Correct responses alleviate the negative evaluation of conflict

Ivan I. Ivanchei, Senne Braem, Luc Vermeulen, Wim Notebaert

Department of experimental psychology, Ghent University

Author Note

Correspondence. Ivan Ivanchei, Department of Experimental Psychology, Henri Dunantlaan, 2, 9000, Ghent, Belgium. E-mail: ivan.ivanchei@ugent.be

Data availability. Preregistration protocols, data and analysis scripts for the reported studies can be found on the Open Science Framework: <https://osf.io/5rdxn/files>.

Acknowledgments. We are grateful to Tom Damen for sharing the details of his studies, which helped us to run a replication. Ivan Ivanchei, Senne Braem, and Wim Notebaert were supported by FWO – Research Foundation Flanders (G.0660.17N), Senne Braem by a European Research Council Starting Grant (852570), and Luc Vermeulen by FWO – Research Foundation Flanders (11H5619N).

Conflicts of interest. The authors have no competing interests to declare.

Overall word count: 8344.

26/11/2020. This paper has not been peer reviewed. Please do not copy or cite without author's permission.

Abstract

Recent studies have demonstrated that cognitive conflict, as experienced during incongruent Stroop-trials, is automatically evaluated as negative in line with theories emphasizing the aversive nature of conflict. However, while this is well-replicated when people only see the conflict stimuli, results are mixed when participants also respond to stimuli before evaluating them. Potentially, the positive surprise people feel when overcoming the conflict, allows them to evaluate the experience as more positive. In this study, we investigated whether task experience can account for contradictory findings in the literature. Across three experiments, we observed that responding to incongruent stimuli was evaluated as negative on the first trials, but this effect disappeared after 32 trials. This contrasted with the results of a fourth experiment showing that the negative evaluation of incongruent trials did not disappear, when participants could not respond to the conflict. A reanalysis of three older experiments corroborated these results by showing that a positive evaluation of conflict only occurred after participants had some experience with the task. These results show that responding to conflict clearly changes its affective evaluation fitting with the idea that creating outcome expectancies (lower expectancies for being correct on incongruent trials) makes the experience of conflict less negative.

Keywords: conflict; reward prediction error; cognitive control; affect; conflict aversiveness.

CORRECT RESPONSES & CONFLICT EVALUATION

Correct responses alleviate the negative evaluation of conflict

In his 2007 update of the conflict monitoring theory, Botvinick proposed that Anterior cingulate cortex (ACC) may not only monitor conflict but serve a broader function of monitoring aversive events that drive avoidance learning (Botvinick, 2007). This idea was later supported by empirical data and it is now well-established that cognitive conflict (e.g. incongruent Stroop stimuli, task-switching, etc) trigger negative affect (Braem et al., 2017; Dignath & Eder, 2015; Dreisbach & Fischer, 2012; Fritz & Dreisbach, 2013, 2015; Schouppe et al., 2012, 2015; Vermeulen, Braem, et al., 2019; Vermeulen, Wisniewski, et al., 2019). For example, using the affective priming technique (Fazio, 2001), Dreisbach and Fischer (Dreisbach & Fischer, 2012) showed that participants process negative pictures faster than positive pictures, when presented immediately after incongruent Stroop stimuli (e.g. “RED” printed in blue) versus congruent Stroop stimuli (“RED” printed in red). This suggests that incongruent stimuli trigger more negative affect than congruent ones. Similarly, using an affect misattribution procedure (Payne et al., 2005), Fritz and Dreisbach (2013, 2015) also showed that neutral pictures following incongruent Stroop stimuli are evaluated as more negative relative to neutral pictures following congruent stimuli. These and other results referenced above were integrated by recent theories suggesting cognitive conflict is registered as an aversive event to drive cognitive control (Botvinick, 2007; Dignath et al., 2020; Dreisbach & Fischer, 2012, 2015; Inzlicht et al., 2015; van Steenbergen, 2015).

However, recent studies suggested that participants, whenever they can resolve the cognitive conflict (i.e., providing accurate responses to incongruent stimuli), might actually experience these events as more positive (Ivanchei et al., 2019; Schouppe et al., 2015). This intrinsically rewarding nature of conflict resolution is also predicted by reinforcement learning theories of conflict processing, which suggest that people experience a more positive performance prediction error following a successful response to an incongruent versus

congruent trial (Alexander & Brown, 2011; Braem et al., 2012, 2015; Chetverikov & Kristjánsson, 2016, Silvetti et al., 2011, 2018).

