American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

0278-7393/20/$12.00

2020, Vol. 46, No. 2, 327–349
http://dx.doi.org/10.1037/xlm0000724

Stroop-Like Effects of Derived Stimulus–Stimulus Relations

Baptist Liefooghe and Sean Hughes
Ghent University

James R. Schmidt
University of Bourgogne Franche-Comte

Jan De Houwer
Ghent University

Automaticity can be established by consistently reinforcing contingencies during practice. During reinforcement learning, however, new relations can also be derived, which were never directly reinforced. For instance, reinforcing the overlapping contingencies A → B and A → C, can lead to a new relation B-C, which was never directly reinforced. Across 5 experiments we investigated if such derived relations can also induce automatic effects. We first trained participants to derive a relation between a nonsense word and a color word, and then used the nonsense words as distractors in a Stroop task. Results indicate that derived color-word associates induce Stroop effects. This effect, however, is present only when sufficient attention is allocated to the distractor words during the Stroop task, and is driven by a response conflict. We conclude that, under the present training conditions, derived color–word associates became related to the corresponding color word at the lexical level, but did not gain direct access to the corresponding semantic color representation.

Keywords: derived stimulus relating, automaticity, Stroop effects, language learning

Automaticity occupies center stage in many areas of psychology (e.g., Moors, 2016; Moors & De Houwer, 2006). Although recent work suggests that automaticity is a multifaceted concept (e.g., Bargh, 1994; Moors, 2016; Moors & De Houwer, 2006 for elaborate discussions), it is often conceived of as fast and uncontrolled behavior, which is unintentionally performed. Automaticity develops over practice. This was, for instance, demonstrated by Shiffrin and Schneider (1977), who presented training phases in which stimuli were either consistently related to the same response, or inconsistently related to different responses. In a subsequent test phase, a different response mapping was imposed and performance deteriorated for stimuli that were consistently related to a particular response in the training phase, compared with stimuli that were inconsistently related to different responses. This basic finding has now been elaborated in many ways (for examples, see Moutsopoulou, Yang, Desantis, & Waszak, 2015; Pfeuffer, Moutsopoulou, Pfister, Waszak, & Kiesel, 2017; Verbruggen & Logan, 2008; Waszak, Hommel, & Allport, 2003), thus emphasizing the importance of taking learning history into account when studying automaticity. Such a learning history can be experimentally induced, as in the aforementioned example, but is evidently also present prior to an experimental session. A classic example is the Stroop task (Stroop, 1935; for reviews, see Kalanthroff, Davelaar, Henik, Goldfarb, & Usher, 2018; MacLeod, 1991), in which participants need to identify the print color of a color word while ignoring the meaning of the word itself (e.g., say "green" to the word RED printed in green). The congruency or Stroop effect is the observation that people are typically slower and less accurate in responding to *incongruent* trials where the meaning of the word and color mismatch (e.g., the word GREEN printed in yellow; $\text{GREEN}_{\text{yellow}}$), relative to *congruent* trials, where the meaning of the word and color match (e.g., $\text{GREEN}_{\text{green}}$). No explicit training phase is present in most Stroop studies. Instead, the automatic effect triggered by the color words is based on an extensive learning history that participants have in reading words (e.g., Jensen & Rohwer, 1966; MacLeod, 1991; see also Liefooghe & De Houwer, 2016 for a discussion).

Although the idea that automaticity depends on practice is self-evident, research on automaticity has mainly focused on automatic effects in the context of contingencies that were *directly* reinforced during practice (e.g., MacLeod & Dunbar, 1988), for instance, by rewarding the pairing of two stimuli using error feedback (e.g., Schmidt, Crump, Cheesman, & Besner, 2007). Yet, a rich vein of research in the reinforcement-learning literature suggests that directly reinforcing some set of contingencies also leads to the emergence of new contingencies, which were never directly reinforced in the past. Such *derived learning* has been investigated by using the matching-to-sample procedure (MTS; e.g., Sidman & Tailby, 1982). During a MTS procedure, partici-

---

Baptist Liefooghe and Sean Hughes, Department of Experimental Clinical and Health Psychology, Ghent University; James R. Schmidt, LEAD-CNRS UMR 5022, University of Bourgogne Franche-Comte; Jan De Houwer, Department of Experimental Clinical and Health Psychology, Ghent University.

This research was supported by grant BOF16/MET_V/002 of Ghent University to Jan De Houwer. Raw data and corresponding analysis scripts of all experiments reported are deposited at https://osf.io/q7ckd/.

Correspondence concerning this article should be addressed to Baptist Liefooghe, Department of Experimental Clinical and Health Psychology, Ghent University, H. Dunantlaan 2, B-9000 Ghent, Belgium. E-mail: baptist.liefooghe@ugent.be

pants are required to perform conditional discriminations by selecting one of two *comparison stimuli* when a *conditional (or sample) stimulus* is presented (see Figure 1 for an illustration). Consider, for instance, the following stimuli “RED,” “GREEN,” “PLESK,” and “KLAMF.” “RED” and “GREEN” are used as comparison stimuli and “PLESK” and “KLAMF” are used as sample stimuli. Participants are tasked with selecting one of the two comparison stimuli, depending on the identity of the sample stimuli. For instance, participants may be reinforced via error feedback to select the comparison stimulus “RED” when the sample stimulus “PLESK” is presented and the comparison stimulus “GREEN” when the sample stimulus “KLAMF” is presented (i.e., PLESK→ RED and KLAMF→ GREEN).

Following such *conditional-discrimination training*, the MTS procedure is presented a second time to assess the extent to which training resulted in the formation of new relations. Sample and comparison stimuli are once again presented and participants perform the conditional-discrimination task without any error feedback. Typically, responses in this *conditional-discrimination test* are in line with the previously trained contingencies. For example, the comparison stimulus “RED” will be selected more often when presenting the sample stimulus “PLESK” and the comparison stimulus “GREEN” will be selected more often when presenting the sample stimulus “KLAMF.” However, when reversing sample and comparison stimuli (i.e., “RED” and “GREEN” now serve as samples; “PLESK” and “KLAMF” as comparisons), the comparison stimulus “PLESK” will be selected more often when presenting the sample stimulus “RED” and, likewise, the comparison stimulus “KLAMF” will be selected more often when presenting the sample stimulus “GREEN.” Whereas the contingencies “PLESK” → “RED” and “KLAMF” → “GREEN” were directly reinforced during training, two additional relations emerge: “RED” → “PLESK” and “GREEN” → “KLAMF.” Put another way, contingencies trained in one direction lead to relations derived in the opposite direction.

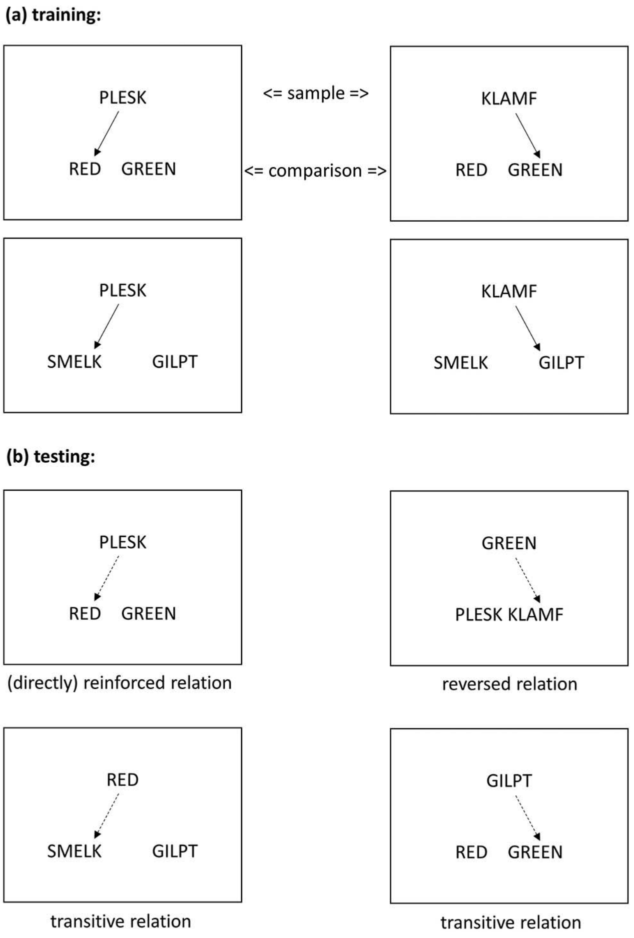

Figure 1. Illustration of a conditional-discrimination training and test. Solid arrows point toward the reinforcement of contingencies during training. Dashed arrows point toward the response pattern that is most likely observed during testing, without reinforcement. Besides responding in accordance to the directly reinforced contingencies, responses will also be based on derived relations, namely: reversed and transitive relations.

The above type of training can be extended by adding two new sets of overlapping contingencies, which are directly reinforced (see Figure 1). For instance, “PLESK” → “RED” and “KLAMF” → “GREEN,” on the one hand, and “PLESK” → “SMELK” and “KLAMF” → “GILPT,” on the other hand. In this case, both sets of contingencies share the same sample stimuli (“PLESK” and “KLAMF”), but involve different comparison stimuli (“RED” and “GREEN” vs. “SMELK” and “GILPT”). During a subsequent test phase in which no reinforcement is provided, responses in line with the contingencies that were previously reinforced (e.g., “KLAMF” → “GREEN”) as well as reversed responding will be observed (e.g., “GREEN” → “KLAMF”). In addition, when combining comparison stimuli of both sets contingencies by using one set of comparison stimuli as sample stimuli (e.g., sample stimuli “RED,” “GREEN”; comparison stimuli: “GILPT,”“SMELK”), the comparison stimulus “SMELK” will be more likely selected when presenting the sample stimulus “RED” and comparison stimulus “GILPT” is more likely to be selected when presenting the sample stimulus “GREEN.” The direct reinforcement of partially overlapping contingencies in the training phase thus results in the formation of several new relations which were never directly reinforced, namely: “RED” → “SMELK,” “SMELK” → “RED,” “GREEN” → “GILPT,” “GILPT” → “GREEN.”

When these different emergent or derived relational responses are observed the stimuli involved are said to participate in an *equivalence relation* (e.g., people “act as if” “GREEN,” “KLAMF,” and “GILPT” are equivalent in some respect) or belong to *equivalence classes* of stimuli (Sidman, 2000, 2009). Sidman (e.g., Sidman, 1990, 1994, 1997) related stimulus equivalence to the presence of three mathematical relations: reflexivity, symmetry, and transitivity. *Reflexivity* indicates that each stimulus is conditionally related to itself (i.e., “if *a*, then *a*”). *Symmetry* requires that the relation between stimuli be reversible (e.g., training “if *a*, then *b*” results in “if *b*, then *a*”). Finally, *transitivity* refers to the fact that when a relation between two stimuli (“if *a*, then *b*”) is recombined with a second relation containing one of those stimuli and a novel stimulus (“if *b*, then *c*”), a novel relation emerges (“if *a*, then *c*”).

A classic demonstration of derived stimulus relating was provided by Sidman and Tailby (1982). The authors trained analphabetic children to select pictures of objects or animals when hearing the corresponding word (e.g., point to the picture of a dog [comparison] in the presence of the word "dog" [sample]). The children were also trained to select written words when hearing the corresponding word (e.g., point to the written word "dog" [comparison] in the presence of the spoken word "dog"). Following such training, the children were not only able to apply the reinforced contingencies, but also to relate the written words to their pictorial counterparts and vice versa. In other words, new relations were derived which were not directly reinforced during training. Since Sidman and Tailbly's (1982) observations, this phenomenon has been extensively documented (e.g., Barnes & Holmes, 1991; Cullinan, Barnes, Hampson, & Lyddy, 1994; Hayes & Hayes, 1992; Sidman, 1986; Wulfert & Hayes, 1988, see Barnes-Holmes et al., 2004; Hughes & Barnes-Holmes, 2016 for reviews) and applied in various educational programs (e.g., de Rose, de Souza, Rossito, & de Rose, 1992).

Derived stimulus–stimulus relations thus indicate that reinforcement learning can lead to more than the emergence of directly reinforced responses. If we combine this idea with our original premise, namely that automaticity is a function of practice, then a new question arises: Can derived stimulus–stimulus relations also induce automatic effects? The present study tackles this issue. To this end, we implemented the aforementioned example on conditional discrimination. More precisely, we used conditional-discrimination training to establish two classes of equivalent stimuli. Each class consisted of a color word and two nonwords (e.g., "GREEN," "KLAMF," "GILPT"; "RED," "SMELK," "PLESK"). Whereas the relation between the color word and the first nonword was directly reinforced via error feedback (i.e., *reinforced associate*), the conditional-discrimination training was constructed in such a way that the relation between the color word and the second nonword emerged via derivation (i.e., *derived associate*). Following training we then presented a modified Stroop task which included the color words and their associates as distractors.

Within the aforementioned procedure, our main question was whether the nonwords which were related to the color words through reinforcement or derivation could also induce a Stroop effect. Previous work has shown that directly reinforced contingencies can trigger a congruency effect when incorporated into a Stroop-like task. For instance, MacLeod and Dunbar (1988) trained participants to name a particular color when a specific shape was presented (e.g., triangle → yellow). After extensive training, these contingencies triggered a Stroop effect when participants were required to name the color in which the shapes were printed. More recently, Schmidt, Crump, Cheesman, and Besner (2007) let participants identify the color of a neutral distractor word (e.g., MOVE), which was presented in a particular color on 75% of the trials. Color identification was faster when the words appeared in the color they were most often presented in relative to when they appeared in another color. Such a finding again indicates that the direct reinforcement of arbitrary contingencies between a color and a color word, leads to automatic response biases. Finally, Geukes, Gaskell, and Zwitserlood (2015; see also Richards & Blanchette, 2004 for a similar approach) presented participants with word–word pairs, each consisting of a color word and a nonword. Some of these pairs were presented more frequently than others and participants thus learned to relate a nonword to a particular color word. These nonwords triggered a congruency effect when used as distractors in a Stroop task. Given these past findings, we expected to observe Stroop effects for directly reinforced associates. Yet, to our knowledge, no evidence is available suggesting that derived associates can also trigger Stroop effects. With this in mind, we set out to provide a first test of the idea that derived stimulus-stimulus relations lead to automatic effects as measured by the Stroop task.

Experiment 1

In Experiment 1, we used conditional-discrimination training to establish two equivalence classes, each consisting of a color word, a reinforced associate, and a derived associate. Following this training, we administered two tests. First, we assessed if training led to the formation of two equivalence classes by using a test in which we probed for the presence of symmetrical (i.e., is SMELK more likely to be selected when RED is presented, following the training of the contingency SMELK → RED) and transitive responding (i.e., is GILPT more likely to be selected when RED is presented, following the training of the contingencies SMELK → RED and SMELK → GILPT) in the absence of reinforcement. Second, we administered a modified Stroop task which included the color words and their associates as distractors. We had two questions: (a) Did training led to the formation of two equivalence classes?; and (b) Do reinforced and associates also induce a Stroop effect?

Method

Participants. Fifty-seven students at Ghent University participated in exchange for 10 euro. Participants were naive to the purposes of the experiment. For each experiment we aimed for a minimal sample of 32 participants, which is advisable for detecting a medium-sized effect with a power of .80.

Materials.

Conditional-discrimination training. On each trial of the MTS task a sample stimulus was presented in the center of the screen, along with two comparison stimuli on the bottom left and right of the screen (see Figure 1). Participants did not receive instructions about the correct sample-comparison mappings and had to learn these on the basis of error feedback. They were instructed to match a sample stimulus to one of the two comparison stimuli by pressing either a left- or a right-key (either the "A" or "P" keys on an AZERTY keyboard). The left-right position of the comparison stimuli varied randomly on a trial-by-trial basis and each alignment occurred equally often. Selecting the correct comparison stimulus led to a 1000ms intertrial interval, followed by the next trial, whereas selecting the incorrect comparison stimulus caused the screen to turn red for 200 ms prior to the intertrial interval. On each trial, the sample and comparison stimuli were presented for a maximum of 5,000 ms or until a response was emitted. Stimuli were the words ROOD and GROEN, which mean "red" and "green" in Dutch (for reasons of clarity we will refer to RED and GREEN in the remainder of the article) and the nonwords SMELK, GILPT, KLAMF, and PLESK. Stimuli were presented in uppercase, black, 16-point Arial font against a white background.

Conditional-discrimination training consisted of three phases. During the first phase, two nonwords (e.g., PLESK and KLAMF) served as sample stimuli while the color words RED and GREEN were used as comparison stimuli. Two color word to nonword relations were established by reinforcing the selection of a specific color word in the presence of a certain nonword. For instance, the participant was reinforced for selecting PLESK in the presence of RED and KLAMF in the presence of GREEN. During the second phase, the sample stimuli from the first phase were related to a second set of comparison stimuli (e.g., SMELK and GILPT) via reinforcement. For instance, the participant was now reinforced for selecting SMELK in the presence of PLESK and GILPT in the presence of KLAMF. The order of both training phases were counterbalanced across participants and followed by a third phase consisting of a mixed block in which each of the aforementioned relations was retrained. The first two phases each consisted of 80 trials (40 per sample stimulus), while the third phase consisted of 160 trials (40 per sample stimulus). In the third practice phase, a small break was introduced after 80 trials. The duration of the break was self-paced. In each practice phase the presentation of sample and comparison stimuli occurred in random sequences, such that the different samples (and the corresponding comparisons) were intermixed.

