
14 be perfect, as a fantasy word could have been paired with one valence in a trial, although it
15 appeared more frequently with the other valence. Participants were first asked to indicate whether
16 a given nonword appeared more often with positive or negative target words over the course of the
17 study by pressing P or N, respectively. Next, we assessed participants confidence on a 5-point
18 Likert scale (1=very unconfident, 2=unconfident, 3=neither, 4=confident, 5=very confident). For
19 exploratory purposes, we also asked at what point during the task they became aware of any
20 relationship between nonwords and target valence (1 to 4: During the first/second/third/fourth
21 block of the task; 5 = I did not detect any relationship).

22 Finally, we assessed memory for instructed contingencies and asked participants whether
23 they still remembered the specific instructions they received before the main task for (truly and

falsely instructed) nonwords 3 to 6. Participants had to respond based on the instruction and should report whether they were informed that a given nonword would appear more often with positive or negative target words by pressing P or N, respectively. Then, participants were fully debriefed on screen, thanked and compensated.

Analytical approach for performance in the VCL task

RT data from the VCL task were analyzed with a linear mixed-effects model (LMM), using trials as unit of analysis (level1 predictor) nested within participants (level2 predictor). In total, we ran three random intercept models in hierarchical order, treating participants as random effects, and predictors and their interactions as fixed effects. Note that coding of predictors and order of analyses follows the analytical approach of Rudolph et al. (2025a) as closely as possible, to allow for an easy comparison of results. Predictors for each model were centered via contrast coding of each predictor with respect to the relative frequencies of its factor levels with a mean of zero and a difference of 1 between the two weights (see Models and Footnote 1 for details). This way, the resulting regression coefficients reflect the mean RT difference (in ms) between the two conditions.

Results

All analyses were conducted with R (see open data Markdown script for details on packages).

Preregistered analyses

Memory for instructed contingencies. We probed participants' memory for instructed contingencies at the end of the study. Participants had to report the instructed contingency for nonword 3 to 6 (i.e., those nonwords that were mentioned in the instructions). The accuracy of these responses was coded for each nonword and averaged per person. We analyzed mean accuracy scores as a function of instruction (true vs. false), which did not differ significantly, $t(80) = -0.70$,

1 $p = .483$, $d_z = 0.08$, $BF_{01}=6.67$, indicating that participants memory for true ($M = 0.75$, $SD = 0.34$)
 2 vs. false instructed ($M = 0.71$, $SD = 0.33$) contingencies was comparable and equally good, as both
 3 scores also significantly different from chance (0.5), $t(80) = 6.59$, $p < .001$, $d_z = 0.73$ and $t(80) =$
 4 5.81 , $p < .001$, $d_z = 0.65$, respectively.

5 *Awareness for experienced contingencies.* Participants were also asked to report their
 6 impression of experienced contingencies for each nonword. The accuracy of these responses was
 7 coded for each nonword and corresponded to the factual contingencies in the VCL, which was
 8 then averaged per person. We analyzed mean accuracy scores as a function of instruction (true vs.
 9 false vs. no) in a one-way ANOVA, which yielded a main effect of instruction, $F(2,160) = 22.73$,
 10 $p < .001$, $\eta_p^2 = .22$. Follow-up tests showed that true instructions significantly boosted awareness
 11 for experienced contingencies ($M = .73$, $SD = .33$) compared with no instructions ($M = .56$, SD
 12 $= .34$), $t(80) = 3.46$, $p = .001$, $d_z = 0.38$, $BF_{10}=27$. In turn, false instructions significantly attenuated
 13 awareness for experienced contingencies ($M = .38$, $SD = .34$) compared with no instructions, $t(80)$
 14 $= -3.94$, $p < .001$, $d_z = 0.44$, $BF_{10}=119$. Accuracy scores were significantly better than chance for
 15 true instructions, $t(80) = 6.46$, $p < .001$, $d_z = 0.72$, but worse² than chance for false instructions,

² Note that the awareness measure for experienced contingencies possibly overestimates unawareness, as the measure effectively only had two response options. Thus, errors in the false instruction condition might reflect actual unawareness of experienced contingencies, or correct memory for the instructed contingency, or guessing. We realized this ambiguity of the measure only after the preregistration and therefore decided against further (preregistered) analyses that included awareness of experienced contingencies. In addition, the study put more emphasis on correctly memorizing the instructed contingencies, at least from the perspective of the participants. Immediately after receiving instructions about the specific contingencies, every participant had to pass an instruction check to guarantee that all instructions were correctly remembered. If this criterion was not met, participants were redirected back to reading the

$t(80) = -3.26, p = .002, d_z = -0.36$, respectively, but did not differ from chance for the no instruction condition, $t(80) = 1.79, p = .078, d_z = 0.20$.

Performance in the VCL task. In agreement with the preregistration, erroneous trials (6.8%) or trials that followed an erroneous trial (6.1%) as well as trials with outlier RT (5.3%, i.e. trials with RT below 150 ms or more than 1.5 interquartile ranges above the 75th percentile of the individual RT distribution, Tukey, 1977) were discarded from further RT analyses.

In line with the analytical approach (see Method section), we ran hierarchical multi-level analyses comprising three models in total. Results of the hierarchical analyses are summarized in Table 2.

In Model 1, *Contingency* was the only predictor (contrast coded³ with high contingency = .25 and low contingency = -.75). As expected, it produced a significant CL effect, $b = -6.01, t(37534.3) = -4.33, p < .001$, indicating that participants were 6 ms faster on high contingency trials, compared with low contingency trials.

instructions and were required to complete the instruction check again. This procedure continued until participants showed complete and accurate retention of the instructions. The memory test for instructed contingencies which was assessed at the end of the study documents that participants successfully remembered both instructions equally well, which we interpret as a successful manipulation check.

³ 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: $w_1 = \frac{f_2}{(f_1 + f_2)}$ 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 (see Rudolph & Rothermund, 2025). 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 will be assigned with 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.

In Model 2, we added *Previous Response* as a predictor to account for systematic variance due to stimulus-specific episodic retrieval of previous responses. We coded the relation between the response required in the current trial and the trial in which the last previous nonword occurrence took place (note that only correct previous responses were considered here). The previous response could either match (i.e., previous response = same response as current trial; contrast weight: .3932) or mismatch (i.e., previous response $\neq$ same response as current trial, contrast weight: -.6068; see Figure 1B) with the current correct response. As expected, *Previous Response* was a significant predictor, $b = -12.58$, $t(37534.9) = -9.06$, $p < .001$, indicating that performance was faster when the current response matched (vs. mismatched) the previous response. Against expectations, *Contingency* no longer had a significant effect on trial RT in Model 2, $b = 0.61$, $t(37534.3) = 0.390$, $p = .697$. Tentatively, one could assume that residual CL effects were absent. However, we want to point out that we expected a reversed CL effect for the false instruction condition, which might have neutralized typical CL effects in the no and true instruction conditions. To test this, additional predictors are needed that code for the respective instructed contingency.