In contrast, some recent studies did not show conflict positivity after overt responses to conflict (Damen et al., 2018; Goller et al., 2019; Pan et al., 2020). Damen et al. (2018) found conflict negativity in both conditions: with and without overt responses to Stroop stimuli. Similarly, Goller et al. (2019) also found conflict negativity using overt responses, but not when the proportion of congruent trials was low (25%). Another study measured affect with facial electromyography and found increased activation of the corrugator (“frowning”) muscles in incongruent trials after correct responses (Berger et al., 2019). However, the effect was there only if trial-by-trial accuracy feedback was presented. Finally, another recent study (Pan et al., 2020) found no clear affective association after overt responses.

In the present study, we want to investigate this effect in more detail and try to resolve these apparent contradictions in previous studies. Similarly, in their recent review, Dignath et al. (2020) described the positive affective association of incongruent trials (in short: the conflict positivity effect) as one of the outstanding questions in the study of affective regulation of cognitive control. To approach this question, we focus on one important experimental detail, namely the amount of participants’ experience with incongruent trials (i.e., the number of trials). Notably, the experiments of Damen et al. (2018) and Goller et al. (2019), who showed a negative evaluation of incongruent trials after overt responses, were quite short (32 – 48 trials and 128 – 144 trials, respectively). In contrast, the studies by Schouppe et al. (2015), who showed conflict positivity, and Ivanchei et al. (2019), were longer (720 and 600 trials, respectively). Berger et al. (2019) present three experiments: a short experiment (192 trials) showed conflict negativity; two longer ones (768 trials) showed conflict negativity and no effect respectively. The study by Pan et al., who found neither a

CORRECT RESPONSES & CONFLICT EVALUATION

positive nor negative evaluation, used 440 trials. Therefore, we reasoned that participants probably need experience with the task before the negative affect associated with conflict trials can turn into positive affect. This idea would fit with the theories of ACC function that interpret conflict processing in terms of unexpected (performance) outcomes (Alexander & Brown, 2011; Silvetti et al., 2011, 2014, 2018). Namely, if the positive evaluation of conflict reflects a positive performance prediction error (Braem et al., 2015), one can expect this effect to occur only after outcome expectancies for congruent and incongruent trials start to diverge (i.e., when participants had first experience with responding to these trials). Until then, the (potentially more dominant) negative evaluation of experiencing a cognitive conflict could explain the results.

With this hypothesis in mind, we re-analyzed previous experiments where the conflict positivity effect was observed (Ivanchei et al., 2019; Schouppe et al., 2015). When we analyzed only the first 90-100 trials of these studies, we also failed to observe conflict positivity: $F(1, 24) = 0.05, p = 0.829$ (Ivanchei et al., 2019); $F(1, 17) = 0.23, p = 0.635$ (Schouppe et al., 2015, Experiment 2A). Instead, the effect becomes apparent only later in the experiment (see Figure 1).

CORRECT RESPONSES & CONFLICT EVALUATION

Figure 1. First three blocks of previous studies showing the conflict positivity effect. Upper panel conflict positivity, lower panel – positive priming for incongruent (IC) and congruent (C) stimuli separately.

This re-analysis offers first evidence in favor of the idea that the positive evaluation of cognitive conflict only develops later in the experiment. If this idea holds, we should also be able to see whether the negative evaluation of cognitive conflict disappears, when replicating the design of studies that showed those (e.g., Damen et al., 2018), but simply making these experiments longer. To this end, we focused on the study that first showed a negative evaluation of responding to conflict by Damen et al. (2018). In Experiment 1, we used the exact same procedure as Damen et al. (2018) but made the experiment six times longer (188 trials). In Experiment 2, we again used the same affective evaluation procedure, but preceded this task with another 188 trials of the Stroop task without an affective evaluation. Experiment 3 replicated Experiment 1 but was yet longer (288 trials). Finally, in Experiment 4 we used the same procedure as Experiment 3 but did not allow participants to respond to the Stroop stimuli. All four experiments were pre-registered.

Experiment 1

In Experiment 1, we replicated the procedure of Experiment 3B from Damen et al. (2012) as close as possible. We introduced two minor changes. First, we used a balanced stimulus set with the same amount of unique congruent and incongruent stimuli, to ensure the effect could not be attributed to a mere exposure effect (Bornstein, 1989). Second, our experiment was six times longer. We expected to observe a negative affect by incongruency in block one (that replicates Damen et al, 2018), and a reversed effect in the rest of the blocks. The pre-registration protocol can be found on OSF: <https://osf.io/u38vp>.