Test phase. The test phase consisted of a conditional-discrimination test and a Stroop task, the order of which was counterbalanced across participants.

Conditional-discrimination test. A notable change in comparison to the *training* phase was that participants did not receive any corrective feedback after responding. This test consisted of three different types of trials that were each presented 40 times (120 trials in total). The first trial-type was designed to assess if participants formed the relations between stimuli that were directly reinforced during the training phase (e.g., select RED given PLESK or GREEN given KLAMF). The second trial type was designed to assess whether the reversed (derived) stimulus relation was also formed (i.e., symmetry). To this end, sample and comparison stimuli were reversed compared to the training phase. For instance, when the selection of RED was reinforced in the presence of PLESK during training, we assessed whether PLESK would be selected in presence of RED as a sample stimulus. The third trial-type was designed to assess whether transitive relations were established. To this end, sample and comparison stimuli of both set of contingencies were recombined. For instance, RED (or GREEN) were presented along with SMELK and GILPT and we examined if participants were more likely to select RED given SMELK or GREEN given GILPT. Participants were provided with a small break after 60 trials.

Stroop task. Participants had to judge the color (red or green) of distractor words (RED, GREEN, KLAMF, GILPT, SMELK, PLESK) by pressing either a left- or a right-key (i.e., “A” or “P”) on an AZERTY keyboard. The color-to-response mapping was counterbalanced across participants. Each trial began with a fixation mark (“+”) presented in the middle of the screen for 500 ms, followed by a distractor word in one of the target colors for 5,000 ms or until a response was made. Following incorrect or late responses, the screen flashed red for 200 ms, before the 1,000-ms intertribal interval (ITI) started.

Three different types of distractor words were used in the Stroop task. The first were the genuine color words RED and GREEN. The second were reinforced associates (e.g., PLESK and KLAMF in the aforementioned example). The third were derived associates (e.g., SMELK and GILPT). Each stimulus was presented 30 times in each color in uppercase, 16-point Arial font against a white background. This resulted in 360 trials wherein color words, reinforced associates, and derived associates were each presented 120 times. For each type of stimuli, half of the trials were congruent with respect to the stimulus color and the other half were incongruent (i.e., 60 congruent trials and 60 incongruent trials). Congruent and incongruent trials were presented in a random order. A small break was provided after 60 trials.

Procedure. Participants were tested in groups of two or three. Each participant was placed in a separate cubicle. The experiment was run by using the T-scope C/C++ library for Windows (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006). Participants signed an informed consent at the onset of the experiment. Specific instructions were provided prior to each training and test phase. Prior to the first conditional-discrimination phase, participants were instructed about the trial structure of the conditional-discrimination task and how these should be responded to, namely, learn to relate the sample stimulus to the correct comparison stimulus on the basis of error feedback. An example was provided, which employed stimuli different to the stimuli used in the experiment. At the onset of the second phase, participants were informed that a new set of sample and comparison stimuli would be trained in exactly the same way. Prior to the third practice phase, participants were instructed that the sets of both previous phases would be intermixed. Prior to the conditional-discrimination test participants were instructed that no error-feedback was provided and that participants could respond to the sample stimuli as they believed was correct. The Stroop task was introduced by providing standard instructions, which indicated that the color was the relevant color and the distractors word were to be ignored. Participants were thus not instructed about the nature of the upcoming phases beforehand (i.e., at the onset of the experiment or during a specific phase). The experiment lasted for approximately 45 min.

Data analysis. All data processing and analyses were performed using R (R Core Team, 2018). For each test, data were analyzed by using repeated measured ANOVAs, with an α-level of .05. We also report Bayes factors and effect sizes (partial eta squared) for all relevant effects and interactions. The ANOVAs were calculated by using the “afex” package (Singmann, Bolker, Westfall, & Aust, 2018), Bayes factors were calculated with the “BayesFactor” package, using the default JZS prior (.707; Morey, Rouder, & Jamil, 2015). For all Bayesian analyses we report only the BF_1 (i.e., the Bayes factor for evidence in favor of the alternative hypothesis), and we interpret them using the classification discussed in Schönbrodt and Wagenmakers (2018). Note that rerunning the Bayesian ANOVAs, may lead to slightly diverging results. This is due to variations in sampling noise. Finally, we performed paired-samples *t* tests to decompose main effects and interactions; Bayes factors, effect sizes (Hedges’ *g*), and 95% confidence intervals are also reported for these analyses.

Results

Conditional-discrimination test. The main dependent variable was the proportion of trials on which a comparison stimulus was selected in line with directly reinforced, symmetry, and transitive relations, which we refer to as the proportion of correct trials or PCs. Five participants had an average PC-score below .70 (.47, .57, .55, .42, .68) and were not considered for further analysis. PCs were then submitted to a repeated measures ANOVA with the factor relation type (reinforced, symmetry, transitivity) as a within-subjects factor. The main effect of relation type was significant, $F(2, 102) = 4.39$, $MSE = .002$, $p < .05$, $\eta_p^2 = .08$, $BF = 2.57 \pm 1.40\%$. Contrasts indicated that PCs were lower for the transitivity relation ($M = .90$; $SD = .07$) compared to the directly reinforced relation ($M = .93$; $SD = .06$), $M_{diff} = .03$, $t(51) = 2.41$, $p < .05$, $BF = 2.09$, $g_{av} = .37$, 95% CI [.00, .05], and the symmetry relation ($M = .93$; $SD = .06$), $M_{diff} = .02$, $t(51) = 2.32$, $p < .05$, $BF = 1.74$, $g_{av} = .35$, 95% CI [.00, .05]. PCs did not differ significantly between the directly reinforced relation and the symmetry relation, $M_{diff} = .002$, $t(51) = .25$, $p = .80$, $BF = .16$, $g_{av} = .03$, 95% CI [−.01, .02].

Stroop task. Reaction times (RTs) of the correct trials and PCs of the Stroop task were each subjected to a 2 (Trial Type: Congruent, Incongruent) × 3 (Distractor Type: Color Words, Reinforced Associate, Derived Associate) repeated measures ANOVA. Cell means and corresponding standard errors are presented in Figure 2. For the RTs, analyses revealed a main effect of trial type, $F(1, 51) = 6.43$, $MSE = 628$, $p < .05$, $\eta_p^2 = .11$, $BF = 3.36 \pm 2.27\%$. RTs were shorter on congruent ($M = 406.81$; $SD = 56.09$) compared with incongruent trials ($M = 413.92$; $SD = 63.16$). Neither the main effect of distractor type, $F(2, 102) = 1.77$, $MSE = 606$, $p = .29$, $\eta_p^2 = .02$, $BF = 0.11 \pm 3.08\%$, nor the two-way interaction between trial type and distractor type were significant, $F(2, 102) = 1.77$, $MSE = 524$, $p = .18$, $\eta_p^2 = .03$, $BF = .27 \pm 8.84\%$. Visual inspection of the interaction, however, indicated that the difference between congruent and incongruent trials was 14 ms for the color words, 6 ms for the reinforced associates, and 2 ms for the derived associates. In view of our research question we conducted additional contrasts to further explore this interaction. These contrasts suggested that the congruency effect was significant for the color words, $M_{diff} = 13.55$, $t(51) = 2.41$, $p < .05$, $BF = 2.09$, $g_{av} = .19$, 95% CI [2.27, 24.83], but not for the reinforced associates, $M_{diff} = 6.36$, $t(51) = 1.41$, $p = .17$, $BF = .38$, $g_{av} = .11$, 95% CI [−2.72, 15.43], or the derived associates, $M_{diff} = 1.68$, $t(51) = .48$, $p = .64$, $BF = .17$, $g_{av} = .03$, 95% CI [−5.41, 8.78].

For the PCs, neither the main effect of trial type, $F(1, 51) = 1.64$, $MSE = .001$, $p = .21$, $\eta_p^2 = .03$, $BF = 0.28 \pm 1.52\%$, nor the main effect of distractor type, $F(2, 102) = .83$, $MSE = .001$, $p = .44$, $\eta_p^2 = .02$, $BF = 0.07 \pm 1.52\%$, were significant. The two-way interaction was also not significant, $F(2, 102) = 1.89$, $MSE = .001$, $p = .16$, $\eta_p^2 = .04$, $BF = 0.29 \pm 2.24\%$. Exploratory analyses produced a marginally significant congruency effect for the color words: $M_{diff} = .01$, $t(51) = 1.89$, $p = .06$, $BF = .79$, $g_{av} = .26$, 95% CI [.00, .02]. This was not the case for the reinforced associates, $M_{diff} = .00$, $t(51) = .49$, $p = .62$, $BF = .17$, $g_{av} = .08$, 95% CI [−.01, .01], and the derived associates, $M_{diff} = .00$, $t(51) = .93$, $p = .36$, $BF = .23$, $g_{av} = .10$, 95% CI [.00, .01].

Additional analysis. Stroop effects are known to become larger for slower responses (e.g., Bub, Masson, & Lalonde, 2006; Christ, Holt, White, & Green, 2007; Pratte, Rouder, Morey, &

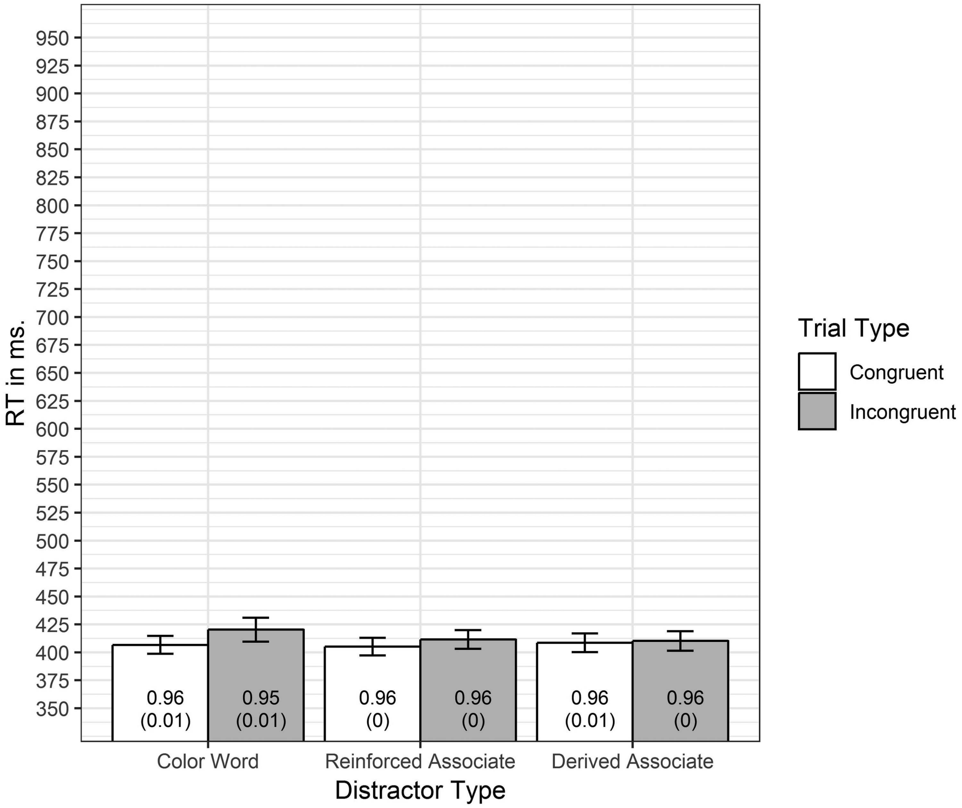

Figure 2. Reaction times (RTs) and PCs of Experiment 1 as a function of trial type and distractor type. Error bars denote the standard errors. Standard errors for PCs are printed between brackets.

Feng, 2010; West, 2003). Accordingly, we tested whether Stroop effects were more likely to be present when considering the right tail of the RT distribution. To this end, the "Vincent averaging" or "Vincentization" technique was used (Ratcliff, 1979; Vincent, 1912). For each participant and each cell of the design, correct RTs were sorted in ascending order and binned in five classes of equal size, using quantiles as boundary values. The mean of each bin was then computed (see also Ambrosi, Servant, Blaye, & Burle, 2019 for a similar approach). For each distractor and trial type, the mean of each bin was averaged across participants to obtain an average RT distribution which is representative of each individual distribution. Delta plots (e.g., Pratte et al., 2010) were then constructed calculating the difference in RT between congruent and incongruent trials within each bin, for each distractor type separately. These differences were then plotted as a function of the mean RT of the congruent and incongruent trials within that bin, per distractor type. The resulting delta plots are presented in Figure 3. For all three distractor types the difference between congruent and incongruent trials increases as a function of mean RT length. When considering the bin containing the 20% slowest RTs, the Stroop effect measured for the color words was significant, $M_{diff} = 50.77$, $t(51) = 2.96$, $p < .01$, $BF = 7.19$, $g_{av} = .31$, 95% CI [16.34, 85.19]. For the reinforced and derived associates the Stroop effect was numerically present, but statistically unreliable (reinforced associates: $M_{diff} = 24.40$, $t(51) = 1.53$, $p = .13$, $BF = .45$, $g_{av} = .17$, 95% CI [−7.62, 56.42]; derived associates: $M_{diff} = 14.47$, $t(51) = 1.08$, $p = .28$, $BF = .26$, $g_{av} = .10$, 95% CI [−12.33, 41.27]).

Discussion

The results of Experiment 1 are twofold. First, participants performed extremely well on the conditional-discrimination test, even though performance was slightly worse on the trials probing for transitivity relations. Based on the conditional-discrimination test, it is thus reasonable to assume that two equivalence classes were formed that each consisted of a color word, a reinforced associate, and a derived associate. Nevertheless, RTs in the Stroop task offer only moderate evidence for the hypothesis that congruent stimuli were responded to more quickly than incongruent stimuli. Additional contrasts indicated the presence of a significant Stroop effect for the color words, but not for the derived and reinforced associates. However, the interaction between trial type and distractor type was not significant. A similar pattern of results was observed for the PCs of the Stroop task, albeit without statistically significant differences. In an additional analysis, we inspected the RT distributions more closely and concentrated on the bin containing the 20% slowest responses. A significant Stroop effect was observed for the color words. Although Stroop effects

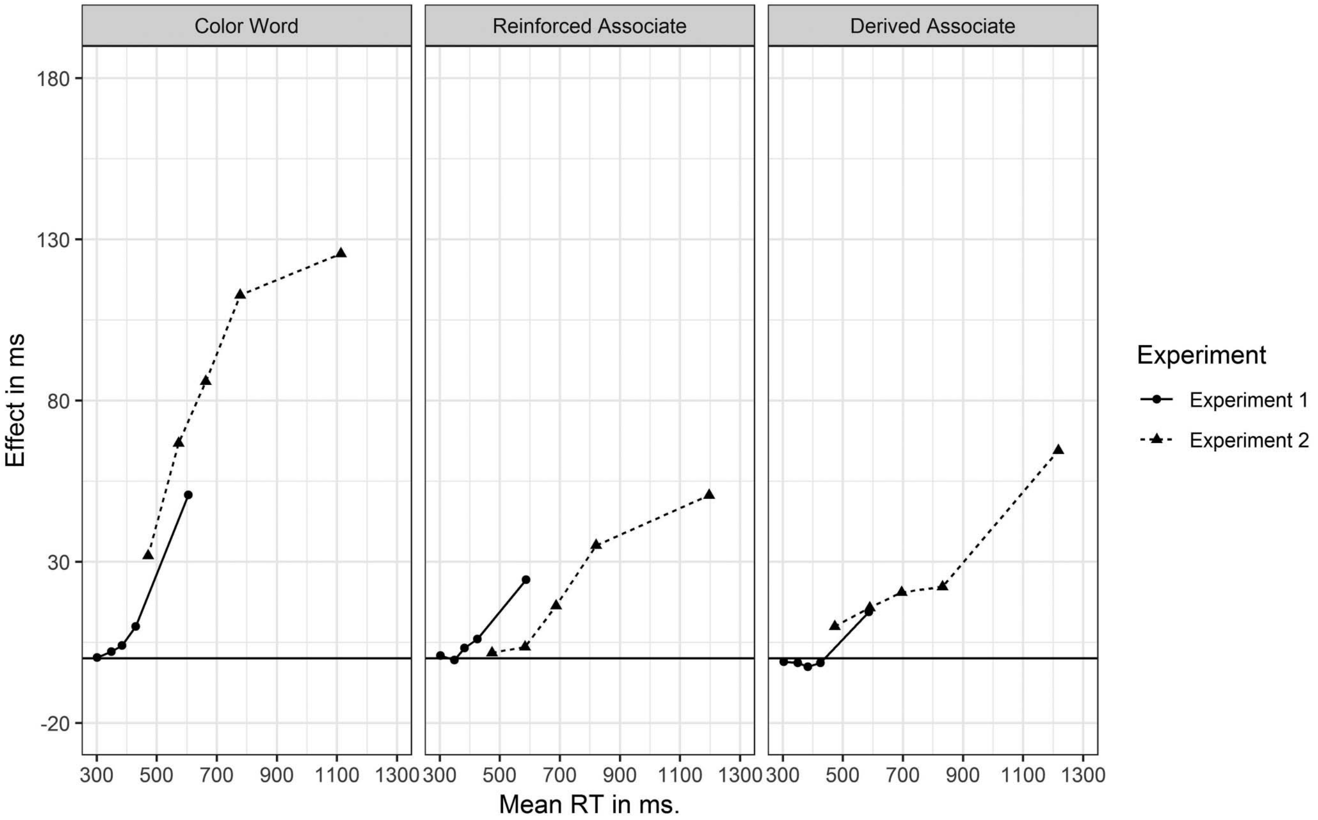

Figure 3. Difference in reaction time (RT) between congruent and incongruent trials of Experiment 1 and Experiment 2, calculated per bin and per distractor type, and plotted as a function of the mean bin RT per distractor type.

were numerically present for the reinforced and derived associates, these effects were not statistically different from zero.