In Model 3, further predictors were added: *True Instructions* (TINS, coded as +.5 if true instructions were provided regarding the high contingency nonword-valence pairing, but coded as -.5 if no instructions were provided regarding nonword-valence pairings; false instructions were coded as 0) and its interaction with *Contingency* were entered to the model. Most central for the preregistered hypotheses, the interaction term was significant, $b = -11.58$, $t(37534.2) = -2.953$, $p = .003$, indicating that true instructions boosted the residual CL effect.

Adding *False Instructions* (FINS, coded as +.5 if false instructions were provided regarding the high-contingency nonword-valence pairing and coded as -.5 if no instructions were provided, true instructions were coded as 0) and its interaction with *Contingency* as predictors in

the model also yielded a significant interaction effect, $b = 16.58$, $t(37534.2) = 4.23$, $p < .001$, which is again in line with our expectations. Note that similarly to Rudolph et al. (2025a), the regression weight is *positive*, which indicates that false instructions about the contingency *reduce* the residual CL effect.

Model comparison. To evaluate model fit, we used the difference in log-likelihoods (see Table 2). The change in model fit from Model 1 (including only Contingency) to Model 2 (including *Contingency* and *Previous Response*) was significant, $\chi^2(1) = 81.98$, $p < .001$. This indicates that adding *Previous Response* as predictor significantly improved the overall model fit. Furthermore, the change in model fit was further improved from Model 2 to Model 3 (which added main effects of instructions and their interactions with *Contingency*), reflected in a significant comparison, $\chi^2(1) = 19.20$, $p < .001$. Thus, Model 3 had a significantly better model fit than Model 2.

Follow-up analyses: Residual CL effects within each instruction condition. To gain a better understanding of the interactions ($CL \times TINS$ and $CL \times FINS$, respectively), we analyzed the residual CL effects. First, we ran separate LMMs with the predictors *Contingency* and *Previous Response* for each instruction condition (Table 3). Descriptively, residual CL effects were larger for the true instruction condition (-3 ms) compared to the no instruction condition (-1.6 ms) but were *reversed* in direction in the false instruction condition (+6.5 ms). However, only the latter residual CL effect was statistically significant from zero, $t(12443.0) = 2.413$, $p = .016$. These effects are in line with our expectations and conceptually replicate findings by Rudolph et al. (2025a).

Second, to assess the impact of the instruction condition, we included memory for instructed contingencies for each nonword (remembered vs. forgotten) and its interactions with

Contingency to the models (see Table 4 and Figure 1). Descriptively, for nonwords of the true instruction condition, residual CL effects were stronger for words for which the item-specific contingency instructions were correctly remembered after the VCL (residual CL: -4 ms, $t[9415]=-1.38$, $p=0.16$), compared to those words for which the instructed contingencies had been forgotten (residual CL: 0.6 ms, $t[3065]=-0.11$, $p=0.91$); however, the interaction of Contingency $\times$ Memory for instructed contingencies was not significant, $b = -4.63$, $t(12502.5) = 0.841$, $p = .400$. Likewise, for nonwords of the false instruction condition, residual CL effects were descriptively reversed for participants who remembered the item-specific (false) contingency instructions after the VCL (residual CL: 9.7 ms, $t[8885]=3.12$, $p=0.002$), whereas a (non-significant) standard CL effect was obtained for words without correct memory for the false instructions (residual CL: -1.1 ms, $t[3530]=-0.21$, $p=0.83$); again, however, the interaction of Contingency $\times$ Memory for instructed contingencies also missed conventional levels of significance, $b = 7.14$, $t(12443.0) = 1.349$, $p = .177$.

Exploratory analyses

We tested whether the instruction manipulation influenced evaluative ratings for nonwords and ran a 2 (high contingency valence: positive vs. negative) $\times$ 3 (Instruction: true vs. no vs. false) repeated measures ANOVA. This analysis yielded only a significant valence $\times$ instruction interaction, $F(2, 160) = 16.46$, $p < .001$, $\eta_p^2 = .17$, indicating that EC effects varied as a function of the instruction condition. To decompose this interaction, we computed EC effects as the difference in ratings for nonwords that were mostly paired with positive targets minus ratings for nonwords mostly paired with negative targets ($\Delta EC = \text{positive} - \text{negative}$) for each instruction condition. EC scores were then tested against zero. This yielded significant EC effects for the true instruction condition ($\Delta = 1.63$), $t(80) = 5.55$, $p < .001$, $d_z = 0.62$, but revealed significant *reversed*

EC effects for the false instruction condition ($\Delta = -0.69$), $t(80) = -2.16$, $p = .033$, $d_z = 0.24$. In turn, no significant EC effect emerged for the no instruction condition, $\Delta = -0.21$, $t(80) = -0.83$, $p = .407$, $d_z = 0.09$.

As explicit memory for high contingency-nonword pairings is discussed as a necessary condition for the emergence of EC effects (Baeyens et al., 1990; Pleyers et al., 2007; Stahl & Unkelbach, 2009), we were interested whether the reversal of EC effects hinged on memory for instructed contingencies. Thus, we ran a LMM analysis on evaluative ratings for each nonword (level1 predictor), nested within participants (level2 predictor). We excluded nonwords from the no instruction condition (as no EC effects were obtained for this condition and to reduce complexity of the analysis). As predictors, we used valence of high contingency nonword pairings (positive valence: +0.5, negative valence: -0.5), instruction (true: +0.5, false: -0.5), and memory accuracy for instructed contingencies⁴ (remembered: +0.44, forgotten: -0.56). This analysis yielded a main effect of valence, $b = 0.411$, $t(241.9) = 2.09$, $p = .037$, which was qualified by a valence $\times$ instruction interaction, $b = 2.331$, $t(241.8) = 5.93$, $p < .001$, reflecting the already known modulation of EC by instruction condition, which was also further qualified by a valence $\times$ instruction $\times$ memory for instructed contingencies interaction, $b = 5.277$, $t(299.9) = 5.83$, $p < .001$. No other effect was significant (Table 5). To decompose the three-way interaction, we conducted separate analyses for remembered versus forgotten instructed contingencies, using valence and instruction as predictors. For remembered instructed contingencies, this yielded a significant main effect of valence, $b = 0.486$, $t(190.7) = 2.09$, $p = .038$, which was qualified by the already known valence $\times$ instruction interaction, $b = 3.74$, $t(181.4) = 8.07$, $p < .001$. In turn, neither the main

⁴ In line with the preregistration, we also ran the same analysis with another predictor that coded awareness for experienced contingencies; however, this predictor had no influence, neither on the reported effects, nor on any other effects, which is why we refrain from reporting this analysis in the main text. Interested readers find the corresponding analysis in the open code Markdown document, though.

effect ($|t| < 1$), nor the interaction ($b = -1.44$, $t(66.57) = -1.927$, $p = .0581$) were significant when instructed contingencies were forgotten. Together, these exploratory analyses indicate that (a) EC effects varied as a function of instructed, *not* experienced contingencies and (b) were only obtained if instructed contingencies were correctly remembered at the end of the study. Both findings are consistent with the idea that propositional processes contribute to EC effects (Gast & De Houwer, 2012, 2013; Hütter & De Houwer, 2017).