Method

Participants. We determined the sample size using sequential Bayesian hypothesis testing (Schönbrodt et al., 2017) by increasing sample size until a decisive Bayes Factor (BF) larger than 6 (or 1/6) was obtained for the paired t-test for the effect of congruency on target evaluation either in block 1 (H1: conflict is negative) or in aggregated blocks 2-6 (H1: conflict is positive). We started with 30 participants and calculated the BF after every new 10 participants. For financial reasons, the upper limit was set to 80 participants (eligible for analysis). Participants were recruited (and the data was collected) online using Prolific Academic. Ninety participants (46 women, min age = 18, max = 35, $M = 27.1$, $SD = 5.1$) took part in the experiment in exchange for a small fee (£ 1.5). The experiment adhered to the general ethical protocol of the ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University.

Stimuli. We used two sets of stimuli for our study. Stroop stimuli served as primes. In the Stroop task, participants are presented with the color words 'GREEN', 'BLUE', 'RED', and 'YELLOW' printed in green, blue, red or yellow color. In a four-color Stroop task, there are four possible congruent stimuli (color words printed in the same color, e.g. YELLOW printed in yellow) and 12 incongruent stimuli (e.g. YELLOW printed in blue). For each

CORRECT RESPONSES & CONFLICT EVALUATION

participant, we randomly selected four incongruent primes. However, the incongruent primes were selected in a way that participants were presented with the same number of different words and the same number of different ink colors. This way, together with the congruent stimuli, there were eight unique primes that were each presented equally often, ensuring our results could not be attributed to a mere exposure effect. Thirty-two pictures of different polygons served as targets. To increase variability in target identity given a limited target set, starting from trial 33, a target picture could be presented upside down with a probability of 0.5. Thus, the first 32 targets were presented the same way as in Damen et al. (2018).

Procedure. At the beginning of the experiment, participants were explained that they would be asked to indicate the print color of color-words. After that, they had a four-trial training, where they were provided with feedback. The response deadline was 4000 ms. After training, the main experiment started. Participants had six blocks of 32 trials. In each trial, participants were presented with a congruent or incongruent prime until response, followed by a target presented for 1000 ms. After the target disappeared, participants had to evaluate how much they liked the target on a 9-point scale. If participants did not provide a response to the prime within 2000 ms, the trial was aborted, and the target was not presented. No additional feedback was provided. The target evaluation time was unlimited. However, participants were encouraged to provide a quick and intuitive evaluation of targets. Participants had to respond with a mouse. All response options were presented as buttons on the screen (see Figure 2).

In each block, participants were presented four times with each prime and once with each target. If a target was paired with a congruent prime in block N, this target was paired with an incongruent prime in block N+1 and vice versa. Participants had short breaks between blocks. The experiment took around 20 minutes.

Figure 2. The sequence of events in a trial in Experiments 1 to 3.

Results

Statistical analyses. For the analysis of Stroop task performance, we applied a repeated-measures ANOVA on aggregated response times and error rates for responses within response deadline (2000 ms). For response times analyses, only correct responses were considered. For the analysis of the affective response evaluation, we applied a repeated-measures ANOVA on aggregated target liking ratings following responses in the Stroop task. Only correct responses were considered. The Greenhouse-Geisser correction was applied when appropriate. We also calculated Bayes factors for the affective responses. As a prior, we used half-normal distribution with $M = 0$ and $SD = 0.22$, which was the raw effect reported in a similar experiment by Damen et al. (2018, Experiment 3B). Bayes factors were calculated using an R function (see analysis files on OSF) based on the calculator provided by Zoltan Dienes at http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm.

Five participants who had more than 30% of missed or erroneous responses in a Stroop task were removed from the analysis. Five more participants were removed because they used only one target evaluation option in more than 90% of trials.

The same analysis approach was applied to all experiments reported in this paper.

Stroop task performance. Participants responded faster on congruent Stroop trials ($M = 1102$ ms, $SD = 137$) compared to incongruent ($M = 1238$ ms, $SD = 142$), $F(1, 79) = 302.64$, $p < 0.001$, $\eta_p^2 = 0.79$, and made less errors on congruent trials (2.8%, $SD = 0.0$) compared to incongruent ($M = 8.2\%$, $SD = 0.1$), $F(1, 79) = 70.42$, $p < 0.001$, $\eta_p^2 = 0.47$, indicating a significant Stroop effect.

Affective response to incongruency in block 1. In block 1, participants provided higher liking ratings for targets following congruent primes ($M = 4.74$, $SD = 1.20$) compared to incongruent ones ($M = 4.59$, $SD = 1.08$), $F(1, 79) = 4.18$, $p = 0.044$, $\eta_p^2 = 0.05$, $BF10 = 3.62$, indicating substantial evidence for the alternative hypothesis (Jeffreys, 1961), i.e. conflict negativity.