Taken together, the results of the conditional-discrimination test indicate that new relations were derived after the training phases. However, the results of the Stroop task are difficult to interpret and the question arises whether the Stroop task used in Experiment 1 was optimal for the present purposes. This issue is tackled in Experiment 2.

Experiment 2

One issue with the Stroop task used in Experiment 1 was that participants might not have allocated sufficient attention to the distractor words when responding to the target colors. As was argued by Besner, Stolz, and Boutilier (1997; see also Bauer & Besner, 1997; Besner, 2001; Besner & Stolz, 1999; Manwell, Roberts, & Besner, 2004), processing of distractor words in the Stroop task is limited in various ways. Of particular interest for the present purposes is that the magnitude of the Stroop effect is drastically reduced when congruent stimuli (RED_{red}) are omitted and replaced by neutral stimuli ($\text{TABLE}_{\text{red}}$). Besner et al. (1997) argued that congruent stimuli encourage participants to read the distractor words. Although this reading strategy may be unconscious, it can bias the Stroop task. As a consequence, ignoring the irrelevant word on incongruent trials also becomes harder, which is evidenced by the observation that the size of the Stroop effect increases as the proportion of congruent trials increases (see, e.g., Tzelgov, Porat, & Henik, 1997). Of the 360 trials presented in the Stroop task of Experiment 1, only 60 were congruent trials, which included a color word as distractor. The remaining 120 congruent trials included a nonword. Within such a task context, participants may thus not have been sufficiently encouraged to process the distractor words. Accordingly, in Experiment 2 we adapted the Stroop task, such that participants were forced to attend the distractor words. To this end, the Stroop task was combined with a go/no-go task.[1] In a subset of the trials, participants were instructed not to respond. No-go trials were cued by the identity of two additional nonwords (filler words). Participants were thus required to process the identity of the distractor words on each trial of the Stroop task. In line with Experiment 1, the central question was whether reinforced and derived associates induce a Stroop effect in this adapted version of the Stroop task.

Method

Participants. Fifty-four new participants were recruited at Ghent University and paid 10 euro for their participation.

Procedure. The training phases and the conditional-discrimination test were similar to Experiment 1. The Stroop task was adapted in two ways. First, two additional nonwords (JULPS and FLIHK) were included. On a subset of the trials, the identity of these two distractor words cued that participants did not have to respond to the print color of the stimulus. These filler words were only used during the Stroop task and were not part of the training phase. Second, we suspected that the colors red and green may be intrinsically related to no-go and go action tendencies. Therefore we changed the target print-colors to blue and yellow. Pilot testing indicated that these latter colors were best perceived on a black background and, as such, we changed the background color of the Stroop task from white to black. Following these changes, the color words BLAUW and GEEL were now used, which mean BLUE and YELLOW, respectively, in Dutch. For reasons of clarity, we use the English translations in the remainder of the article. The Stroop task now consisted of 480 trials (120 trials per distractor type). A break was provided after each 60-trial block.

Results

Conditional-discrimination test. Two participants failed to meet the inclusion criterion of an accuracy proportion of .70 (.52, .58) and were excluded from subsequent analyses. The main effect of relation type was significant, $F(2, 102) = 5.59$, $MSE = .002$, $p < .01$, $\eta_p^2 = .10$, $BF = 6.75 \pm 1.49\%$. PCs were significantly higher for the directly reinforced relation ($M = .94$; $SD = .06$) compared with the symmetry relation ($M = .92$; $SD = .07$), $M_{diff} = .02$, $t(51) = 2.73$, $p < .01$, $BF = 4.18$, $g_{av} = .35$, 95% CI [.01, .04], or the transitivity relation ($M = .90$; $SD = .11$), $M_{diff} = .04$, $t(51) = 2.88$, $p < .01$, $BF = 5.99$, $g_{av} = .48$, 95% CI [.01, .04]. PCs did not differ significantly between the transitivity relation and the symmetry relation, $M_{diff} = .02$, $t(51) = 1.47$, $p = .15$, $BF = .41$, $g_{av} = .22$, 95% CI [−.01, .05].

Stroop task. Cell means and corresponding standard errors are presented in Figure 4. For the RTs, the main effect of trial type was significant, $F(1, 51) = 64.79$, $MSE = 2384$, $p < .001$, $\eta_p^2 = .56$, $BF > 1{,}000 \pm 3.48\%$. RTs were shorter on congruent trials compared to incongruent trials. The main effect of distractor type was also significant, $F(2, 102) = 8.74$, $MSE = 5{,}489$, $p < .001$, $\eta_p^2 = .15$, $BF > 1{,}000 \pm 2.02\%$. RTs were significantly shorter for color words compared with reinforced associates, $M_{diff} = 33.48$, $t(51) = 3.09$, $p < .01$, $BF = 9.79$, $g_{av} = .18$, 95% CI [11.69, 55.26], and derived associates, $M_{diff} = 41.73$, $t(51) = 3.78$, $p < .01$, $BF = 62.14$, $g_{av} = .22$, 95% CI [19.55, 63.89]. RTs did not differ significantly between reinforced and derived associates, $M_{diff} = 8.25$, $t(51) = .96$, $p = .34$, $BF = .23$, $g_{av} = .04$, 95% CI [−8.97, 25.47].

A two-way interaction between trial type and distractor type was also obtained, $F(2, 102) = 11.54$, $MSE = 2746$, $p < .001$, $\eta_p^2 = .18$, $BF = 84.46 \pm 4.63\%$. A significant congruency effect was observed for all three distractor types, color words: $M_{diff} = 84.70$, $t(51) = 8.41$, $p < .01$, $BF > 1{,}000$, $g_{av} = .48$, 95% CI [64.48, 104.91]; reinforced associates: $M_{diff} = 21.61$, $t(51) = 2.06$, $p < .05$, $BF = 1.06$, $g_{av} = .11$, 95% CI [.55, 42.66]; derived associates: $M_{diff} = 27.23$, $t(51) = 2.85$, $p < .01$, $BF = 5.47$, $g_{av} = .13$, 95% CI [8.02, 46.43]. The congruency effect was larger for the color words compared with the reinforced associates, $M_{diff} = 63.09$, $t(51) = 4.99$, $p < .01$, $BF > 1{,}000$, $g_{av} = .85$, 95% CI [37.69, 88.49], and derived associates, $M_{diff} = 57.47$, $t(51) = 3.78$, $p < .01$, $BF = 63.19$, $g_{av} = .81$, 95% CI [26.98, 87.97]. The congruency effect did not differ between the reinforced and derived associates, $M_{diff} = 5.62$, $t(51) = .36$, $p < .72$, $BF = .16$, $g_{av} = .08$, 95% CI [−25.67, 36.91].

For the PCs, a main effect was observed for trial type, $F(1, 51) = 36.28$, $MSE = .001$, $p < .001$, $\eta_p^2 = .42$, $BF > 1{,}000 \pm 1.49\%$. PCs were higher on congruent compared to incongruent trials. The main effect of distractor type was also significant, $F(2, 102) = 3.59$, $MSE = .001$, $p < .05$, $\eta_p^2 = .07$, $BF = 0.74 \pm 1.68\%$.

---

[1] The authors are indebted to Adriaan Spruyt for this suggestion.

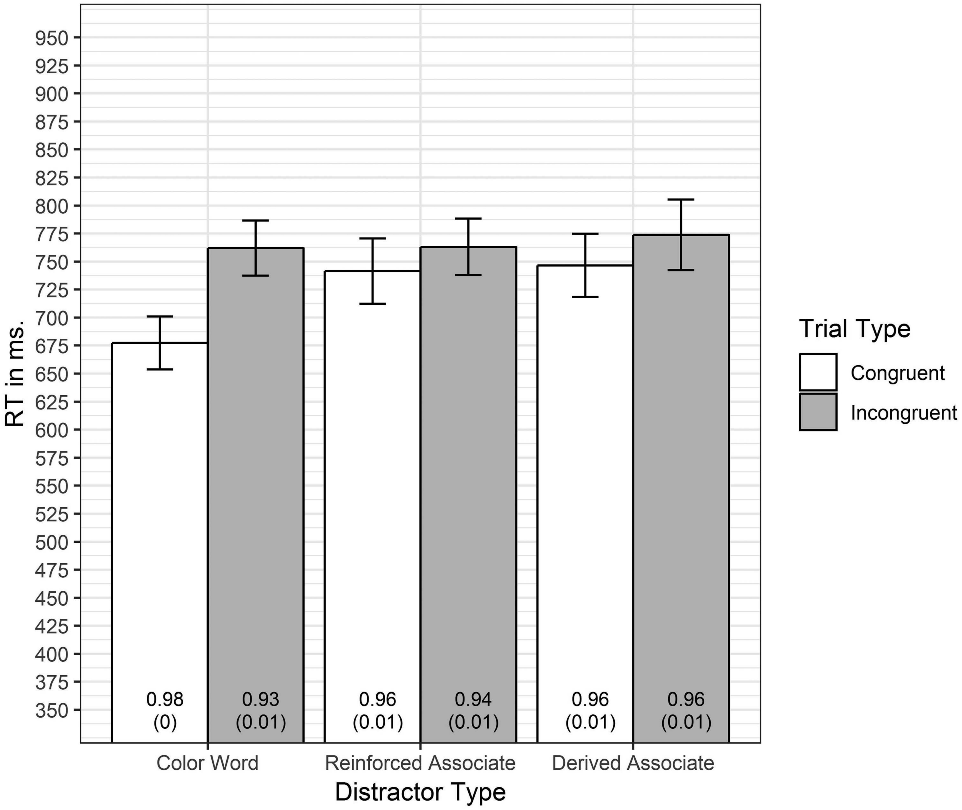

Figure 4. Reaction times (RTs) and PCs of Experiment 2 as a function of trial type and distractor type. Error bars denote the standard errors. Standard errors for PCs are printed between brackets.

PCs were significantly higher on the derived associates compared with the reinforced associates, $M_{diff} = .01$, $t(51) = 2.82$, $p < .01$, $BF = 5.18$, $g_{av} = .29$, 95% CI [.00, .02], and color words, $M_{diff} = .01$, $t(51) = 2.06$, $p < .05$, $BF = 1.06$, $g_{av} = .24$, 95% CI [.00, .02]. PCs did not differ significantly between the color words and reinforced associates, $M_{diff} = .00$, $t(51) = .67$, $p = .51$, $BF = .19$, $g_{av} = .09$, 95% CI [−.01, .01].

The interaction between trial type and distractor type was also significant, $F(2, 102) = 9.68$, $MSE = .001$, $p < .001$, $\eta_p^2 = .16$, $BF = 168.03 \pm 2.25\%$. The congruency effect was significant for the color words, $M_{diff} = .04$, $t(51) = 5.84$, $p < .01$, $BF > 1{,}000$, $g_{av} = 1.04$, 95% CI [.03, .06], and the reinforced associates, $M_{diff} = .02$, $t(51) = 3.44$, $p < .01$, $BF = 24.56$, $g_{av} = .44$, 95% CI [.01, .04], but not for the derived associates, $M_{diff} = .00$, $t(51) = .87$, $p = .39$, $BF = .22$, $g_{av} = .11$, 95% CI [−.01, .02]. The congruency effect for the color words was also larger than that observed for the reinforced associates, $M_{diff} = .02$, $t(51) = 2.26$, $p < .05$, $BF = 1.55$, $g_{av} = .41$, 95% CI [.00, .04], and the derived associates, $M_{diff} = .04$, $t(51) = 4.23$, $p < .01$, $BF = 231.33$, $g_{av} = .86$, 95% CI [.02, .06].

Discussion

Results indicated that performance during the conditional-discrimination test was significantly better for directly reinforced compared with symmetry and transitivity relations. This is slightly different compared to Experiment 1, in which we observed that performance was somewhat lower for the transitivity relation compared to the other two relation types that were probed. Nevertheless, in general, performance on the conditional-discrimination test was good. Hence, it is safe to conclude that two equivalence classes were again formed after the conditional-discrimination training.

In contrast to Experiment 1, a large Stroop-effect was now present for the color words, thus replicating previous findings in the literature (see MacLeod, 1991). More importantly, we also observed Stroop effects for the reinforced and derived associates. For the reinforced associates, the Stroop effect was significant for the RTs and the PCs. For the derived associates, the Stroop effect was significant for the RTs, but not for the PCs. Whereas the Stroop effect was larger for the color words compared to the reinforced and derived associates, it did not differ between reinforced and derived associates.

It could be argued that the presence of Stroop effects in Experiment 2 is not driven by the additional demand to process the distractor words, but simply by the fact that response times were longer in Experiment 2 than in Experiment 1. Because Stroop effects increase for longer RTs, any secondary task that slows down responding may have induced the same pattern of results. In order to investigate this issue, we also constructed delta plots for Experiment 2. These are presented along with the delta plots of Experiment 1 in Figure 3. For the color words, the Stroop effect in Experiment 2 is larger, even for bins which fall within the RT range covered by the delta plot of Experiment 1. For the reinforced associates, the Stroop effect is numerically larger for the slower bins in Experiment 1 compared to the early bins of Experiment 2, which again cover the same range of RTs. For the derived associates, however, the delta plots of both experiments are almost completely aligned. This rather mixed pattern of results thus suggests that the Stroop effects obtained in Experiment 2, may in part be induced by the need to focus attention on the distractor words, but also by shifts in RT distributions. Note, however, that regardless of the reasons for the difference in results of Experiments 1 and 2, it is important to see that Stroop effects can be found not only for reinforced but also for derived associates.

Taken together, the results of Experiment 2 suggest that new stimulus–stimulus relations that are formed through derivation can induce a response bias upon the condition that participants are encouraged to attend to the distractor words. In addition, derived stimulus-stimulus relations bias responding to the same extent as directly reinforced stimulus-stimulus relations. The aim of Experiment 3 was to replicate and further examine the nature of these effects.

Experiment 3

Past work has frequently pointed to the idea that the Stroop effect has two sources. On the one hand, conflict can occur between the meaning of the word and the color, for instance, the word "GREEN" and the color "yellow" (i.e., *stimulus conflict,* Glaser & Glaser, 1982; Mackinnon, Geiselman, & Woodward, 1985; Stirling, 1979). On the other hand, the response engendered by the word (i.e., pronouncing the word "green") and the response engendered by the color (e.g., pronouncing the word "yellow" or pressing an arbitrary response linked to the color "yellow") also compete for selection (i.e., *response conflict*; Klein, 1964). The general consensus is that both stimulus and response conflict contribute to the standard Stroop effect (Augustinova & Ferrand, 2014; Augustinova, Silvert, Ferrand, Llorca, & Flaudias, 2015; Ferrand & Augustinova, 2014). The separate contribution of stimulus conflict and response conflict to the Stroop effect was evidenced by two-to-one mapping experiments. For instance, De Houwer (2003; see also, Chen, Bailey, Tiernan, & West, 2011; Hasshim & Parris, 2015; Jongen & Jonkman, 2008; Schmidt & Cheesman, 2005; van Veen & Carter, 2005) presented participants with a Stroop task in which four target colors were present: half of the target colors were mapped to one response key (e.g., green and yellow on a left key) while the other half were mapped to a second key (e.g., pink and blue on a right key). In this task, there were three different types of trials: (a) *identity* trials in which target color and color word converged (e.g., $\text{GREEN}_{\text{green}}$); (b) *same-response* trials in which target color and color word differed, but both pointed toward the same response (e.g., $\text{PINK}_{\text{blue}}$); and (c) *different-response* trials in which the target color and color word both required a different response (e.g., $\text{GREEN}_{\text{pink}}$). The difference in performance between identity and same-response trials reflects a *stimulus-congruency effect*, which is assumed to offer a proxy of stimulus conflict, whereas the difference in performance between same- and different-response trials reflects a *response-congruency effect*, which is taken as proxy of response conflict. De Houwer (2003) reported evidence for both stimulus and response conflicts, with same-response trials being slower than identity trials (stimulus-congruency effect), but faster than different-response trials (response-congruency effect).