Discussion

With the present study, we aimed to demonstrate that manipulating contingency instructions affects residual CL effects. This would replicate findings from Rudolph et al. (2025a), but in the context of a valence contingency learning (VCL) task. In line with our expectations, we obtained CL effects in the VCL task (Giesen et al., 2025). These effects were eliminated as soon as we statistically controlled for stimulus-based episodic response retrieval (Model 2). However, in Model 3, residual CL effects were affected by the manipulation of contingency instructions, reflected in significant interaction effects between contingency and the respective instruction contrast. Residual CL effects were descriptively stronger for truly instructed nonword-valence pairings (compared with no instructions). In turn, we obtained a significant *reversed* residual CL effect for falsely instructed nonword-valence pairings (compared with no instructions). These findings closely correspond with findings from Rudolph et al. (2025a). They observed boosted residual CL effects for true instructions but a reversal of residual CL effects for an induction of false contingency beliefs in color-word contingency learning. Importantly, within the present study, memory for instructed contingencies contributed to the effects. We found that the expected effect of instructed contingencies on VCL was obtained when participants memorized the instructed contingencies correctly at the end of the experiment. No significant residual VCL effect was found

1 for words for which the instructions could not be remembered. Overall, this corroborates the
2 conclusion that propositional beliefs had a causal effect on residual CL effects.

3 Another noteworthy finding concerns the impact of instructed contingencies on EC effects.
4 Note that EC effects were not the primary concern of our study. Still, findings are indeed
5 illuminating and nicely match the pattern of findings that was obtained for VCL effects. First, we
6 observed that instructions were sufficient to produce typical EC effects. This is reflected in more
7 positive ratings for nonwords that were mostly paired with positive valence, compared with
8 negative valence, when true instructions were given. Importantly, EC effects were *reversed* if false
9 instructions were given (i.e., more positive ratings for nonwords that were mostly paired with
10 negative valence, compared with negative valence). This implies that valence ratings were
11 predominantly based on instructions, and not on the experienced contingencies (cf. Benedict et al.,
12 2019). Further analyses suggested that EC effects were limited to those cases in which participants
13 correctly remembered the instructed contingencies at the end of the study (for a similar finding,
14 see also Giesen et al., 2025). Together, these findings indicate that EC effects were driven by
15 instructed, not experienced contingencies, which is consistent with other studies documenting EC
16 effects by instruction (Benedict et al., 2019; Gast & De Houwer, 2012, 2013; Hütter & De Houwer,
17 2017). In addition, EC effects were only obtained if instructed contingencies were correctly
18 remembered at the end of the study, which is also consistent with studies attesting to the role of
19 memory processes in EC (Baeyens et al., 1990; Gast, 2018; Pleyers et al., 2007; Stahl & Unkelbach,
20 2009). Both observations are therefore consistent with the idea that propositional processes
21 contribute to EC effects.

22 **Theoretical implications**

1 The present findings bear several important theoretical implications. First and foremost,
2 our findings suggest that CL effects can be understood as the outcome of two independent
3 processes. These are stimulus-based episodic retrieval of the previously executed response and the
4 application of contingency knowledge or propositional beliefs. The first process is documented by
5 the observation that CL effects are reduced or even eliminated as soon as episodic response
6 retrieval is controlled for (either statistically: Giesen et al., 2020; Schmidt et al., 2020; or
7 experimentally by measuring CL effects for distractors that do not have a predecessor in the
8 respective block: Rothermund et al., 2025). The second process is documented by the observation
9 that residual CL effects remain after controlling for episodic retrieval of previous responses
10 (Rudolph et al., 2025a; Xu & Mordkoff, 2020). Importantly, residual CL effects have been shown
11 to be influenced by manipulations that affect propositional knowledge or beliefs. For instance, this
12 is the case for general instructions about upcoming contingencies (Giesen et al., 2025). Another
13 example stems from conveying true or false instructions about specific contingencies between
14 distractors and targets (Rudolph et al., 2025a). The present result pattern for CL effects and for EC
15 effects further corroborates this assumption. Thus, the present study represents a successful
16 conceptual replication of Rudolph et al. (2025a) in the realm of valence contingency learning. In
17 this regard, the false instruction condition yielded particularly relevant findings, because instructed
18 contingencies were in exact *opposition* to participants' experiences. Yet, participants showed
19 reversed CL effects (i.e., they were faster in low frequency nonword-valence combinations,
20 compared with high frequency combinations) and additionally showed reversed EC effects (e.g.,
21 they rated nonwords more positively although these nonwords were in fact more often paired with
22 negative valence responses, and vice versa). Hence, evaluative learning effects investigated in our
23 study did not reflect the actual exposure to nonword-valence pairings but resulted from participants'

1 sustained belief in the instructions they received prior to the task. This observation is consistent
2 with other studies showing that human learning processes are under the influence of propositional
3 beliefs and are often resistant to actual contingencies (e.g., Mitchell et al., 2009; for a detailed
4 discussion, see Rudolph et al., 2025a). Our findings highlight that this conclusion not only holds
5 for contingency learning, but also holds for evaluative learning of nonword-valence pairings in
6 humans, which are then reflected in CL and EC effects.

7 *The relationship between (valence) contingency learning effects and EC effects*

8 In our view, valence contingency learning and EC are related as they both represent forms
9 of evaluative learning. It is also compelling that both result from the same procedure (if one
10 considers the VCL as learning phase and the rating task as test phase). Still, there are clear
11 discrepancies between both phenomena: For instance, CL is measured on-task, whereas EC is
12 measured off-task (but see Richter & Giesen, 2025). CL is typically attributed to stimulus-response
13 co-occurrences (explicitly tested by Schmidt et al., 2007), whereas EC is attributed to stimulus-
14 stimulus (CS-US) co-occurrences (De Houwer, 2007), although there is evidence for EC that is
15 based on SR co-occurrences (Gast & Rothermund, 2011; Blask et al., 2016). Furthermore, whereas
16 awareness is currently discussed as a prerequisite of EC, this is not the case for CL, as CL effects
17 can emerge independently of contingency awareness (Schmidt et al., 2007; Schmidt, 2021). The
18 present approach, which introduces two possible processes that might underlie CL, explains the
19 latter discrepancy, as CL without item awareness seems to be due to retrieval of SR episodes (i.e.,
20 an automatic process that operates independently of awareness). In principle, awareness of
21 nonword-valence co-occurrences can then independently contribute to CL effects (above and
22 beyond SR retrieval processes, cf. Giesen et al., 2025). Yet the exact nature of these representations
23 is currently unclear, as nonword-valence co-occurrences could represent evaluations of the CS.

Alternatively, they could represent knowledge that a specific CS signals a positive or negative response (Gast & Rothermund, 2011; Blask et al., 2016). Future research is needed to follow up on these open issues.