Affective response to incongruency in blocks 2-6. Collapsed data from blocks 2-6 did not provide any evidence for the difference between liking ratings for targets following congruent primes ($M = 4.49$, $SD = 1.16$) compared to incongruent ones ($M = 4.52$, $SD = 1.16$), $F(1, 79) = 1.01$, $p = 0.318$, $\eta_p^2 = 0.01$, $BF10 = 0.34$, indicating uncertainty about the alternative hypothesis, i.e. conflict positivity.

A model with prime congruency and block indicated a significant effect of block, $F(5, 395) = 2.67$, $p = 0.035$, $\eta_p^2 = 0.08$, a non-significant effect of prime congruency, $F(1, 79) = 0.01$, $p = 0.903$, $\eta_p^2 = 0.00$, and a marginally significant effect of block $\times$ prime congruency interaction, $F(5, 395) = 2.54$, $p = 0.057$, $\eta_p^2 = 0.06$.

CORRECT RESPONSES & CONFLICT EVALUATION

Figure 3. Upper two rows (plots A-D): conflict positivity (the difference between ratings for targets following incongruent (IC) and congruent (C) primes) by block in four experiments. Lower row (E-F): ratings for targets following IC and C primes; E – combined data from experiments 1-3; F – Experiment 4. Error bars represent 95% within-subject confidence intervals.

Discussion

The results of Experiment 1 replicated the findings by Damen et al. (2018, Experiment 3B) by showing that incongruent stimuli facilitated the negative evaluation of neutral stimuli in the first 32 trials (i.e., block 1). Importantly, however, this effect was no longer significant later in the experiment. In other words, this result supported our hypothesis in showing that the negative evaluation following correct responses disappeared over time, possibly because practice with the task allowed the participants to learn they can also resolve the conflict. However, although the effect disappeared over time, it never reversed, as in the studies of Schouppe et al. (2015) and Ivanchei et al. (2019) where correct responses to incongruent stimuli were found to induce positive affect. Schouppe et al. (2015) and Ivanchei et al. (2019) also used additional practice with the Stroop task prior to the “Stroop + target

evaluation” trials (i.e., 24 trials in Schouppe et al., 2015; and until criterion in Ivanchei et al., 2019). Therefore, observing a more positive evaluation following correct responses to incongruent trials (relative to congruent trials) may necessitate more experience with the conflict task. In Experiment 2, participants had extended Stroop task training (six blocks of 32 trials), before the “Stroop + target evaluation” blocks.

Experiment 2

The pre-registration protocol for Experiment 2 can be found on OSF <https://osf.io/uvqds>.

Method

Participants. We determined the sample size using sequential Bayesian hypothesis testing by increasing sample size until a decisive Bayes Factor (BF) larger than 6 (or 1/6) was obtained for the effect of congruency on target evaluation (conflict positivity) in main experimental blocks (aggregated). For financial reasons, the upper limit was set to 80 participants (eligible for analysis)¹. Eighty-four participants (45 women, min age = 18, max = 35, $M = 27.1$, $SD = 5.1$) took part in the experiment in exchange for a small fee (£ 2). The experiment adhered to the general ethical protocol of the ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University.

Stimuli. Stimuli were the same as in Experiment 1.

Procedure. In the first part of the experiment, participants performed six blocks of 32 trials of a typical Stroop task (without the presentation of a neutral stimulus that participants had to evaluate). The response deadline was 2000 ms. Corresponding feedback was provided

---

¹ The BF for the conflict positivity crossed the 1/6 threshold when we collected 50 participants ($BF_{10}=0.11$), but we decided not to stop and go up to the upper bound of 80 participants. It does not increase the false discovery rate within the Bayesian framework, i.e. we just gain more confidence in our conclusions with more data.

after errors and responses exceeding the response deadline. The second part of the experiment was exactly the same as the procedure of Experiment 1.

Results

Three participants had more than 30% of missed or erroneous responses in the Stroop task in the second part of the experiment and were removed from the analysis. The remaining participants performed better than 70%, in both the second and first part of the experiment.

One more participant was removed because they used only one target evaluation option in more than 90% of trials. The same analysis techniques were applied in Experiment 2.

Stroop task performance in the first part of the experiment. In the first part of the experiment (Stroop only), participants responded faster on congruent trials ($M = 977$ ms, $SD = 143$) compared to incongruent ($M = 1118$ ms, $SD = 154$), $F(1, 79) = 354.91$, $p < 0.001$, $\eta_p^2 = 0.82$, and made less errors on congruent trials ($M = 1.1\%$, $SD = 0.0$) compared to incongruent trials ($M = 5.1\%$, $SD = 0.0$), $F(1, 79) = 65.27$, $p < 0.001$, $\eta_p^2 = 0.45$, indicating significant Stroop effect.