In Experiment 3, we applied the two-to-one mapping approach in order to investigate the nature of the conflict induced by reinforced and derived associates, and by doing so, the characteristics of the representations mediating the equivalence classes we established through conditional-discrimination training. For the color words, we expected to replicate previous findings, which indicate the presence of both stimulus- and response-congruency effects (i.e., RT identity trials $<$ RT same-response trials $<$ RT different-response trials; e.g., De Houwer, 2003). Yet the more critical question was: which properties of the color words are acquired by reinforced and derived associates following conditional-discrimination training? This issue can be related to the distinction between lexical and semantic word representations. Lexical representations contain information about word forms, whereas semantic representations correspond to word meanings. In the context of language acquisition, Kroll and Stewart (1994) assume that second-language word forms (lexical representations) do not have direct access to semantic representations, as it is the case for first-language words. Instead, lexical representations of second-language words are first translated into the lexical representation of their first-language equivalents. Following Kroll and Stewart (1994), second-language words only directly access semantics when second-language proficiency is sufficiently high. When considering reinforced and derived color words associates as newly learned "foreign language" words, two possibilities thus arise. On the one hand, the conditional-discrimination training we provided might not have been extensive enough for the associates to be directly linked to the semantic representation of the corresponding color. For example, nonwords trained to be equivalent with the color word GREEN, would not activate a semantic representation "green." In this case, the Stroop effects observed in Experiment 2 for the reinforced and derived associates are based solely on response conflicts. More specifically, the reinforced and derived associates may be automatically mapped onto their corresponding lexical equivalent, namely the color word, after which the response assigned to the corresponding color in the Stroop task is retrieved. In this scenario, a response-congruency effect but no stimulus-congruency effect should be observed (i.e., RTs identity trials $=$ RTs same-response trials $<$ RTs different-response trials).

On the other hand, several studies indicate that newly acquired foreign words have almost immediate access to semantic representations even at early stages of novel language learning (see Brysbaert & Duyck, 2010, for a review). In addition, Schmidt, Hartsuiker, and De Houwer (2018) presented Dutch-speaking participants with French color words in a Stroop task with a two-to-one mapping and observed both a stimulus-congruency effect and a response-congruency effect. Based on these findings, it could be argued that conditional-discrimination training is sufficient for the associates to be directly related to the semantic representation of their corresponding color word. In line with Schmidt et al. (2018), both a stimulus- and a response-congruency effect are expected to emerge (i.e., RTs identity trials $<$ RTs same-response trials $<$ RTs different-response trials). Both hypotheses were contrasted in Experiment 3.

Method

Participants. Fifty-nine new participants were recruited from Ghent University and paid 10 euro for taking part.

Procedure. A similar procedure was used as in Experiment 2 with three changes. First, in the Stroop task participants now had to react to four target print-colors: blue, magenta, yellow, and gray. Blue and magenta were mapped to one response key, yellow and gray to another. The left-right assignment of the responses was counterbalanced across participants. The Stroop task consisted of 576 trials, thus 144 trials per distractor type (color words, reinforced associates, derived associates, filler words). For each distractor type, there were 48 identity trials, 48 same-response trials, and 48 different-response trials. The Stroop task was subdivided in six blocks of 96 trials, with a small break after each block. In order to keep the experiment within a 60 min timeframe, the training phases were shortened to 50 trials per contingency. As a result, the third training phase only contained 100 trials and no break was provided. The memory test was also reduced to 96 trials (i.e., 32 trials for testing each relation), without any break. Finally, the 5,000-ms response deadline was reduced to 2,500 ms.

Results

Conditional-discrimination test. Four participants failed to meet the inclusion criteria of .70 (.44, .47, .65, and .68) and were excluded from subsequent analyses. Overall accuracy on the memory test was .87. The main effect of relation type was not significant, $F(2, 108) = .18$, $MSE = .00$, $p = .83$, $\eta_p^2 = .00$, $BF = .07 \pm .95\%$. Performance was comparable for the directly reinforced ($M = .88$, $SD = .10$), symmetry ($M = .87$, $SD = .08$) and transitivity relations ($M = .87$, $SD = .10$).

Stroop task. Cell means and corresponding standard errors are presented in Figure 5. For the RTs, the main effect of trial type was significant, $F(2, 108) = 79.56$, $MSE = 4{,}296$, $p < .001$, $\eta_p^2 = .60$, $BF > 1{,}000 \pm 6.02\%$. The stimulus-congruency effect (i.e., the difference in performance between identity trials and same-response trials) was not significant, $M_{diff} = 10.11$, $t(54) = 1.70$, $p = .09$, $BF = .57$, $g_{av} = .08$, 95% CI [−1.79, 22.01]. The response-congruency effect (i.e., the difference in performance between same-response and different-response trials) was significant: $M_{diff} = 73.25$, $t(54) = 9.88$, $p < .001$, $BF > 1{,}000$, $g_{av} = .55$, 95% CI [58.38, 88.17].

The main effect of distractor type was also significant, $F(2, 108) = 60.25$, $MSE = 3456$, $p < .001$, $\eta_p^2 = .53$, $BF > 1{,}000 \pm 4.04\%$. RTs were significantly shorter for color words compared with reinforced associates, $M_{diff} = 62.71$, $t(54) = 10.00$, $p < .001$, $BF > 1{,}000$, $g_{av} = .48$, 95% CI [50.14, 75.28], and derived associates, $M_{diff} = 63.30$, $t(54) = 9.67$, $p < .001$, $BF > 1{,}000$, $g_{av} = .49$, 95% CI [50.18, 76.43]. RTs did not differ between reinforced and derived associates, $M_{diff} = .59$, $t(54) = .09$, $p = .93$, $BF = .15$, $g_{av} = .00$, 95% CI [−12.18, 13.37].

The interaction between trial type and distractor type was significant, $F(4, 216) = 17.43$, $MSE = 1{,}837$, $p < .001$, $\eta_p^2 = .25$, $BF > 1{,}000 \pm 2.59\%$. For the color words, the stimulus-

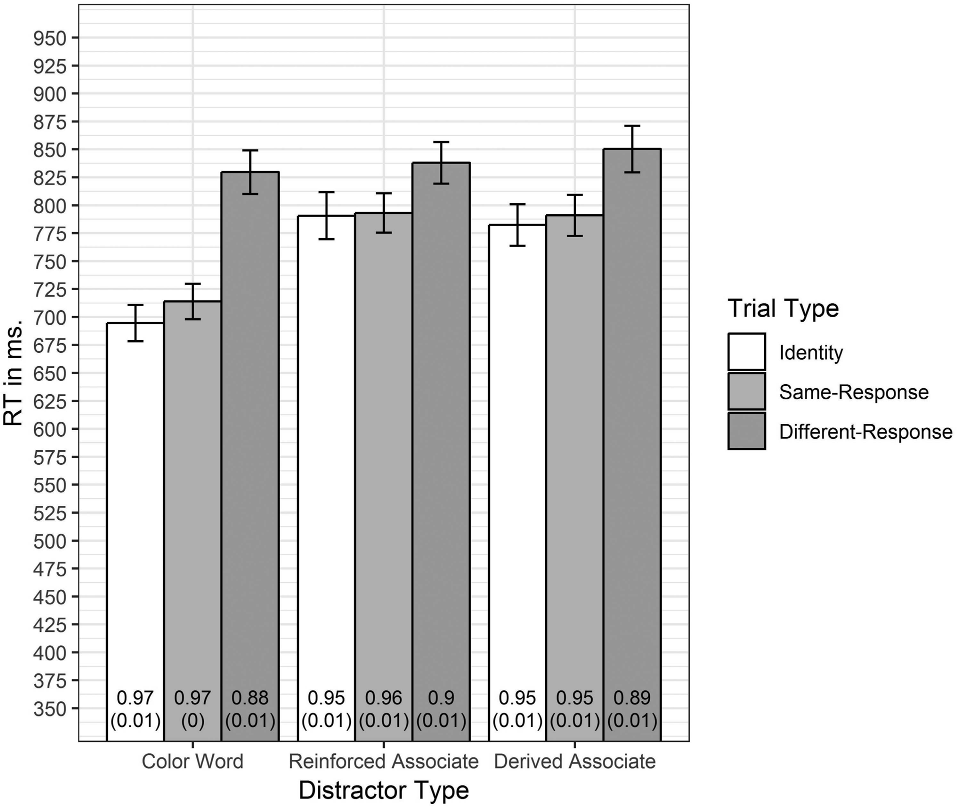

Figure 5. Reaction times (RTs) and PCs of Experiment 3 as a function of trial type and distractor type. Error bars denote the standard errors. Standard errors for PCs are printed between brackets.

congruency effect, $M_{diff} = 19.33$, $t(54) = 2.44$, $p < .05$, $BF = 2.17$, $g_{av} = .16$, 95% CI [3.42, 35.25], and the response-congruency effect were significant, $M_{diff} = 115.67$, $t(54) = 13.69$, $p < .001$, $BF > 1{,}000$, $g_{av} = .88$, 95% CI [98.74, 132.59]. For the reinforced associates, the stimulus-congruency effect was not significant, $M_{diff} = 2.49$, $t(54) = .24$, $p = .81$, $BF = .15$, $g_{av} = .02$, 95% CI [−18.58, 23.58], whereas the response-congruency effect was significant, $M_{diff} = 44.88$, $t(54) = 4.58$, $p < .01$, $BF = 731.19$, $g_{av} = .33$, 95% CI [25.24, 64.51]. A similar pattern was observed for the derived associates: A stimulus-congruency effect did not emerge, $M_{diff} = 8.58$, $t(54) = 1.05$, $p = .30$, $BF = .24$, $g_{av} = .06$, 95% CI [−7.71, 24.87]; whereas a response-congruency effect did emerge, $M_{diff} = 59.28$, $t(54) = 5.28$, $p < .001$, $BF > 1{,}000$, $g_{av} = .41$, 95% CI [36.77, 81.79].

Follow-up contrasts revealed that the stimulus-congruency effect observed for the color words did not differ significantly from the nonsignificant stimulus-congruency effects observed for the reinforced, $M_{diff} = 16.84$, $t(54) = 1.57$, $p = .12$, $BF = .47$, $g_{av} = .25$, 95% CI [−4.61, 38.28], and derived associates, $M_{diff} = 10.76$, $t(54) = 1.03$, $p = .31$, $BF = .24$, $g_{av} = .18$, 95% CI [−10.17, 31.68]. Furthermore, the response-congruency effect was larger for color words compared to reinforced associates, $M_{diff} = 70.79$, $t(54) = 6.34$, $p < .001$, $BF > 1{,}000$, $g_{av} = 1.04$, 95% CI [48.43, 93.15], or derived associates, $M_{diff} = 56.39$, $t(54) = 5.19$, $p < .001$, $BF > 1{,}000$, $g_{av} = .77$, 95% CI [34.64, 78.14]. The response-congruency effect did not differ between the latter two distractor types, $M_{diff} = 14.40$, $t(54) = 1.27$, $p = .21$, $BF = .32$, $g_{av} = .18$, 95% CI [−8.32, 37.13].

With respect to the PCs, the main effect of trial type was significant, $F(2, 108) = 44.38$, $MSE = .01$, $p < .001$, $\eta_p^2 = .45$, $BF > 1{,}000 \pm 2.08\%$. Whereas no evidence was obtained for a stimulus-congruency effect, $M_{diff} = .00$, $t(54) = .65$, $p = .52$, $BF = .18$, $g_{av} = .09$, 95% CI [−.01, .02], the response-congruency effect was again significant, $M_{diff} = .07$, $t(54) = 7.54$, $p < .001$, $BF > 1{,}000$, $g_{av} = 1.31$, 95% CI [.05, .09]. The main effect of distractor type was on the verge of being significant, $F(2, 108) = 3.17$, $MSE = .001$, $p = .05$, $\eta_p^2 = .06$, $BF = .10 \pm 2.09\%$. PCs were lower for the derived associates compared to the color words, $M_{diff} = .01$, $t(54) = 2.18$, $p < .05$, $BF = 1.31$, $g_{av} = .21$, 95% CI [.00, .02], and the reinforced associates, $M_{diff} = .01$, $t(54) = 2.05$, $p < .05$, $BF = 1.01$, $g_{av} = .15$, 95% CI [.00, .01]. PCs did not differ between color words and reinforced associates, $M_{diff} = .00$, $t(54) = .73$, $p = .47$, $BF = .19$, $g_{av} = .07$, 95% CI [−.01, .01].

Finally, the interaction between trial type and distractor type was also significant, $F(4, 216) = 6.43$, $MSE = .001$, $p < .001$, $\eta_p^2 = .11$, $BF = 4.89 \pm 2.18\%$. The response-congruency effect was significant for all three types of distractors; color words: $M_{diff} = .10$, $t(54) = 8.48$, $p < .001$, $BF > 1{,}000$, $g_{av} = 1.58$, 95% CI [.08, .12]; reinforced associates: $M_{diff} = .06$, $t(54) = 5.13$, $p < .01$, $BF > 1{,}000$, $g_{av} = .96$, 95% CI [.04, .08]; derived associates: $M_{diff} = .06$, $t(54) = 4.67$, $p < .01$, $BF = 967.76$, $g_{av} = .85$, 95% CI [.03, .08]. The response-congruency effect was larger for color words compared to reinforced associates, $M_{diff} = .04$, $t(54) = 3.49$, $p < .01$, $BF = 28.47$, $g_{av} = .49$, 95% CI [.02, .07], and derived associates, $M_{diff} = .04$, $t(54) = 2.85$, $p < .01$, $BF = 5.54$, $g_{av} = .48$, 95% CI [.01, .07]. The difference between reinforced and derived associates was not significant, $M_{diff} = .00$, $t(54) = .22$, $p = .83$, $BF = .15$, $g_{av} = .03$, 95% CI [−.02, .02]. Stimulus-congruency effects were absent for all three types of distractors: color words, $M_{diff} = .01$, $t(54) = .72$, $p = .47$, $BF = .19$, $g_{av} = .11$, 95% CI [−.01, .02]; reinforced associates, $M_{diff} = .01$, $t(54) = .85$, $p = .39$, $BF = .21$, $g_{av} = .13$, 95% CI [−.01, .03]; derived associates, $M_{diff} = .00$, $t(54) = .11$, $p = .91$, $BF = .15$, $g_{av} = .02$, 95% CI [−.01, .01].

Additional analyses. Delta plots were again constructed (see Experiment 1 for a description of how this was done) to have a more fine-grained view on the entire RT distribution and see whether the stimulus-congruency effects were not cloaked in the main analysis. As can be seen in Figure 6, the stimulus-congruency effect only slightly increased as a function of mean RT for the derived associates: For the Bin containing the 20% slowest responses, this effect was around 25 ms. However, the corresponding contrast was not significant: $M_{diff} = 24.75$, $t(54) = 1.24$, $p = .22$, $BF = .31$, $g_{av} = .10$, 95% CI [−15.18, 64.68].

Discussion

Performance on the conditional-discrimination test was again relatively good, which can be taken as evidence that participants formed two equivalence classes after the conditional-discrimination training. When considering performance on the Stroop task, results indicated the presence of a response-congruency effect: performance was inferior on different-response trials compared with same-response trials. This effect was more pronounced for color words compared with reinforced and derived associates. The pattern of results were similar for RTs and PCs. The response-congruency effect was similar for reinforced and derived associates.

With regard to the stimulus-congruency effect, the results are more ambiguous. Consider, for instance, the RTs. For the color words, the stimulus-congruency effect was significant but only anecdotal evidence was obtained for the alternative hypothesis. Nevertheless, this pattern is in line with previous studies indicating the presence of stimulus-congruency effects in the Stroop task when using color words as distractors (e.g., De Houwer, 2003; Schmidt & Cheesman, 2005; Schmidt et al., 2018). For the reinforced and derived associates, the stimulus-congruency effect was not significant and the corresponding Bayes factors indicated that the null hypothesis was very likely. The Stroop effect induced by reinforced and derived associates is thus mainly driven by response conflict.

Taken together, the results of Experiment 3 suggest that reinforced and derived associates are connected to lexical representations of the color words. During the Stroop task, the associates may be automatically translated into their corresponding color words, which in turn leads to the activation of the response associated to that color via the instructions of the Stroop task. However, the associates do not directly access a semantic representation of the color word they are related to.