Conclusions

The present study represents a conceptual replication of Rudolph et al. (2025a), but in the context of valence contingency learning. We manipulated participants' beliefs about contingencies in the VCL task. Participants received either true, false or no instructions regarding the actual nonword-valence contingencies in the VCL task. Effects of contingency learning (CL) and evaluative conditioning (EC) for nonwords were assessed. As expected, contingency beliefs modulated contingency learning. True instructions boosted residual CL effects, whereas false instructions reduced residual CL effects. If participants memorized the falsely instructed contingencies, the residual CL effect was reversed. Exploratory analyses also revealed a modulatory influence of contingency beliefs on EC effects, which varied solely as a function of (remembered) contingency instruction and were unaffected by experienced contingencies. Together, the modulation of both, residual CL effects and EC effects by instructed contingencies highlights the dominant role of propositional beliefs in producing these evaluative learning effects.

Acknowledgements

We thank Amelie Hespeler, Florine Güttler, Pia Hölken, Annika Nigrin, Jakob Schmidt, and Lene Weinert for their help in data collection, and Sophie Pesch for her help in manuscript preparation. The study was funded by grants from the German Research Foundation to Carina G. Giesen (GI 1295/2-1,2).

Declaration of Interests

The authors report there are no competing interests to declare.

References

- Baeyens, F., Eelen, P., & van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition & Emotion*, 4(1), 3–18. <https://doi.org/10.1080/02699939008406760>
- Blask, K., Frings, C., & Walther, E. (2016). Doing is for feeling. *Journal of Experimental Psychology: General*, 145(10), 1263–1268. <https://doi.org/10.1037/xge0000211>
- Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*, 23(2), 161-189. <https://doi.org/doi.org/10.1177/1088868318763>
- De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, 10(2), 230–241. <https://doi.org/10.1017/S1138741600006491>
- De Houwer, J., & Hughes, S. (2020). *The psychology of learning: An introduction from a functional-cognitive perspective*. The MIT Press. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2020-31499-000&site=ehost-live>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Foerster, A., Rothermund, K., Parmar, J. J., Moeller, B., Frings, C., & Pfister, R. (2021). Goal-Based Binding of Irrelevant Stimulus Features for Action Slips. *Experimental Psychology*, 68(4), 206–213. <https://doi.org/10.1027/1618-3169/a000525>
- Gast, A. (2018). A declarative memory model of Evaluative Conditioning. *Social Psychological Bulletin*, 13(3), Article e28590. <https://doi.org/10.5964/spb.v13i3.28590>

- Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *The Quarterly Journal of Experimental Psychology*, 65(9), 1657–1674.
<https://doi.org/10.1080/17470218.2012.665061>
- Gast, A., & De Houwer, J. (2013). The influence of extinction and counterconditioning instructions on evaluative conditioning effects. *Learning and Motivation*, 44(4), 312–325. <https://doi.org/10.1016/j.lmot.2013.03.003>
- Gast, A., Richter, J., & Ruzspel, B. (2020). Is there evidence for unaware evaluative conditioning in a valence contingency learning task? *Cognition and Emotion*, 34(1), 57–73. <https://doi.org/10.1080/02699931.2019.1652147>
- Gast, A., & Rothermund, K. (2011). I like it because I said that I like it: Evaluative conditioning effects can be based on stimulus-response learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 37(4), 466–476.
<https://doi.org/10.1037/a0023077>
- Gawronski, B., & Bodenhausen, G. V. (2009). Operating principles versus operating conditions in the distinction between associative and propositional processes. *Behavioral and Brain Sciences*, 32, 207-208.
<https://doi.org/10.1017/S0140525X09000958>
- Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative–propositional evaluation model. *Social and Personality Psychology Compass*, 8(8), 448-462. <https://doi.org/10.1111/spc3.12124>
- Giesen, C. G., Duderstadt, H., Richter, J., & Rothermund, K. (2025). Dissociating the roles of episodic retrieval and contingency awareness in valence contingency learning. *Cognition & Emotion*, 1–17. <https://doi.org/10.1080/02699931.2025.2456608>

- 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), Article 2927, 2927. <https://doi.org/10.3389/fpsyg.2019.02927>
- Hütter, M., & Houwer, J. de (2017). Examining the contributions of memory-dependent and memory-independent components to evaluative conditioning via instructions. *Journal of Experimental Social Psychology*, 71, 49–58. <https://doi.org/10.1016/j.jesp.2017.02.007>
- 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>
- Parmar, J., Foerster, A., Pfister, R., & Rothermund, K. (2022). Frankly, My Error, I Don't Give a Damn: Retrieval of Goal-Based but Not Coactivation-Based Bindings after Erroneous Responses. *Journal of Cognition*, 5(1), 34. <https://doi.org/10.5334/joc.224>
- Parmar, J., & Rothermund, K. (2024). Nothing else matters: Stimulus-response binding and retrieval is independent of affective consequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 50(3), 362–382. <https://doi.org/10.1037/xlm0001288>
- Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 33(1), 130–144. <https://doi.org/10.1037/0278-7393.33.1.130>

Richter, J., & Giesen, C. G. (2025). The role of episodic retrieval in evaluative conditioning:

Evaluative conditioning effects differ depending on the temporal distance to the last stimulus pairing. *Cognition & Emotion*, 1–18.

<https://doi.org/10.1080/02699931.2025.2481108>

Rothermund, K., Kapinos, L., Houwer, J. de, & Schmidt, J. R. (2025). Long-term

Contingency Learning Depends on Contingency Awareness. *Journal of Cognition*, 8(1), 23. <https://doi.org/10.5334/joc.433>

Rudolph, M., Giesen, C. G., & Rothermund, K. (2025a). False contingency knowledge

reverses the color-word contingency learning effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 51(7), 1023–1033.

<https://doi.org/10.1037/xlm0001413>

Rudolph, M., Giesen, C. G., & Rothermund, K. (2025b). Stimulus-response binding and

retrieval operates independently of contingency awareness: A mega-analysis. *Consciousness and Cognition*, 135, 103936.

<https://doi.org/10.1016/j.concog.2025.103936>.

Rudolph, M., & Rothermund, K. (2025). Two sources of color-word contingency learning:

Episodic retrieval of stimulus-response bindings and propositional knowledge. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 51(2), 209–217.

<https://doi.org/10.1037/xlm0001353>

Schmidt, J. R. (2021). Incidental learning of simple stimulus-response associations: A review

of colour-word contingency learning research. *Année Psychologique*, 121, 77–127.

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning

without awareness: Evidence for implicit control. *Consciousness and Cognition: An International Journal*, 16(2), 421–435. <https://doi.org/10.1016/j.concog.2006.06.010>

Schmidt, J. R., & De Houwer, J. (2012a). Contingency learning with evaluative stimuli:

Testing the generality of contingency learning in a performance paradigm.

Experimental Psychology, 59(4), 175–182. <https://doi.org/10.1027/1618-3169/a000141>

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

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.

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>

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology. Animal Behavior Processes*, 35(2), 286–291. <https://doi.org/10.1037/a0013255>

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

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

Appendix

German (English translation) target words, grouped by valence, that were used during practice vs. experimental trials.

Phase	Positive		Negative	
practice	Hoffung (hope)	Lachen (laughter)	Armut (poverty)	Diebstahl (thievery)
	Freiheit (freedom)	Respekt (respect)	Betrug (fraud)	Koma (coma)
	Sommer (summer)	Erfolg (success)	Kummer (grief)	Sadist (sadist)
	Humor (humor)	Treue (loyalty)	Parasit (parasite)	Verlust (loss)
	Genuss (enjoyment)	Hochzeit (wedding)	Abscheu (disgust)	Gestank (stench)
Main experiment	Geschenk (gift)	Leben (life)	Bomben (bombs)	Krieg (war)
	Traum (dream)	Musik (music)	Drogen (drugs)	Mord (murder)
	Frieden (peace)	Jubel (joy)	Tod (death)	Tumor (tumor)
	Geburt (birth)	Umarmung (hug)	Gewehre (guns)	Unfall (accident)
	Blumen (flowers)	Liebe (love)	Krankheit (sickness)	Verbrechen (crime)
	Freund (friend)	Urlaub (holiday)	Krebs (cancer)	Virus (virus)

Tables

Table 1

Examples for experimental design and instructed and experienced contingencies in the valence classification task.

Instruction Condition	Nonword	Specific instruction	Experienced Contingency	
			Positive target	Negative target
No instruction	kadirga [nonword1]	-	3 (hc)	1 (lc)
	lokanta [nonword2]	-	1 (lc)	3 (hc)
True instruction	nijaron [nonword3]	mostly precedes positive target words	3 (hc)	1 (lc)
	fevkani [nonword4]	mostly precedes negative target words	1 (lc)	3 (hc)
False instruction	marenes [nonword5]	mostly precedes negative target words	3 (hc)	1 (lc)
	gowange [nonword6]	mostly precedes positive target words	1 (lc)	3 (hc)

Note. Hc=high contingency; lc = low contingency. For a given participant, the specific nonwords (kadirga, lokanta, etc.) were randomly assigned to each nonword variable (represented in brackets). Participants pressed a left or right key to classify the valence of target words.

Table 2

Results of a hierarchical multi-level regression analysis, predicting trial RT based on trial contingency (high vs. low), stimulus-based episodic retrieval of previous responses (match vs. mismatch) and instructions (coded as true vs. no and false vs. no) and their interactions with contingency.

	Model 1			Model 2			Model 3		
<i>Predictors</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
(Intercept)	588.06	7.40	< 0.001	588.10	7.40	< 0.001	588.11	7.40	< 0.001
Contingency (C)	-6.01	1.39	< 0.001	0.61	1.57	0.697	0.60	1.57	0.701
Previous Response				-12.58	1.39	< 0.001	-12.57	1.39	< 0.001
TINS [true vs. no]							0.62	1.69	0.714
FINS [false vs. no]							0.41	1.69	0.811
C × TINS							-11.58	3.92	0.003
C × FINS							16.58	3.92	< 0.001
Random Effects									
σ^2	13490.83			13461.41			13454.52		
τ_{00}	4409.55 ID			4406.53 ID			4407.67 ID		
ICC	0.25			0.25			0.25		
N	81 ID			81 ID			81 ID		
Observations	37615			37615			37615		
Marginal R ² / Conditional R ²	0.000 / 0.247			0.002 / 0.248			0.002 / 0.249		
log-Likelihood	-232431.836			-232390.844			-232381.245		

Note. SE=standard error. TINS=predictor contrasting true vs. no instruction; FINS=predictor contrasting false vs. no instruction. Regression weights (b) reflect the difference in milliseconds between the conditions that define a contrast. *** $p < .001$.

Table 3

Test for residual CL effects separately for each instruction condition: Results of multi-level regression analyses, predicting trial RT based on trial contingency (high vs. low) and stimulus-based episodic retrieval of previous responses (match vs. mismatch).

<i>Predictors</i>	No Instruction			True Instruction			False Instruction		
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
(Intercept)	587.59	7.37	<0.001	588.46	7.54	<0.001	588.32	7.49	<0.001
cont	-1.58	2.72	0.561	-3.16	2.73	0.247	6.51	2.70	0.016
prevRESP	-13.16	2.40	<0.001	-16.24	2.42	<0.001	-7.88	2.40	0.001
Random Effects									
σ^2	13411.09			13585.71			13302.64		
τ_{00}	4312.68 _{ID}			4515.73 _{ID}			4457.13 _{ID}		
ICC	0.24			0.25			0.25		
N	81 _{ID}			81 _{ID}			81 _{ID}		
Observations	12509			12583			12523		
Marginal R ² / Conditional R ²	0.003 / 0.245			0.004 / 0.253			0.001 / 0.251		
AIC	154697.118			155776.285			154771.607		

Note. SE=standard error. Regression weights (b) reflect the difference in milliseconds between predictor levels.

Table 4

Results of multi-level regression analyses predicting trial RT within each instruction condition based on trial contingency (high vs. low), stimulus-based episodic retrieval of previous responses (match vs. mismatch), and memory for instructed contingencies (remembered vs. forgotten) and its interaction with contingency.

<i>Predictors</i>	True Instruction			False Instruction		
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>
(Intercept)	588.46	7.51	<0.001	588.32	7.49	<0.001
Contingency (C)	-3.16	2.73	0.248	6.50	2.70	0.016
Previous Response	-16.23	2.42	<0.001	-7.87	2.40	0.001
Memory for instructed contingency (M)	-4.13	3.94	0.295	1.09	3.38	0.746
C × M	-4.63	5.59	0.407	7.14	5.29	0.177
Random Effects						
σ^2	13584.48			13300.51		
τ_{00}	4478.79 _{ID}			4460.64 _{ID}		
ICC	0.25			0.25		
N	81 _{ID}			81 _{ID}		
Observations	12583			12523		
Marginal R ² / Conditional R ²	0.004 / 0.251			0.001 / 0.252		
AIC	155768.647			154764.240		

Note. SE=standard error. Regression weights (b) reflect the difference in milliseconds between predictor levels.

Table 5

Results of a multi-level regression analysis evaluations of nonwords based on high contingency valence (positive vs. negative), Instruction condition (true vs. false), and memory for instructed contingencies (remembered vs. forgotten) and their interactions

<i>Predictors</i>	<i>b</i>	<i>SE</i>	<i>p</i>
(Intercept)	4.24	0.11	<0.001
High contingency valence (V)	0.41	0.20	0.037
Instruction (I)	0.05	0.20	0.780
Memory for instructed contingency (M)	-0.09	0.23	0.694
V×I	2.33	0.39	<0.001
I × M	0.23	0.46	0.621
V × M	0.28	0.46	0.537
V× I × M	5.28	0.91	<0.001
Random Effects			
σ^2	3.10		
τ_{00} ID	0.17		
ICC	0.05		
N ID	81		
Observations	324		
Marginal R ² / Conditional R ²	0.185 / 0.228		
AIC	1326.168		

Note. SE=standard error. Regression weights (b) reflect the difference in valence rating between predictor levels; higher scores reflect more positive ratings for nonwords that were paired more often with positive valence.