Experiment 4

In Experiment 4, we tested if stimulus conflict is more likely to occur when drastically increasing the length of the conditional-discrimination training, which now included 800 trials. Moreover, the Stroop task now always preceded the conditional-discrimination test. The reason for doing so is that in this test, discrimination on the basis of transitivity was tested by presenting derived associates and color words on the same trial. Although no reinforcement was provided, the contingent presentation of both types

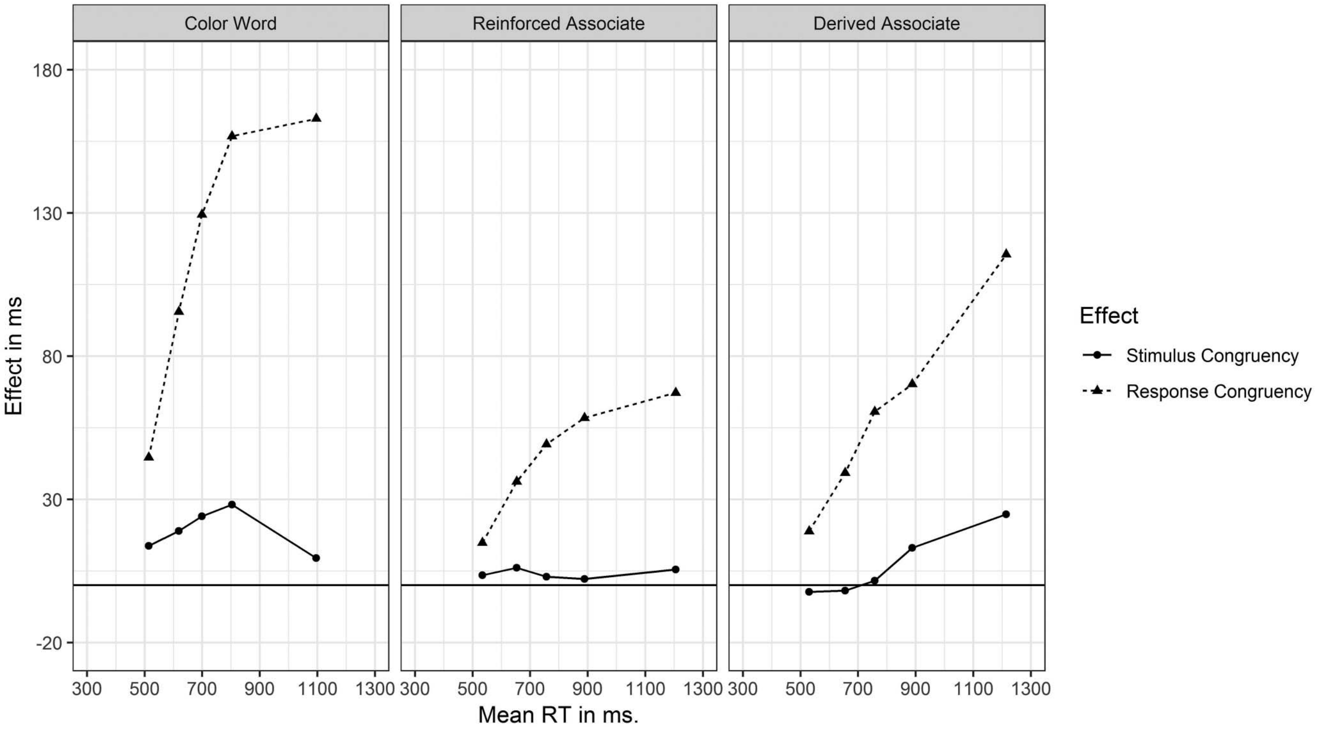

Figure 6. Difference in reaction time (RT) between congruent and incongruent trials of Experiment 3, calculated per bin and per distractor type, and plotted as a function of the mean bin RT per distractor type.

of stimuli may have inflated the effect of derived associates in the Stroop task. In Experiment 4, we excluded any possibility for such effects by systematically presenting the conditional-discrimination test after the Stroop task. Experiment 4 was concerned with two questions: (a) Do the response-congruency effects observed in Experiment 3 replicate?; and (b) Will a stimulus-congruency effect emerge for the reinforced and derived color associates following more extensive conditional-discrimination training?

Method

Participants. Forty-nine new participants were recruited and paid 20 euro for taking part.

Procedure. The procedure was similar to that used in Experiment 3 with three exceptions. First, the number of trials in the conditional-discrimination training was quadrupled so that participants now encountered a total of 800 trials (i.e., 200 trials per contingency). In all three training phases, small breaks were provided after each block of 100 trials. Second, the Stroop task was now presented first, followed by the conditional-discrimination test. Third, the conditional-discrimination test again included 120 trials, with a small break after 60 trials (see Experiments 1 and 2). Overall, the experiment took approximately 2 hr.

Results

Conditional-discrimination test. Three participants failed to meet the .70 accuracy inclusion criterion (.50, .50, and .67) and were excluded from subsequent analyses. The main effect of relation type was significant, $F(2, 90) = 8.56$, $MSE = .001$, $p < .001$, $\eta_p^2 = .14$, $BF = 67.36 \pm 0.56\%$. PCs were significantly higher for the directly reinforced ($M = .94$; $SD = .06$), compared with the symmetry ($M = .88$; $SD = .01$), $M_{diff} = .06$, $t(45) = 5.43$, $p < .01$, $BF > 1{,}000$, $g_{av} = .81$, 95% CI [.04, .08], and transitivity relations ($M = .89$; $SD = .13$), $M_{diff} = .05$, $t(45) = 3.01$, $p < .01$, $BF = 8.09$, $g_{av} = .54$, 95% CI [.02, .09]. PCs for symmetry and transitivity relations did not differ from one another, $M_{diff} = .01$, $t(45) = .37$, $p = .71$, $BF = .17$, $g_{av} = .06$, 95% CI [−.03, .04].

Stroop task. Cell means and corresponding standard errors are presented in Figure 5. For the RTs, the main effect of trial type was significant, $F(2, 90) = 72.50$, $MSE = 4496$, $p < .001$, $\eta_p^2 = .62$, $BF > 1{,}000 \pm 4.12\%$. Whereas the stimulus-congruency effect was not significant, $M_{diff} = 14.84$, $t(45) = 1.76$, $p = .09$, $BF = .66$, $g_{av} = .10$, 95% CI [−2.11, 31.80], the response-congruency effect was, $M_{diff} = 75.10$, $t(45) = 10.36$, $p < .001$, $BF > 1{,}000$, $g_{av} = .50$, 95% CI [60.49, 89.69]. The main effect of distractor type was also significant, $F(2, 90) = 13.48$, $MSE = 5511$, $p < .001$, $\eta_p^2 = .23$, $BF > 1{,}000 \pm 4.08\%$. RTs were significantly shorter on color words compared with reinforced associates, $M_{diff} = 34.93$, $t(45) = 4.32$, $p < .01$, $BF = 269.39$, $g_{av} = .24$, 95% CI [18.63, 51.23], and derived associates, $M_{diff} = 49.09$, $t(45) = 4.75$, $p < .01$, $BF = 974.86$, $g_{av} = .33$, 95% CI [28.26, 69.91]. RTs did not differ significantly between reinforced and derived associates, $M_{diff} = 14.15$, $t(45) = 1.92$, $p = .06$, $BF = .86$, $g_{av} = .10$, 95% CI [.72, 29.03].

The interaction between trial type and distractor type was significant, $F(4, 180) = 14.27$, $MSE = 2719$, $p < .001$, $\eta_p^2 = .24$, $BF > 1{,}000 \pm 5.97\%$ (see Figure 7). For the color words, both the stimulus-congruency, $M_{diff} = 34.49$, $t(45) = 3.13$, $p < .01$, $BF =$

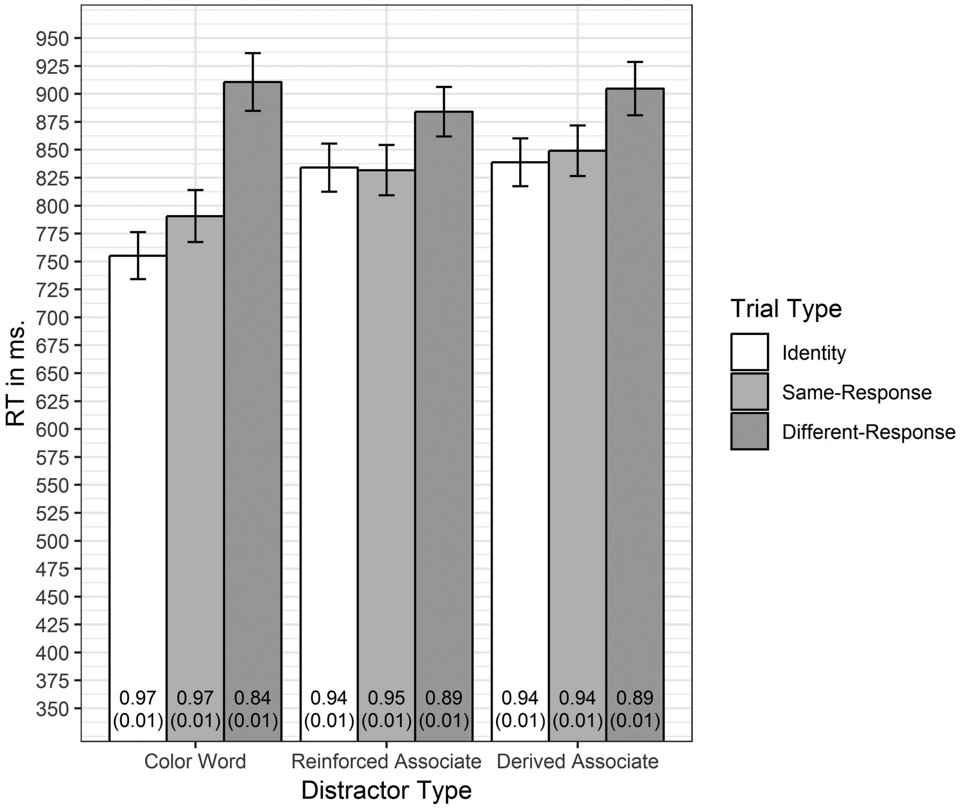

Figure 7. Reaction times (RTs) and PCs of Experiment 4 as a function of trial type and distractor type. Error bars denote the standard errors. Standard errors for PCs are printed between brackets.

10.96, g_{av} = .23, 95% CI [12.69, 58.30], and the response-congruency effect were significant, M_{diff} = 120.01, $t(45)$ = 7.92, $p < .01$, $BF > 1{,}000$, g_{av} = .72, 95% CI [89.49, 150.54]. For the reinforced associates, the response-congruency effect was significant, M_{diff} = 52.29, $t(45)$ = 5.24, $p < .01$, $BF > 1{,}000$, g_{av} = .34, 95% CI [32.19, 72.39], but not the stimulus-congruency effect, M_{diff} = 2.17, $t(45)$ = .19, p = .85, BF = .16, g_{av} = .02, 95% CI [−22.43, 26.98]. Similarly, for the derived associates, the response-congruency effect was significant, M_{diff} = 55.60, $t(45)$ = 5.55, $p < .01$, $BF > 1{,}000$, g_{av} = .35, 95% CI [35.42, 75.77], but not the stimulus-congruency effect, M_{diff} = 10.38, $t(45)$ = 1.13, p = .27, BF = .28, g_{av} = .07, 95% CI [−8.39, 29.15].

The response-congruency effect observed for the color words was significantly larger than that observed for the reinforced associates, M_{diff} = 67.72, $t(45)$ = 3.72, $p < .01$, BF = 50.24, g_{av} = .79, 95% CI [31.08, 104.36], and derived associates, M_{diff} = 64.42, $t(45)$ = 3.46, $p < .01$, BF = 24.83, g_{av} = .75, 95% CI [26.89, 101.94]. The response-congruency effect did not differ between reinforced and derived associates, M_{diff} = 3.30, $t(45)$ = 0.28, p = .78, BF = .17, g_{av} = .05, 95% CI [−20.43, 27.04]. Similarly, the stimulus-congruency effect was significantly larger for the color words compared with the reinforced associates, M_{diff} = 37.77, $t(45)$ = 2.73, $p < .01$, BF = 4.21, g_{av} = .47, 95% CI [9.87, 65.67], and the derived associates, M_{diff} = 25.12, $t(45)$ = 2.24, $p < .05$, BF = 1.53, g_{av} = .36, 95% CI [2.53, 47.72]. The stimulus-congruency effect did not differ between reinforced and derived associates, M_{diff} = 12.65, $t(45)$ = 1.09, p = .28, BF = .28, g_{av} = .17, 95% CI [−10.66, 35.95].

With respect to the PCs, the main effect of trial type was significant, $F(2, 90)$ = 41.55, MSE = .01, $p < .001$, η_p^2 = .48, $BF > 1{,}000 \pm 2.34\%$. A response-congruency effect was present, M_{diff} = .08, $t(45)$ = 8.03, $p < .001$, $BF > 1{,}000$, g_{av} = 1.29, 95% CI [−.10, −.06], but no stimulus-congruency effect, M_{diff} = .01, $t(45)$ = .97, p = .33, BF = .25, g_{av} = .13, 95% CI [−.01, .02]. Although the main effect of distractor type was not significant, $F(2, 90)$ = .67, MSE = .00, p = .51, η_p^2 = .01, $BF = 0.04 \pm 2.35\%$, the interaction between trial type and distractor type was significant, $F(4, 180)$ = 12.68, MSE = .00, $p < .001$, η_p^2 = .22, $BF > 1{,}000 \pm 2.47\%$.

A significant response-congruency effect was observed for all three types of distractor words: color words: M_{diff} = .13, $t(45)$ = 8.86, $p < .001$, $BF > 1{,}000$, g_{av} = 1.75, 95% CI [.09, .16]; reinforced associates: M_{diff} = .07, $t(45)$ = 5.36, $p < .001$, $BF >$ 1,000, g_{av} = .95, 95% CI [.04, .09]; and derived associates: M_{diff} = .05, $t(45)$ = 3.96, $p < .01$, BF = 96.38, g_{av} = .64, 95% CI [.03, .08]. Once again, the response-congruency effect was larger for color words compared with the reinforced associates, M_{diff} = .06, $t(45)$ = 4.50, $p < .01$, BF = 456.05, g_{av} = .66, 95% CI [.03, .09], and derived associates, M_{diff} = .08, $t(45)$ = 4.43, $p < .01$, BF = 380.55, g_{av} = .82, 95% CI [.04, .11]. The response-congruency effect did not differ between reinforced and derived associates, M_{diff} = .02, $t(45)$ = 1.18, p = .25, BF = .31, g_{av} = .18, 95% CI

[−.01, .04]. The stimulus-congruency effect was not significant for any of the distractor types: color words M_{diff} = .003, $t(45)$ = .46, p = .65, BF = .18, g_{av} = .06, 95% CI [−.01, .02]; reinforced associates, M_{diff} = .01, $t(45)$ = 1.55, p = .13, BF = .48, g_{av} = .22, 95% CI [.00, .03]; derived associates: M_{diff} = .004, $t(45)$ = .34, p = .73, BF = .17, g_{av} = .05, 95% CI [−.02, .03].

Additional analysis. As in the previous experiments, we also explore the RT distribution by using delta plots. As shown in Figure 8, the stimulus-congruency effects observed for the reinforced and derived associates did not vary consistently as a function of mean RT. Although this effect was numerically present in the last two bins of the derived associates, additional contrasts indicated that these differences were not reliable (Bin 4: M_{diff} = 16.06, $t(45)$ = 1.14, p = .36, BF = .30, g_{av} = .08, 95% CI [−12.24, 44.35]; Bin 5: M_{diff} = 16.95, $t(45)$ = .75, p = .46, BF = .31, g_{av} = .06, 95% CI [−28.35, 62.23]).

Discussion

In line with the previous experiments, performance on the conditional-discrimination test suggests that two equivalence classes were formed. The results of the Stroop task replicate those of Experiment 3. With regard to the response-congruency effect, RTs and PCs converged toward the same pattern of results. For the color words a sizable response-congruency effect was obtained. Response-congruency effects were also observed for derived and reinforced associates. The response-congruency effects for the associates did not differ in size, but were substantially smaller compared to the response-congruency effect observed for the color words. With respect to the stimulus-congruency effect, only RTs measured on the color words offered moderate evidence for the presence of such an effect. RTs did not offer any evidence for the presence of stimulus-congruency effects on the reinforced and derived associates. In addition, the stimulus-congruency observed for the color words was significantly larger than the stimulus-congruency effects observed for the reinforced and derived associates. Finally, PCs did not provide support for a stimulus-congruency effect on either of the distractor types.

The results of Experiment 4 thus seem to suggest that even drastically increasing the amount of conditional-discrimination training does not result in the presence of stimulus-congruency effects for reinforced and derived associates. Accordingly, the response bias triggered by these associates seems mainly driven by response conflict and not by stimulus conflict. This again suggests that associates do not access a semantic representation of the color they were related to. Before discussing the implications of these results, we first consider an alternative explanation of the current results in a final experiment.