Figure 1

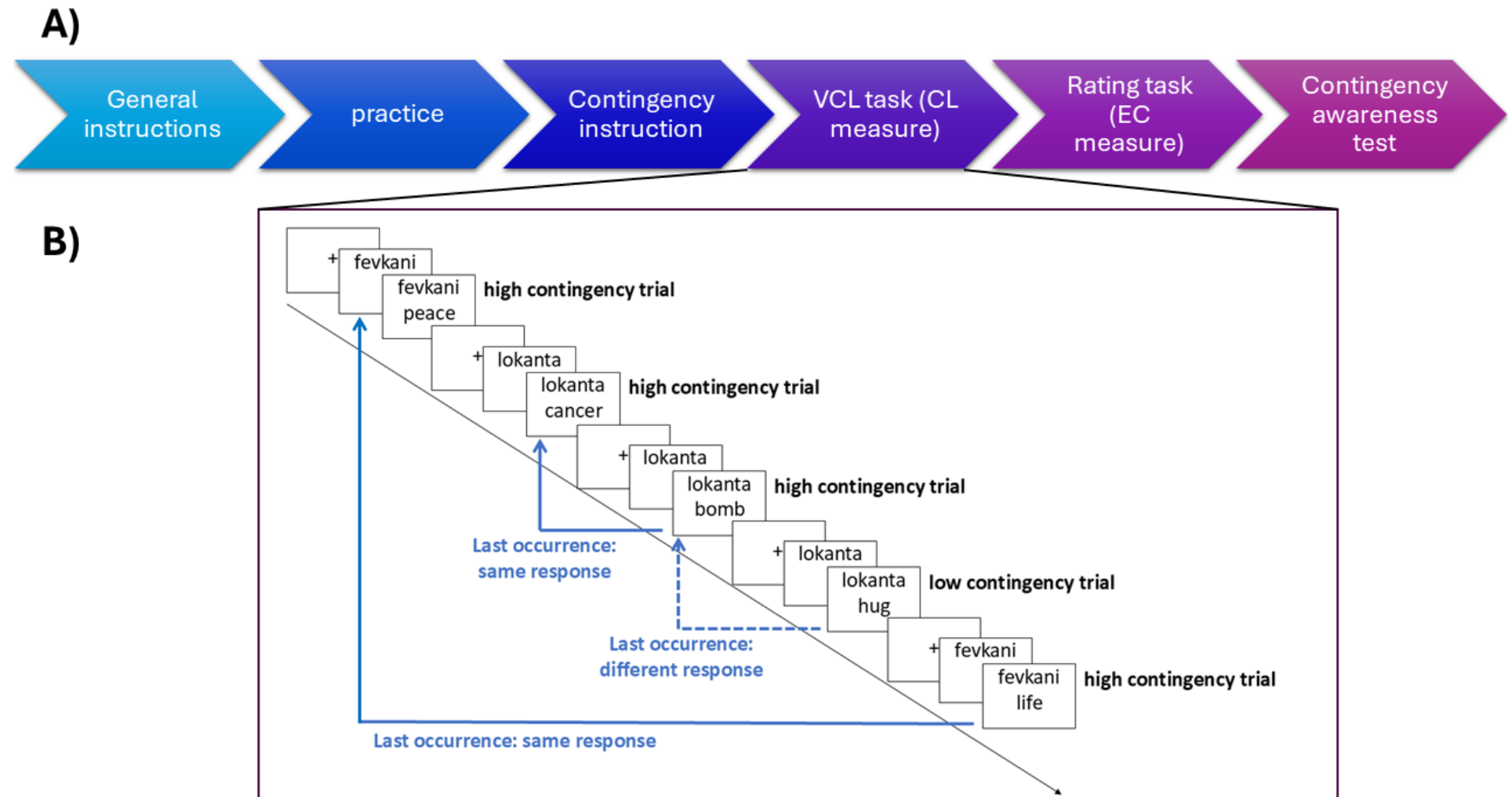


Figure 1. Schematic overview of the experiment, trial structure, and logic of the LMM analysis. (A) Overview of the different experimental phases. VCL = Valence Contingency learning task. CL = Contingency learning. EC= Evaluative Conditioning. (B) Example of trial sequence in the valence contingency learning task. For illustrative purposes and due to space limits, blank trials are omitted; nonwords appeared either above or below the fixation (omitted from figure). Blue arrows illustrate the logic of the LMM analysis that controls for episodic retrieval processes by tracing each nonword presentation back to the last previous occurrence, which either required the same response (solid lines) or a different response (dashed lines).

**False contingency beliefs reverse contingency learning effects in the valence contingency
learning task**

Carina G. Giesen, Matthäus Rudolph & Klaus Rothermund

Supplementary analyses

Serial position analyses

As suggested by an anonymous reviewer, we ran additional analyses of contingency learning (CL) effects and analyzed potential influences of serial position of trials. According to Schmidt et al. (2007), CL effects emerge early and can be detected after only a few trials, even when participants receive no explicit instructions about the underlying contingencies. This may be due to episodic retrieval processes that contribute to the formation of CL effects at an early stage, that is, before contingency knowledge is established. We therefore performed additional analyses to follow the development of CL over the course of the experiment. Rather than serial position of individual trials, we computed a block factor, because CL effects are computed as a difference of low vs high contingency trials. The block factor comprised of 48 trials, resulting in 12 blocks in total per participant. We ran the analysis on the overall dataset (including stimuli from all instruction conditions). However, besides producing a strong main effect (indicating that participants become faster over the course of the task), the block predictor did not interact with the contingency predictor, nor did we find higher order interactions with the contrasts coding true or false instructions, all $|t| < 1.11$, all $p > .26$.

The present analyses therefore converge with the observation by Schmidt et al. (2007) that CL effects emerge early in the task. The following plots illustrate this, as CL effects are present already for block 1 in the overall data set (Figure S1). Note that high contingency trials are mostly faster than the low contingency trials, Importantly, this is already the case for Block 1 and 2, i.e. very early on in the task.

Figure S1

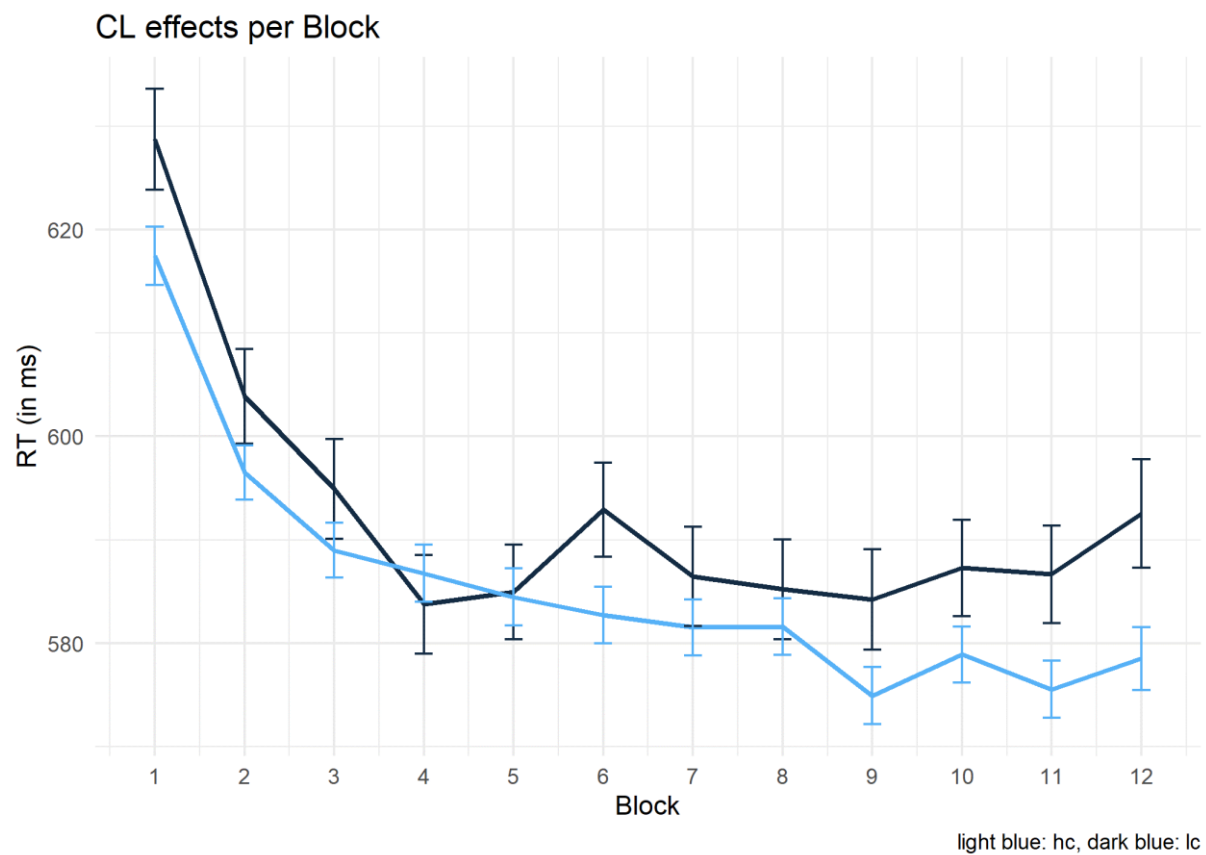


Figure S1. Reaction times (RT, in ms) as a function of contingency level (high, low) and block. Light blue lines depict high contingency trials, dark blue lines depict low contingency trials. Error bars reflect mean standard errors.

The same holds true for the no instruction condition (Figure S2) and for the true instruction condition (Figure S3).

Figure S2

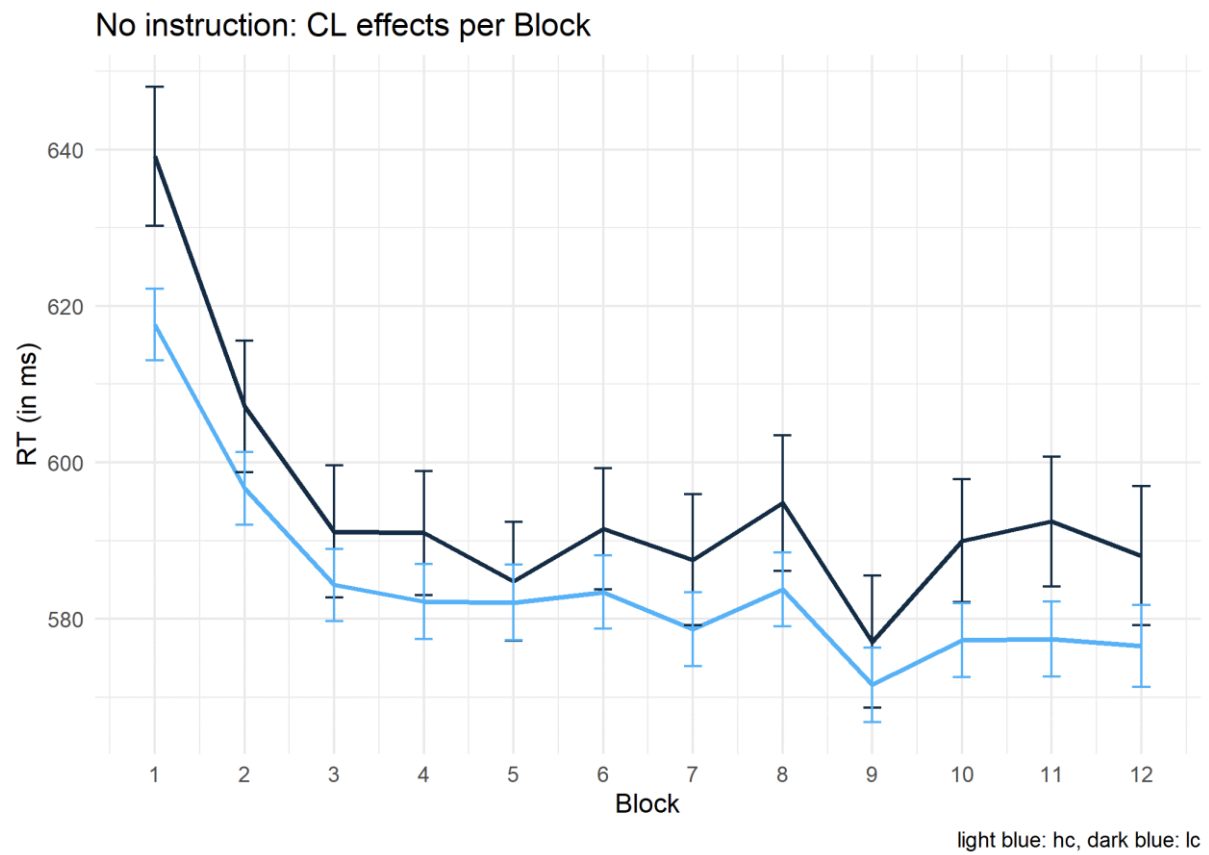


Figure S2. Reaction times (RT, in ms) as a function of contingency level (high, low) and block for the no instruction condition. Light blue lines depict high contingency trials, dark blue lines depict low contingency trials. Error bars reflect mean standard errors.