Experiment 5

Closer inspection of the Stroop effects obtained in Experiments 3 and 4 indicates that the observed response-congruency effects were exceptionally large in comparison to the Stroop effects observed in Experiment 2. One possibility is that the response-congruency effects in these latter experiments were inflated by online learning during the Stroop task. More precisely, when fully crossing target colors and distractor words in a two-to-one Stroop task, one quarter of the trials are identity trials, one quarter same-response trials, and one half are different response trials. In Experiments 3 and 4, however, the Stroop task consisted of one third of identity trials, one third of same-response trials, and one third of different-response trials. This balance was created by reducing the number of different-response trials. Consequently, distractor words may have come to elicit the response that they were more frequently related to, namely, the response required by the identity

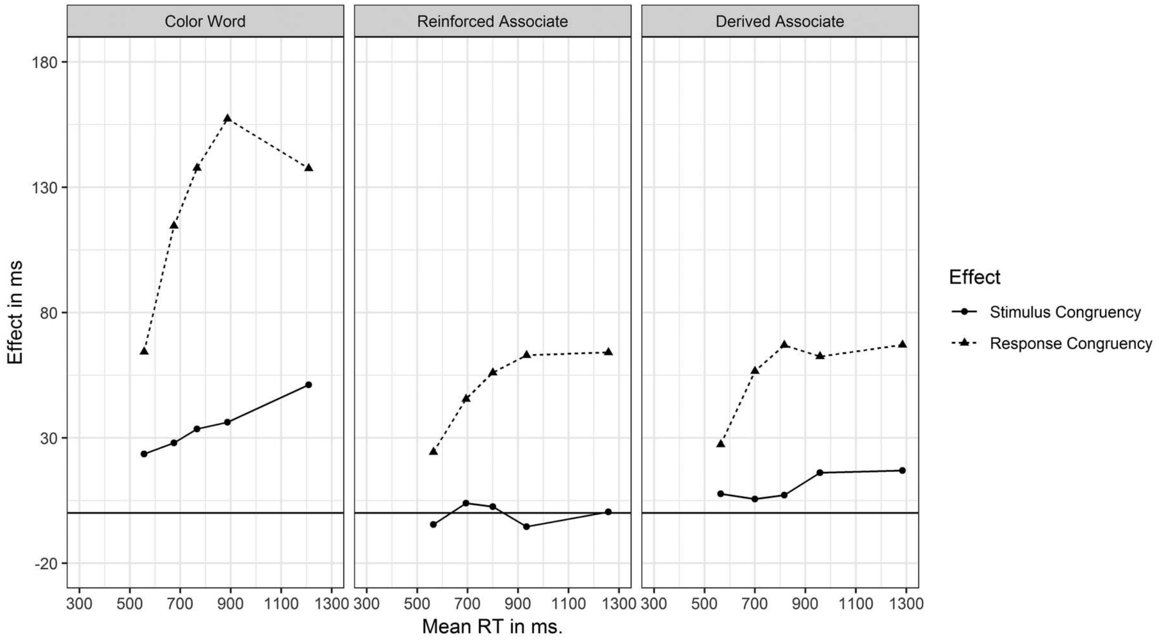

Figure 8. Difference in reaction time (RT) between congruent and incongruent trials of Experiment 4, calculated per bin and per distractor type, and plotted as a function of the mean bin RT per distractor type.

and same-response trials (i.e., two thirds of trials). Such contingency learning has been documented extensively before by Schmidt and colleagues (e.g., Schmidt & Besner, 2008; Schmidt, De Houwer, & Besner, 2010; Schmidt & De Houwer, 2012a, 2012b, 2012c) and may have biased the findings of Experiments 3 and 4. Indeed, the response-congruency effect in these experiments may not only reflect the interference between the response linked to the distractor word and the response linked to the target color, but also the fact that distractor words were selectively paired to a greater extent with one response (one third identity trials, one third same-response trials) compared with another (one third different-response trials). Concerned with this issue, we conducted an additional experiment that is reported in the Appendix. In this experiment, there was no training phase and only the Stroop task used in Experiments 3 and 4 was administered. However, as can be seen in the Appendix, the results of this additional experiment mirror the findings of Experiments 3 and 4. As no conditional-discrimination training preceded the Stroop task, these results indicate that contingency learning may have inflated the effects that we attributed to reinforced and derived associates. In view of the results of Experiment 2, in which no contingency learning could have biased the results because colors and distractor words were fully crossed, it seems unlikely that contingency learning may have completely induced the Stroop-effects observed for reinforced and derived associates in Experiments 3 and 4. In order to control for the bias induced by contingency learning, the Stroop-task used in Experiment 5 was adapted such that an equal number of distractor word/target color combinations was presented, which resulted in one fourth of identity trials, one fourth of same-response trials, and one half of different-response trials. Consequently, each distractor word was related to each response an equal number of times, meaning that contingency learning could not bias our results. The crucial question was whether we would observe the same pattern of results for the reinforced and derived associates as we did in the previous two experiments (i.e., a response-congruency effect but no stimulus-congruency effect).

Method

Participants. Fifty-six new participants took part in exchange for 20 euros.

Procedure. Experiment 5 was similar to Experiment 4 with one exception. Each distractor word during the Stroop task (two color words, two reinforced associates, two derived associates, two filler words) was presented in each color (blue, magenta, yellow, gray) an equal number of times. Doing so required a total of 768 trials, in which 48 identity trials, 48 same-response trials, and 96 different-response trials were obtained for each type of distractor word as well as 192 no-go trials (i.e., filler words as distractor). The Stroop task was subdivided in six blocks of 128 trials, with a small break after each block.

Results

Conditional-discrimination test. Three participants failed to meet the inclusion criterion (.63, .49, .62) and were excluded from subsequent analyses. A main effect of relation type was significant, $F(2, 104) = 8.56$, $MSE = .001$, $p < .001$, $\eta_p^2 = .14$, $BF = 70.28 \pm 0.98\%$. PCs were higher for the directly reinforced ($M = .93$; $SD = .07$), compared with symmetry ($M = .90$; $SD = .07$), $M_{diff} = .04$, $t(52) = 4.58$, $p < .01$, $BF = 686.56$, $g_{av} = .57$, 95% CI [.02, .06], and transitivity relations ($M = .88$; $SD = .11$), $M_{diff} = .05$, $t(52) = 3.25$, $p < .01$, $BF = 14.92$, $g_{av} = .53$, 95% CI [.02, .08], while the latter two relations did not differ from one another, $M_{diff} = .01$, $t(52) = .79$, $p = .43$, $BF = .20$, $g_{av} = .11$, 95% CI [$-.02$, .04].

Stroop task. Cell means and corresponding standard errors are presented in Figure 6. For the RTs, the main effect of trial type was significant, $F(2, 104) = 41.66$, $MSE = 2{,}300$, $p < .001$, $\eta_p^2 = .44$, $BF > 1{,}000 \pm 2.39\%$. The stimulus-congruency effect was not significant, $M_{diff} = 2.48$, $t(52) = .41$, $p = .68$, $BF = .16$, $g_{av} = .02$, 95% CI [-9.58, 15.54], whereas the response-congruency effect was significant, $M_{diff} = 41.56$, $t(52) = 8.27$, $p < .001$, $BF > 1{,}000$, $g_{av} = .30$, 95% CI [31.47, 51.65]. The main effect of distractor type was also significant, $F(2, 104) = 35.36$, $MSE = 3343$, $p < .001$, $\eta_p^2 = .40$, $BF > 1{,}000 \pm 1.66\%$. RTs were shorter on color words compared to reinforced associates, $M_{diff} = 35.23$, $t(52) = 5.28$, $p < .01$, $BF > 1{,}000$, $g_{av} = .25$, 95% CI [21.84, 48.63], and derived associates, $M_{diff} = 35.73$, $t(52) = 5.14$, $p < .001$, $BF > 1{,}000$, $g_{av} = .26$, 95% CI [21.77, 49.68]. RTs did not differ significantly between reinforced and derived associates, $M_{diff} = .50$, $t(52) = .09$, $p = .93$, $BF = .15$, $g_{av} = .00$, 95% CI [-10.42, 11.41].

The interaction between trial type and distractor type was significant, $F(4, 208) = 22.53$, $MSE = 1737$, $p < .001$, $\eta_p^2 = .30$, $BF > 1{,}000 \pm 1.78\%$ (see Figure 9). For the color words, both the stimulus-congruency effect, $M_{diff} = 21.18$, $t(52) = 2.21$, $p < .05$, $BF = 1.40$, $g_{av} = .16$, 95% CI [-1.97, 40.38], and the response-congruency effect, $M_{diff} = 79.91$, $t(52) = 10.96$, $p < .001$, $BF > 1{,}000$, $g_{av} = .58$, 95% CI [65.28, 94.53], were significant. For the reinforced associates, the response-congruency effect was significant, $M_{diff} = 26.15$, $t(52) = 3.75$, $p < .01$, $BF = 57.68$, $g_{av} = .18$, 95% CI [12.15, 40.15], whereas the stimulus-congruency effect was not, $M_{diff} = 2.90$, $t(52) = .37$, $p = .72$, $BF = .16$, $g_{av} = .02$, 95% CI [-13.03, 18.83]. Likewise, for the derived associates the response-congruency effect was significant, $M_{diff} = 17.87$, $t(52) = 2.23$, $p < .05$, $BF = 1.43$, $g_{av} = .13$, 95% CI [1.75, 33.99], but not the stimulus-congruency effect, $M_{diff} = 11.29$, $t(52) = 1.32$, $p = .19$, $BF = .34$, $g_{av} = .08$, 95% CI [-5.87, 28.46].

The response-congruency effect was larger for the color words than reinforced associates, $M_{diff} = 53.76$, $t(52) = 6.07$, $p < .001$, $BF > 1{,}000$, $g_{av} = 1.03$, 95% CI [35.97, 71.54], and the derived associates, $M_{diff} = 62.04$, $t(52) = 6.35$, $p < .01$, $BF > 1{,}000$, $g_{av} = 1.10$, 95% CI [42.42, 81.65]. The response-congruency effect did not differ between derived and reinforced associates, $M_{diff} = 8.28$, $t(52) = .85$, $p = .40$, $BF = .21$, $g_{av} = .15$, 95% CI [-11.30, 27.87]. The stimulus-congruency effect was larger for the color words compared to the reinforced associates, $M_{diff} = 24.08$, $t(52) = 2.24$, $p < .05$, $BF = 1.48$, $g_{av} = .38$, 95% CI [2.52, 45.63], and derived associates, $M_{diff} = 32.47$, $t(52) = 2.89$, $p < .01$, $BF = 6.13$, $g_{av} = .49$, 95% CI [9.96, 54.98]. The stimulus-congruency effect did not differ between reinforced and derived associates, $M_{diff} = 8.39$, $t(52) = .77$, $p = .45$, $BF = .19$, $g_{av} = .14$, 95% CI [-30.31, 13.52].

With respect to the PCs, the main effect of trial type was significant, $F(2, 104) = 28.75$, $MSE = .001$, $p < .001$, $\eta_p^2 = .36$, $BF > 1{,}000 \pm 4.17\%$. The stimulus-congruency effect was not

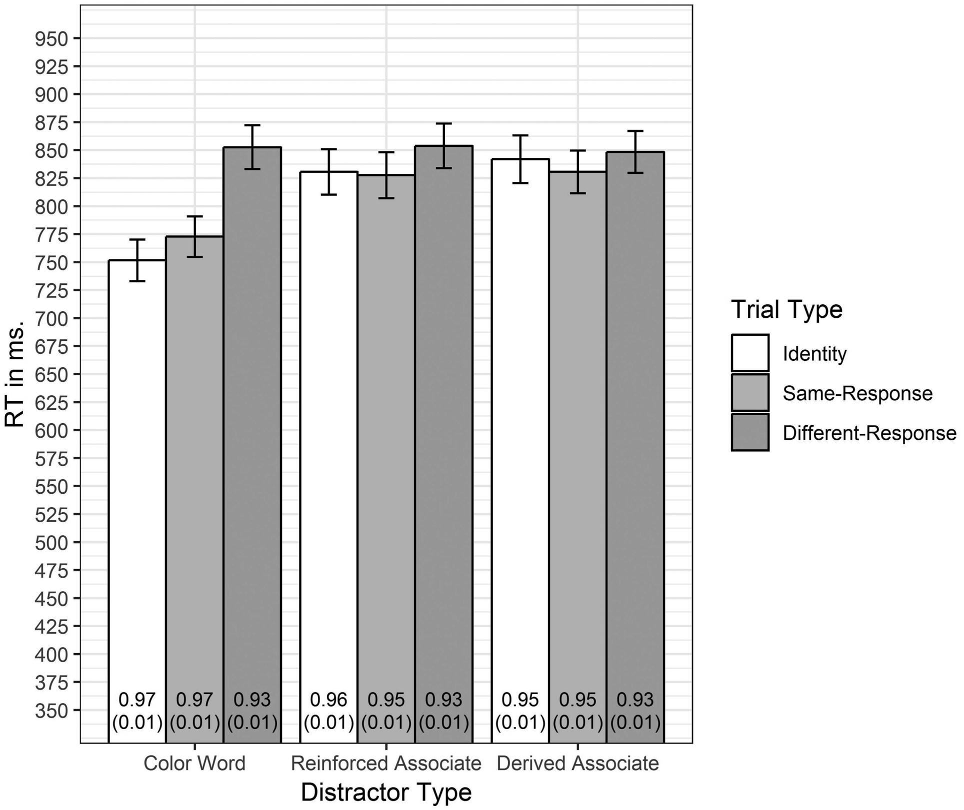

Figure 9. Reaction times (RTs) and PCs of Experiment 5 as a function of trial type and distractor type. Error bars denote the standard errors. Standard errors for PCs are printed between brackets.

significant, M_{diff} = .00, $t(52)$ = .13, p = .89, BF = .15, g_{av} = .02, 95% CI [−.01, .01], while the response-congruency effect was, M_{diff} = .03, $t(52)$ = 6.05, $p < .01$, $BF > 1{,}000$, g_{av} = .77, 95% CI [.02, .04]. The main effect of distractor type was also significant, $F(2, 104)$ = 3.69, MSE = .001, $p < .05$, η^2_p = .07, BF = 0.42 ± 4.15%. PCs were higher for color words compared with derived associates, M_{diff} = .03, $t(52)$ = 6.20, $p < .01$, $BF > 1{,}000$, g_{av} = .80, 95% CI [.02, .04], but not compared to reinforced associates, M_{diff} = .00, $t(52)$ = .13, p = .89, BF = .15, g_{av} = .02, 95% CI [−.01, .01]. PCs did not differ significantly between reinforced and derived associates, M_{diff} = .00, $t(52)$ = .95, p = .34, BF = .23, g_{av} = .10, 95% CI [.00, .01].

This time the interaction between trial type and distractor type was not significant, $F(4, 208)$ = 1.71, MSE = .001, p = .15, η^2_p = .03, BF = 0.17 ± 4.46%. Given our research questions, we explored whether a stimulus- and/or response-congruency effect was reliably present for each type of distractor word. The stimulus-congruency effect did not reach statistical significance for any of the distractor words: color words: M_{diff} = .00, $t(52)$ = .63, p = .53, BF = .18, g_{av} = .09, 95% CI [−.01, .02]; reinforced associates: M_{diff} = .01, $t(52)$ = .94, p = .35, BF = .23, g_{av} = .15, 95% CI [.01, .02]; derived associates: M_{diff} = .00, $t(52)$ = .56, p = .58, BF = .17, g_{av} = .09, 95% CI [−.01, .02]. In contrast, the response-congruency effect was significant in all three cases: color words: M_{diff} = .04, $t(52)$ = 6.43, $p < .01$, $BF > 1{,}000$, g_{av} = .97, 95% CI [.03, .06]; reinforced associates: M_{diff} = .02, $t(52)$ = 2.44, $p < .05$, BF = 2.31, g_{av} = .43, 95% CI [.00, .04]; derived associates: M_{diff} = .02, $t(52)$ = 2.77, $p < .01$, BF = 4.62, g_{av} = .43, 95% CI [.01, .04].

Additional analysis. Delta plots (see Figure 10) indicate that the absence of a stimulus-congruency effect was consistent across RT-bins, both for reinforced and derived associates. If anything, a slightly reversed stimulus-congruency effect was present for the slower RT bins of both distractor types. These reversed effects were, however, not reliable (reinforced associates: M_{diff} = −14.25, $t(52)$ = .71, p = .48, BF = .19, g_{av} = .06, 95% CI [−54.69, 26.19]; derived associates: M_{diff} = −13.58, $t(52)$ = .57, p = .57, BF = .17, g_{av} = .05, 95% CI [−61.69, 34.54]).

Discussion

Performance on the conditional-discrimination test again suggested that we succeeded in establishing two equivalence classes. More importantly, Experiment 5 made it impossible for online contingency learning to bias the results of the Stroop task. We first consider the RT data. For the color words, a sizable response-congruency effect was observed. The stimulus-congruency effect was significant, but only anecdotal evidence was obtained for the alternative hypothesis (see also Experiment 3). For the reinforced associates, moderate evidence was obtained in support of a response-congruency effect, while the stimulus-congruency effect was genuinely absent. For the derived associates, the response-

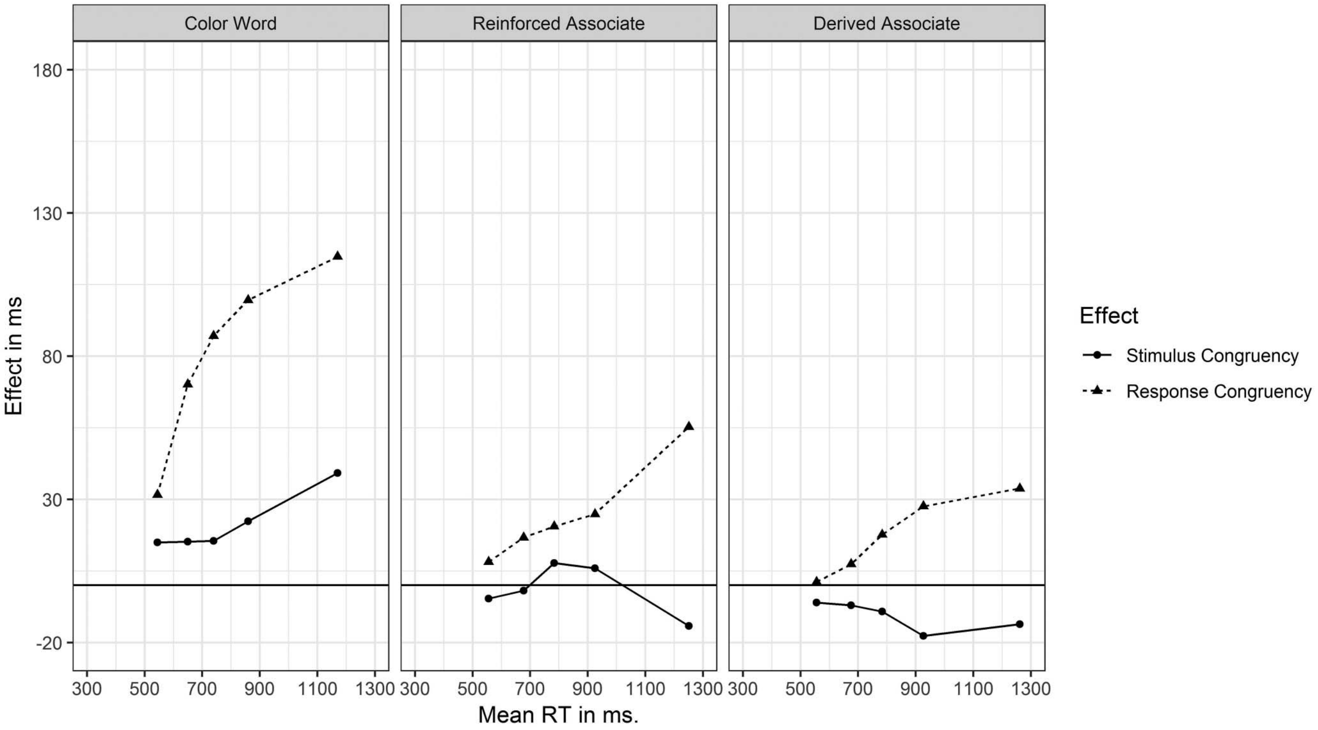

Figure 10. Difference in reaction time (RT) between congruent and incongruent trials of Experiment 5, calculated per bin and per distractor type, and plotted as a function of the mean bin RT per distractor type.

congruency effect was significant, but again only anecdotal evidence was obtained. The stimulus-congruency effect was not significant and the corresponding Bayes factor indicated that the null hypothesis was very likely. The response-congruency effect observed for the color words was larger than the response-congruency effect observed for reinforced and derived associates. The latter effects did not differ. The stimulus-congruency effect observed for the color words was also larger compared with the stimulus-congruency effect on the associates.