Figure S3

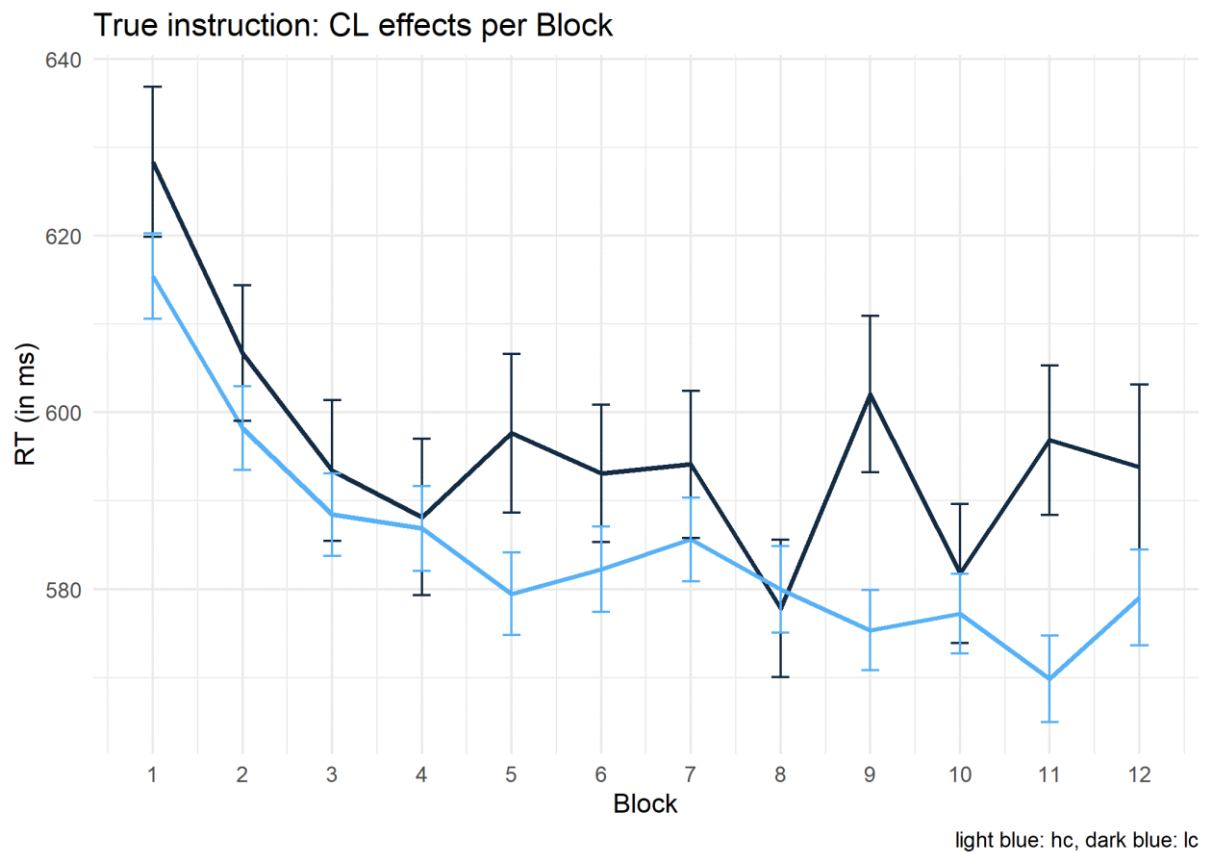


Figure S3. Reaction times (RT, in ms) as a function of contingency level (high, low) and block for the true instruction condition. Light blue lines depict high contingency trials, dark blue lines depict low contingency trials. Error bars reflect mean standard errors.

Figure S4

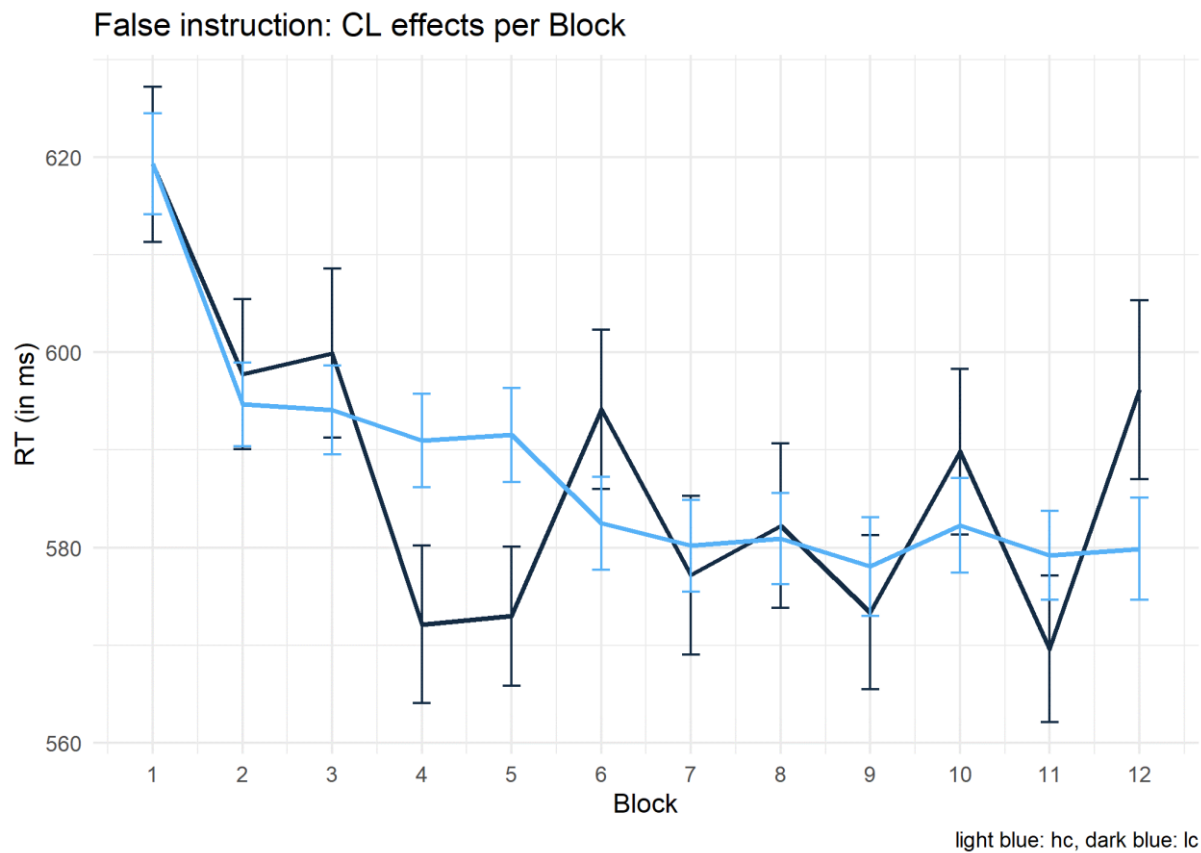


Figure S4. Reaction times (RT, in ms) as a function of contingency level (high, low) and block for the false instruction condition. Light blue lines depict high contingency trials, dark blue lines depict low contingency trials. Error bars reflect mean standard errors.

For the false instruction condition, the course of CL effects is somewhat different (Figure S4): As the data of the false instruction condition already indicate that - on average – participants do not change their expectancy as a function of actual experience. Otherwise, it would be unclear why CL effects are reversed for the false instruction condition when we look at the averaged data. In fact, the block x CL interaction is not significant for the false instruction condition. However, the pattern appears more irregular compared to the plots for the no and correct instruction conditions, indicating that both instructions and experienced contingencies

influence the CL effect (especially in blocks 4 and 5, participants seem to rely on the falsely instructed contingencies, possibly because they encountered some nonword-valence combinations [low contingency trials] that matched and possibly confirmed their false knowledge). However, the fluctuations do not follow a systematic time-based pattern, so the findings seem to suggest that there is a lasting effect of the false instructions for a substantial percentage of participants in the sample.