For the PCs, the response-congruency effect was significant for all three distractor types. Strong evidence was obtained for a response-congruency effect in the color words, and anecdotal (reinforced associates) to moderate (derived associates) evidence for the associates. Finally, PCs did not offer evidence for a stimulus-congruency effect. Taken together, the findings of Experiment 5 are consistent with those of Experiments 3–4: Derived and reinforced associates can induce a Stroop effect, which is driven by response conflict and not by stimulus conflict. Once again this supports the conclusion that conditional-discrimination training leads to connections between the lexical representation of color words and their associates, but these associates are not able to directly access the semantic representation of their corresponding color.

Bayesian Meta-Analysis

A potential concern is that Experiments 3–5 were not sensitive enough to detect stimulus-congruency effects for reinforced and derived associates. Although the Bayesian *t* tests indicated that the absence of such effect was more likely given the observed data, we have to acknowledge that the samples in Experiments 3–5 may not have been large enough to detect small effects. A sensitivity analysis was conducted to assess the minimum effect size (dz) that could be detected in Experiments 3–5 with 80% power. The minimum effect sizes were .34 (Experiment 3), .37 (Experiment 4), and .35 (Experiment 5), which suggests that our experiments were primarily sensitive to detect medium-sized effects. In view of this, we conducted a Bayesian Meta-Analysis (e.g., Rouder & Morey, 2011) with the package “BayesFactor,” using the default JZS prior (.707; Morey et al., 2015). The Bayes factors obtained through this analysis are based on a combination of the samples of Experiments 3–5. As can be seen in Table 1, strong evidence was obtained for response-congruency effects for all three distractor types and this both for the RTs and the PCs. RTs also provided considerable evidence in support of the presence of a stimulus-congruency effect in the color words. PCs did not offer such

Table 1
Bayes Factors of the Meta-Analysis Performed on the Data of Experiments 3–5

Congruency effect	Color words	Reinforced	Derived
Stimulus congruency			
RTs	681.44	.06	.07
PCs	.11	.38	.08
Response congruency			
RTs	>1,000	>1,000	>1,000
PCs	>1,000	>1,000	>1,000

support. Importantly, even combining the data of Experiments 3–5 did not offer convincing evidence for the presence of a stimulus-congruency effect for the reinforced and derived associates, neither in terms of RTs or PCs. In conclusion, this small meta-analysis confirms the findings obtained in each experiment separately.

General Discussion

The present study aimed to extend current knowledge on the relation between automaticity and practice by testing whether automatic effects can also be induced by new stimulus-stimulus relations, which were formed on the basis of derivation, and in the absence of direct reinforcement. To test this idea, we first administered conditional-discrimination training with the goal of establishing two equivalence classes. Each class consisted of a color word, a reinforced associate, and a derived associate. Following this training, a conditional-discrimination test and a Stroop task were administered. In all five experiments, the results of the conditional-discrimination test indicated that we succeeded in establishing two equivalence classes. The results of the Stroop task can be summarized as follows. In Experiment 1, only negligible Stroop effects were observed, which led to the conclusion that the distractor words were not sufficiently processed during the Stroop task. Accordingly, in Experiment 2 the Stroop task was combined with a go/no-go task in which participants had to refrain from responding to Stroop stimuli when specific distractors were presented (i.e., participants were now required to process the distractor words). In this experiment, substantial Stroop effects were obtained for the color words, as well as for the reinforced and derived associates. In Experiment 3, we tested whether these effects were driven by stimulus conflict, response conflict, or both. The results indicated the presence of both types of conflict for color words, but only evidence for a response conflict was obtained for the associates. A similar pattern of results was obtained when substantially increasing the amount of training (Experiment 4) and when controlling for biases produced by online contingency learning (Experiment 5). Finally, the Stroop effects obtained for derived associates were always similar to the effects obtained for the reinforced associates.

Based on these findings we propose that conditional-discrimination training resulted in the formation of connections between lexical representations of the nonwords and their corresponding color words. When sufficient attention is allocated to the nonwords during the Stroop task, they are automatically (in the sense of fast and unintentionally) translated into the corresponding color word, which triggers the retrieval of the response assigned to that color via the instructions of the Stroop task. The observation that overall RTs were similar for the reinforced and derived color associates suggests that both associates were directly connected to the color words and that the translation occurred in a single step (i.e., derived associate → color word; reinforced associate → color word). This again indicates that directly reinforced relations and derived relations are characterized by similar representations in memory. However, in both cases we failed to observe a stimulus conflict as indexed by the stimulus-congruency effect. Although the conditional-discrimination training was extensive in Experiments 4 and 5, the associates did not seem to be directly related to a semantic representation of the color that the corresponding color word was referring to. In addition, RT distribution analyses or a small meta-analysis did not offer additional support for the presence of stimulus-congruency effects reinforced and derived associates. Taken together, our findings fit within the proposals of Kroll and Stewart (1994) on second-language learning, who argue that newly acquired second-language words are only connected at the lexical level with their first-language equivalents, but can only access semantic representations when language proficiency is sufficiently high.

Following the interpretation we offer for the results of Experiments 3–5, it could be argued that the response-congruency effects observed for color words and their associates are based on the same processing route. For color words, a lexical representation leads to the activation of a semantic representation of the corresponding color, which in turns leads to response activation. For the associates, a lexical representation of the associated nonword first activates a lexical representation of the corresponding color word, from which the same processing route is activated as for the color words. Following this line of reasoning, it could be predicted that the response-congruency effects for the color words are highly correlated with the response-congruency effects observed for their associates. Additional analyses, however, do not seem to confirm this hypothesis. In general, correlations between both types of effects was low.[2] When combining the results of all three experiments by using a Bayesian meta-analysis, we obtained no evidence in support of a correlation between the color words and the reinforced associates ($BF < 1$) and only anecdotal evidence for a correlation between color words and derived associates ($BF = 1.35$). Although one should be careful in interpreting correlations between difference scores as they tend to be unreliable (e.g., see Hedge, Powell, & Sumner, 2018 for a discussion of this issue in the context of cognitive-control research), the absence of such correlations may indicate that the processes underlying the Stroop effect and the processes underlying the automatic effect of derived and reinforced associates are perhaps more distinct than we initially assumed.

When considering previous research on the contribution of response conflict and stimulus conflict to congruency effects in different variations of the Stroop task, our findings are especially in line with research investigating automatic effects of color word contingencies. As mentioned in the Introduction, color–word contingency learning consists of presenting distractor words more often in one color than in another color (e.g., Schmidt et al., 2007; Schmidt & Besner, 2008; Schmidt & De Houwer, 2011, 2012a, 2012b, 2012c). Typically, faster responses are observed when a frequent color word contingency is presented as compared to when a less frequent color word contingency is presented. Schmidt et al. (2007; see also, Schmidt & De Houwer, 2012a, 2012b, 2012c) observed that such congruency effects are mainly driven by response conflict, which led to the conclusion that the biased word–color contingencies resulted in the distractor word becoming directly related to the response required by the color-judgment task. Similarly, the present results indicate that reinforced and derived associates became related to the response of the Stroop task. On the other hand, Schmidt and Cheesman (2005) used semantic-color associates as distractor words in a Stroop task with a two-to-one response mappings, such as the word SKY as an associate for the color blue. Schmidt and Cheesman (2005) observed that semantic associates only triggered a semantic conflict but no response

---

[2] Color word versus reinforced associate: .25 (Experiment 3), .01 (Experiment 4), .23 (Experiment 5). Color word versus derived associate: .42 (Experiment 3), .06 (Experiment 4), .19 (Experiment 5).

conflict. Based on these findings, these authors concluded that semantic associates spread activation to related concepts in semantics, producing semantic conflict with the target color concept, but are not potent enough to indirectly bias a potential response (e.g., "sky" facilitating "blue" strongly enough to retrieve a left-key response linked to "blue" via instructions). Finally, our results are at odds with findings of Schmidt et al. (2018), who observed both a stimulus and a response conflict when presenting French color words as distractors to Dutch-speaking participants.

The aforementioned findings may suggest that the current findings are limited by the specific conditional-discrimination training we used. First, the amount of conditional-discrimination training used in the current experiments may not have been extensive enough for participants to consider the nonwords as being semantically equivalent to the color words. Semantic associates (e.g., SKY for the color blue) are the product of a lifelong training history. Similarly, the Dutch-speaking participants in the study of Schmidt et al. (2018) already had some prior knowledge of French prior to the experiment, perhaps a sufficient amount for stimulus conflict to emerge. Although the conditional-discrimination training in Experiments 4–5 consisted of up to 800 trials, we cannot rule out the possibility that more extensive training may result in stimulus conflicts. For instance, Logan and Klapp (1991) had to administer multiday training schemes in order for participants to solve an alpha-numerical task automatically. In such a task, participants are required to add a letter to a number (e.g., $4 + F = ?$). To do so, they need to substitute the letter with its corresponding rank order in the alphabet (i.e., $F = 6$). Only extended practice enabled participants to do this automatically. Second, the nature of the conditional-discrimination training may also be of importance, independent of its length. The training task used here required participants to select a nonword or a color word upon presentation of another nonword. When considering the distinction between lexical and semantic representations of words (e.g., Kroll & Stewart, 1994), this task mainly requires lexical processing: one word form needs to be related to another word form. Accordingly, the conclusion that the Stroop effect we observed for reinforced and derived associates is based on the translation of these associates into their corresponding color word may not be surprising. In the end, the conditional-discrimination training mainly targeted the formation of relations between lexical representations of color words and nonwords. In addition, the nonwords we used were selected arbitrarily and are unlikely to be considered as semantic associates by the participants, even after extensive training. In other words, the presence of stimulus conflict may arise when adapting the training so that the nonwords are more explicitly related to the conceptual representation of a particular color and a context is adopted, which offers a more ecological valid analogue of language learning.

Based on the assumption that the automatic effects of color words and their associates rely on the same processing route, it could be assumed that improving conditional-discrimination training results in the elimination of the differences between the Stroop-effects observed for color words and their associates. At the same time, the alternative hypothesis needs to be considered, namely that both types of effects are underlain by different processes. In the present study, the automatic influence of color words and color associates was assessed by contrasting congruent and incongruent trials. We note, however, that the Stroop literature also addresses much more complex data patterns. The comparison between color words and associates could thus be extended in several ways. For instance, Stroop effects are supposedly underlain by response facilitation on congruent trials and response interference in incongruent trials (e.g., Lindsay & Jacoby, 1994; but see Brown, 2011 for a discussion). The question thus arises whether both components are also present for associates? Furthermore, Stroop effects are sequentially modulated: Stroop effects are smaller on trials following a Stroop-incongruent trial (e.g., Notebaert, Gevers, Verbruggen, & Liefooghe, 2006). Although the nature of this modulation effect is still under debate (see Egner, 2007 for a review), the sequential analysis of properly constructed Stroop tasks may also offer the basis of a more fine-grained comparison between color words and associates. Both examples thus indicate that future research will be needed to further investigate the communalities and differences between Stroop-effects and automatic effects of derived stimulus–stimulus relations.

To conclude, our results indicate that derived stimulus-stimulus relations can trigger an automatic effect to the same extent as directly reinforced stimulus-stimulus relations. This finding indicates that direct reinforcement is not a prerequisite for inducing automaticity on the basis of novel contingencies. Such a finding is clearly challenging for current theories of automaticity. For instance, the question arises how we can account for the current findings in terms of episodic-retrieval models of automaticity (e.g., Logan, 1985, 1988; Schmidt, De Houwer, & Rothermund, 2016). At the same time, we acknowledge the limits of the current findings. Automatic effects of derived associates only induced a response conflict when participants were encouraged to process the distractor words. It becomes clear that overcoming the current limits of reinforced and derived stimulus-stimulus relations also imposes a challenge for future research. In addition, the fact that research on second-language learning frequently demonstrated that second-language words can access their corresponding semantic representations even when second-language proficiency is low (e.g., Altarriba & Mathis, 1997; Duyck & De Houwer, 2008; see Brysbaert & Duyck, 2010, for a in depth discussion) suggests that there should be conditions under which derived learning can result in semantic effects. Future research is needed to establish what these conditions are.

References

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36,* 550–568. http://dx.doi.org/10.1006/jmla.1997.2493

Ambrosi, S., Servant, M., Blaye, A., & Burle, B. (2019). Conflict processing in kindergarten children: New evidence from distribution analyses reveals the dynamics of incorrect response activation and suppression. *Journal of Experimental Child Psychology, 177,* 36–52. http://dx.doi.org/10.1016/j.jecp.2018.06.006

Augustinova, M., & Ferrand, L. (2014). Automaticity of word reading: Evidence from the semantic Stroop paradigm. *Current Directions in Psychological Science, 23,* 343–348. http://dx.doi.org/10.1177/0963721414540169

Augustinova, M., Silvert, L., Ferrand, L., Llorca, P. M., & Flaudias, V. (2015). Behavioral and electrophysiological investigation of semantic and response conflict in the Stroop task. *Psychonomic Bulletin & Review, 22,* 543–549. http://dx.doi.org/10.3758/s13423-014-0697-z

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Barnes, D., & Holmes, Y. (1991). Radical behaviorism, stimulus equivalence, and human cognition. *The Psychological Record, 41,* 19–31. http://dx.doi.org/10.1007/BF03395091

Barnes-Holmes, D., Barnes-Holmes, Y., Smeets, P. M., Cullinan, V., & Leader, G. (2004). Relational frame theory and stimulus equivalence: Conceptual and procedural issues. *International Journal of Psychology & Psychological Therapy, 4,* 181–214.

Bauer, B., & Besner, D. (1997). Processing in the Stroop task: Mental set as a determinant of performance. *Canadian Journal of Experimental Psychology, 51,* 61–68. http://dx.doi.org/10.1037/1196-1961.51.1.61

Besner, D. (2001). The myth of ballistic processing: Evidence from Stroop's paradigm. *Psychonomic Bulletin & Review, 8,* 324–330. http://dx.doi.org/10.3758/BF03196168

Besner, D., & Stolz, J. A. (1999). Unconsciously controlled processing: The Stroop effect reconsidered. *Psychonomic Bulletin & Review, 6,* 449–455. http://dx.doi.org/10.3758/BF03210834

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review, 4,* 221–225. http://dx.doi.org/10.3758/BF03209396

Brown, T. L. (2011). The relationship between Stroop interference and facilitation effects: Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 85–99. http://dx.doi.org/10.1037/a0019252

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition, 13,* 359–371. http://dx.doi.org/10.1017/S1366728909990344

Bub, D. N., Masson, M. E., & Lalonde, C. E. (2006). Cognitive control in children: Stroop interference and suppression of word reading. *Psychological Science, 17,* 351–357. http://dx.doi.org/10.1111/j.1467-9280.2006.01710.x

Chen, A., Bailey, K., Tiernan, B. N., & West, R. (2011). Neural correlates of stimulus and response interference in a 2–1 mapping Stroop task. *International Journal of Psychophysiology, 80,* 129–138. http://dx.doi.org/10.1016/j.ijpsycho.2011.02.012

Christ, S. E., Holt, D. D., White, D. A., & Green, L. (2007). Inhibitory control in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37,* 1155–1165. http://dx.doi.org/10.1007/s10803-006-0259-y

Cullinan, V. A., Barnes, D., Hampson, P. J., & Lyddy, F. (1994). A transfer of explicitly and non-explicitly trained sequence responses through equivalence relations: An experimental demonstration and connectionist model. *The Psychological Record, 44,* 559–585. http://dx.doi.org/10.1007/BF03395144

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition, 31,* 353–359. http://dx.doi.org/10.3758/BF03194393

de Rose, J. C., de Souza, D. G., Rossito, A. L., & de Rose, T. M. S. (1992). Stimulus equivalence and generalization in reading after matching to sample by exclusion. In S. C. Hayes & L. J. Hayes (Eds.), *The International Institute on Verbal Relations: Understanding verbal relations* (pp. 69–82). Reno, NV: Context Press.

Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: Evidence from the semantic Simon paradigm. *Psychonomic Bulletin & Review, 15,* 961–966. http://dx.doi.org/10.3758/PBR.15.5.961

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective & Behavioral Neuroscience, 7,* 380–390. http://dx.doi.org/10.3758/CABN.7.4.380

Ferrand, L., & Augustinova, M. (2014). Differential effects of viewing positions on standard versus semantic Stroop interference. *Psychonomic Bulletin & Review, 21,* 425–431. http://dx.doi.org/10.3758/s13423-013-0507-z

Geukes, S., Gaskell, M. G., & Zwitserlood, P. (2015). Stroop effects from newly learned color words: Effects of memory consolidation and episodic context. *Frontiers in Psychology, 6,* 278. http://dx.doi.org/10.3389/fpsyg.2015.00278

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 875–894. http://dx.doi.org/10.1037/0096-1523.8.6.875

Hasshim, N., & Parris, B. A. (2015). Assessing stimulus–stimulus (semantic) conflict in the Stroop task using saccadic two-to-one color response mapping and preresponse pupillary measures. *Attention, Perception, & Psychophysics, 77,* 2601–2610.

Hayes, S. C., & Hayes, L. J. (1992). Some clinical implications of contextualistic behaviorism: The example of cognition. *Behavior Therapy, 23,* 225–249. http://dx.doi.org/10.1016/S0005-7894(05)80383-1

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50,* 1166–1186. http://dx.doi.org/10.3758/s13428-017-0935-1

Hughes, S., & Barnes-Holmes, D. (2016). Relational Frame Theory: Implications for the study of human language and cognition. In S. Hayes, D. Barnes-Holmes, R. Zettle, & T. Biglan (Eds.), *Handbook of contextual behavioral science*. New York, NY: Wiley-Blackwell.

Jensen, A. R., & Rohwer, W. D., Jr. (1966). The Stroop color-word test: A review. *Acta Psychologica, 25,* 36–93. http://dx.doi.org/10.1016/0001-6918(66)90004-7

Jongen, E. M., & Jonkman, L. M. (2008). The developmental pattern of stimulus and response interference in a color-object Stroop task: An ERP study. *BMC Neuroscience, 9,* 82. http://dx.doi.org/10.1186/1471-2202-9-82

Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review, 125,* 59–82. http://dx.doi.org/10.1037/rev0000083

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology, 77,* 576–588. http://dx.doi.org/10.2307/1420768

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149–174. http://dx.doi.org/10.1006/jmla.1994.1008

Liefooghe, B., & De Houwer, J. (2016). A functional approach for research on cognitive control: Analysing cognitive control tasks and their effects in terms of operant conditioning. *International Journal of Psychology, 51,* 28–32. http://dx.doi.org/10.1002/ijop.12179

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 219–234. http://dx.doi.org/10.1037/0096-1523.20.2.219

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39,* 367–386. http://dx.doi.org/10.1037/h0080066

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527. http://dx.doi.org/10.1037/0033-295X.95.4.492

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195. http://dx.doi.org/10.1037/0278-7393.17.2.179

MacKinnon, D. P., Geiselman, R. E., & Woodward, J. A. (1985). The effects of effort on Stroop interference. *Acta Psychologica, 58,* 225–235. http://dx.doi.org/10.1016/0001-6918(85)90022-8

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203. http://dx.doi.org/10.1037/0033-2909.109.2.163

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimen-*

tal Psychology: Learning, Memory, and Cognition, 14, 126–135. http://dx.doi.org/10.1037/0278-7393.14.1.126

Manwell, L. A., Roberts, M. A., & Besner, D. (2004). Single letter coloring and spatial cuing eliminates a semantic contribution to the Stroop effect. *Psychonomic Bulletin & Review, 11,* 458–462. http://dx.doi.org/10.3758/BF03196595

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology, 67,* 263–287. http://dx.doi.org/10.1146/annurev-psych-122414-033550

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297–326. http://dx.doi.org/10.1037/0033-2909.132.2.297

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *BayesFactor: Computation of Bayes factors for common designs* (R package Version 0.9). Retrieved from https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf

Moutsopoulou, K., Yang, Q., Desantis, A., & Waszak, F. (2015). Stimulus-classification and stimulus-action associations: Effects of repetition learning and durability. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 68,* 1744–1757. http://dx.doi.org/10.1080/17470218.2014.984232

Notebaert, W., Gevers, W., Verbruggen, F., & Liefooghe, B. (2006). Top-down and bottom-up sequential modulations of congruency effects. *Psychonomic Bulletin & Review, 13,* 112–117. http://dx.doi.org/10.3758/BF03193821

Pfeuffer, C. U., Moutsopoulou, K., Pfister, R., Waszak, F., & Kiesel, A. (2017). The power of words: On item-specific stimulus-response associations formed in the absence of action. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 328–347. http://dx.doi.org/10.1037/xhp0000317

Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception & Psychophysics, 72,* 2013–2025. http://dx.doi.org/10.3758/APP.72.7.2013

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461. http://dx.doi.org/10.1037/0033-2909.86.3.446

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Richards, A., & Blanchette, I. (2004). Independent manipulation of emotion in an emotional Stroop task using classical conditioning. *Emotion, 4,* 275–281. http://dx.doi.org/10.1037/1528-3542.4.3.275

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18,* 682–689.

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

Schmidt, J. R., & Cheesman, J. (2005). Dissociating stimulus-stimulus and response-response effects in the Stroop task. *Canadian Journal of Experimental Psychology, 59,* 132–138. http://dx.doi.org/10.1037/h0087468

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

Schmidt, J. R., & De Houwer, J. (2011). Now you see it, now you don't: Controlling for contingencies and stimulus repetitions eliminates the Gratton effect. *Acta Psychologica, 138,* 176–186. http://dx.doi.org/10.1016/j.actpsy.2011.06.002

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,* 175–182. http://dx.doi.org/10.1027/1618-3169/a000141

Schmidt, J. R., & De Houwer, J. (2012b). Does temporal contiguity moderate contingency learning in a speeded performance task? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 65,* 408–425. http://dx.doi.org/10.1080/17470218.2011.632486

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

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

Schmidt, J. R., Hartsuiker, R. J., & De Houwer, J. (2018). Interference in Dutch-French Bilinguals. *Experimental Psychology, 65,* 13–22. http://dx.doi.org/10.1027/1618-3169/a000384

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

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25,* 128–142. http://dx.doi.org/10.3758/s13423-017-1230-y

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190. http://dx.doi.org/10.1037/0033-295X.84.2.127

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale, NJ: Erlbaum, Inc.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors Cooperative.

Sidman, M. (1997). Equivalence relations. *Journal of the Experimental Analysis of Behavior, 68,* 258–266. http://dx.doi.org/10.1901/jeab.1997.68-258

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74,* 127–146. http://dx.doi.org/10.1901/jeab.2000.74-127

Sidman, M. (2009). Equivalence relations and behavior: An introductory tutorial. *Analysis of Verbal Behavior, 25,* 5–17. http://dx.doi.org/10.1007/BF03393066

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22. http://dx.doi.org/10.1901/jeab.1982.37-5

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *afex: Analysis of factorial experiments* (R package Version). Retrieved from https://cran.r-project.org/package=afex

Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS windows platform. *Behavior Research Methods, 38,* 280–286. http://dx.doi.org/10.3758/BF03192779

Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *The Quarterly Journal of Experimental Psychology, 31,* 121–132. http://dx.doi.org/10.1080/14640747908400712

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662. http://dx.doi.org/10.1037/h0054651

Tzelgov, J., Porat, Z., & Henik, A. (1997). Automaticity and consciousness: Is perceiving the word necessary for reading it? *The American Journal of Psychology, 110,* 429–448. http://dx.doi.org/10.2307/1423567

van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI study. *Neu-*

roImage, 27, 497–504. http://dx.doi.org/10.1016/j.neuroimage.2005.04.042

Verbruggen, F., & Logan, G. D. (2008). Automatic and controlled response inhibition: Associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology: General, 137,* 649–672. http://dx.doi.org/10.1037/a0013170

Vincent, S. B. (1912). The function of the viborissae in the behavior of the white rat. *Behavioral Monographs, 1,* 7–81.

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46,* 361–413. http://dx.doi.org/10.1016/S0010-0285(02)00520-0

West, R. (2003). Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. *Neuropsychologia, 41,* 1122–1135. http://dx.doi.org/10.1016/S0028-3932(02)00297-X

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50,* 125–144. http://dx.doi.org/10.1901/jeab.1988.50-125

Appendix

Additional Experiment

In order to evaluate the extent to which online-contingency learning contributed to the findings of Experiments 3 and 4, the Stroop-task used in these experiments was administered without any conditional-discrimination training.

Method

Participants

Thirty-three new participants were recruited and paid 10 euro to take part.

Procedure

Participants only completed the manual Stroop task as described in Experiment 3. The experiment took approximately 45 min to complete.

Results

Three participants failed to perform the task in adequate way and had an overall accuracy of .00, .00, and .02. These participants were considered as outliers. Cell means and corresponding standard errors are presented in Figure A1. With respect to the RTs, a main effect of trial type emerged, $F(2, 64) = 50.27$, $MSE = 3354$, $p < .001$, $\eta^2_p = .61$, $BF > 1{,}000$. ± 2.01%. The stimulus-congruency effect was not significant, $M_{diff} = 15.72$, $t(32) = 1.63$, $p = .11$, $BF = .61$, $g_{av} = .10$, 95% CI [−35.39, 3.95], while the response-congruency effect was, $M_{diff} = 62.01$, $t(32) = 8.85$, $p < .001$, $BF > 1{,}000$, $g_{av} = .38$, 95% CI [47.74, 76.28]. The main effect of distractor type was also significant, $F(2, 64) = 9.29$, $MSE = 3354$, $p < .001$, $\eta^2_p = .22$, BF = 595.39 ± 1.88%. RTs were significantly shorter for color words compared to reinforced associates, $M_{diff} = 35.79$, $t(32) = 4.35$, $p < .001$, $BF = 199.17$, $g_{av} = .22$, 95% CI [19.02, 52.56] and derived associates, $M_{diff} = 34.98$, $t(32) = 3.76$, $p < .01$, $BF = 44.77$, $g_{av} = .21$, 95% CI [16.01, 53.94]. RTs did not differ between reinforced and derived associates, $M_{diff} = .81$, $t(32) = .10$, $p = .92$, $BF = .19$, $g_{av} = .01$, 95% CI [−16.27, 17.90].

The interaction between trial type and distractor type was significant, $F(4, 128) = 26.18$, $MSE = 1534$, $p < .001$, $\eta^2_p = .45$, $BF > 1{,}000$ ± 4.61%. For the color words, a stimulus-congruency effect, $M_{diff} = 22.75$, $t(32) = 2.12$, $p < .05$, $BF = 1.33$, $g_{av} = .15$, 95% CI [−.87, 44.62], and a response-congruency effect were observed, $M_{diff} = 127.61$, $t(32) = 10.94$, $p < .001$, $BF > 1{,}000$, $g_{av} = .78$, 95% CI [103.86, 151.37]. For the reinforced associates, the response-congruency effect was significant, $M_{diff} = 26.65$, $t(32) = 2.80$, $p < .05$, $BF = 4.97$, $g_{av} = .16$, 95% CI [7.27, 46.04], but not the stimulus-congruency effect, $M_{diff} = 5.18$, $t(32) = .41$, $p = .69$, $BF = .20$, $g_{av} = .03$, 95% CI [−20.91, 31.28]. Similarly, no stimulus-congruency effect was present for the derived associates, $M_{diff} = 18.30$, $t(32) = 1.58$, $p = .12$, $BF = .57$, $g_{av} = .10$, 95% CI [−5.30, 41.91], but a significant response-congruency effect was observed, $M_{diff} = 33.31$, $t(32) = 2.81$, $p < .01$, $BF = 5.05$, $g_{av} = .19$, 95% CI [9.15, 57.47].

A similar pattern of findings was evident for the PCs. A main effect of trial type was obtained, $F(2, 64) = 34.05$, $MSE = .001$, $p < .001$, $\eta^2_p = .52$, $BF > 1{,}000$ ± 1.31%. The response-congruency effect was significant, $M_{diff} = .07$, $t(32) = 6.71$, $p < .001$, $BF > 1{,}000$, $g_{av} = 1.47$, 95% CI [.05, .09], but not the stimulus-congruency effect, $M_{diff} = .01$, $t(32) = 1.14$, $p = .26$, $BF = .34$, $g_{av} = .19$, 95% CI [−.02, .00]. The main effect of distractor type was not significant, $F(2, 64) = .94$, $MSE = .001$, $p = .40$, $\eta^2_p = .03$, $BF = .05$ ± 1.27%.

The interaction between trial type and distractor type was significant, $F(4, 128) = 12.82$, $MSE = .001$, $p < .001$, $\eta^2_p = .29$, $BF > 1{,}000$ ± 1.52%. For all three distractor types a response-congruency effect was observed: color words, $M_{diff} = .11$, $t(32) = 6.51$, $p < .001$, $BF > 1{,}000$, $g_{av} = 1.61$, 95% CI [.08, .15]; reinforced associates, $M_{diff} = .05$, $t(32) = 4.05$, $p < .001$, $BF = 92.08$, $g_{av} = .96$, 95% CI [.03, .08]; derived associates, $M_{diff} = .05$, $t(32) = 4.81$, $p < .001$, $BF = 676.65$, $g_{av} = 1.01$, 95% CI [.03, .07]. In contrast, no significant stimulus-congruency effects were observed: $M_{diff} = .01$, $t(32) = .96$, $p = .34$, $BF = .29$, $g_{av} = .22$, 95% CI [−.01, .02]; reinforced associates, $M_{diff} = .02$, $t(32) = 1.84$, $p = .08$, $BF = .84$, $g_{av} = .43$, 95% CI [.00, .04]; derived associates, $M_{diff} = .01$, $t(32) = .92$, $p = .37$, $BF = .28$, $g_{av} = .17$, 95% CI [−.01, .02].

(*Appendix continues*)

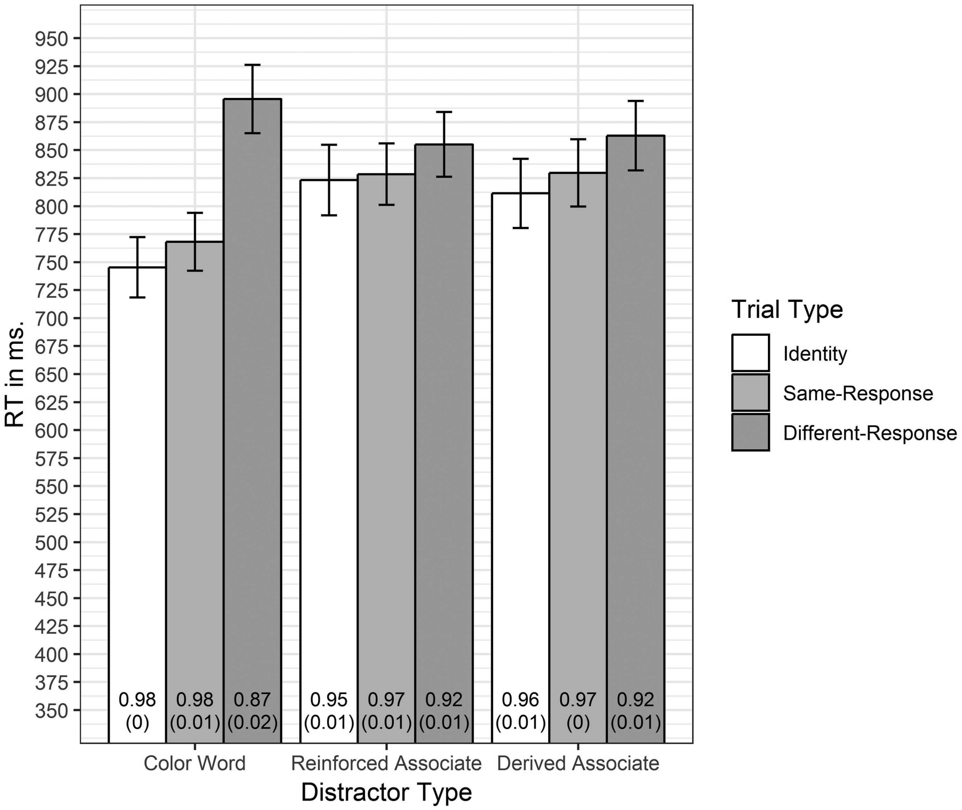

Figure A1. RTs and PCs of the experiment reported in the Appendix, as a function of trial type and distractor type. Error bars denote the standard errors. Standard errors for PCs are printed between brackets.

Received October 23, 2018
Revision received February 21, 2019
Accepted March 27, 2019 ■