Stroop task performance in the second part of the experiment. Similarly, when the Stroop task was combined with target evaluation trials, participants still responded faster on congruent Stroop trials ($M = 1099$ ms, $SD = 160$) compared to incongruent ($M = 1211$ ms, $SD = 157$), $F(1, 79) = 222.00$, $p < 0.001$, $\eta_p^2 = 0.74$, and made less errors on congruent trials ($M = 4.1\%$, $SD = 0.0$) compared to incongruent ($M = 10.8\%$, $SD = 0.1$), $F(1, 79) = 80.49$, $p < 0.001$, $\eta_p^2 = 0.50$.

Affective responses. We hypothesized that the experience with the Stroop task would trigger positive affect in incongruent trials. However, liking ratings for targets following congruent trials ($M = 4.30$, $SD = 1.05$) and incongruent trials ($M = 4.28$, $SD = 1.07$) did not differ significantly, $F(1, 79) = 0.47$, $p = 0.494$, $\eta_p^2 = 0.01$, $BF10 = 0.08$, indicating strong evidence for the null hypothesis.

CORRECT RESPONSES & CONFLICT EVALUATION

The model including block indicated no significant effects of block, $F(5, 395) = 0.71$, $p = 0.520$, $\eta_p^2 = 0.01$, prime congruency, $F(1, 79) = 0.54$, $p = 0.464$, $\eta_p^2 = 0.01$, or the block $\times$ prime congruency interaction, $F(5, 395) = 1.65$, $p < 0.174$, $\eta_p^2 = 0.02$.

Although we did not pre-register this analysis for Experiment 2, we decided to test conflict negativity in the first block after the Stroop-training. Repeated-measures ANOVA exclusively for block 1 indicated marginally significant conflict negativity with higher ratings for targets following congruent trials ($M = 4.45$, $SD = 1.15$) compared to incongruent trials ($M = 4.27$, $SD = 1.00$), $F(1, 79) = 3.89$, $p = 0.052$, $\eta_p^2 = 0.05$, $BF_{10} = 3.89$, indicating substantial evidence for the alternative hypothesis (conflict negativity). There was no significant effect of congruency in blocks 2-6), $F(1, 79) = 0.08$, $p = 0.774$, $\eta_p^2 < 0.01$. $BF_{10} = 0.15$ indicated strong evidence for the null hypothesis. Both results closely replicate the findings from Experiment 1.

Discussion

We did not observe a significant difference in target evaluation following correct responses to congruent and incongruent primes. Interestingly, however, the pattern of results was similar to the results of Experiment 1 in that congruent primes did trigger more positive affect compared to incongruent primes in the first block. This suggests that practice with the Stroop task did not modulate the affective responses to conflict, while practice with the combined Stroop + target evaluation trials does seem to do so. A speculative, post-hoc interpretation of this finding can be that people only start affectively evaluating incongruent stimuli when brought into a context that promotes the affective evaluation of Stroop stimuli (i.e., the affect misattribution procedure). Similarly, only after they start affectively evaluating Stroop stimuli, they can also consider that resolving conflict can cancel this negative evaluation. Another possible interpretation is that participants perceived the combined Stroop and affect misattribution blocks as a completely new task. This explanation

can be supported by the dramatic change of Stroop task performance in the blocks with affect misattribution compared to the first training blocks (see [Appendix A](#)), indicating a substantial redistribution of cognitive resources starting from the first combined Stroop-evaluation block. Both interpretations suggest that we should extend the training on both the Stroop and target evaluation trials. This is what we did in Experiment 3.

Experiment 3

The pre-registration protocol for Experiment 3 can be found on the OSF page: <https://osf.io/vta94>.

Method

Participants. We determined the sample size using sequential Bayesian hypothesis testing by increasing sample size until a decisive Bayes Factor (BF) larger than 6 (or 1/6) was obtained for the effect of congruency on target evaluation either in block 1 (H1: conflict is negative) or in aggregated blocks 2-9 (H1: conflict is positive). For financial reasons, the upper limit was set to 80 participants (eligible for analysis)². Ninety-five participants (58 women, min age = 18, max = 35, $M = 26.9$, $SD = 4.7$) took part in the experiment in exchange for a small fee (£ 2). The experiment adhered to the general ethical protocol of the ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University.

Stimuli and procedure. Stimuli and procedure were the same as in Experiment 1. The only difference was that there were nine blocks of 32 trials instead of six blocks.

---

² As in Experiment 2, The *BF* for the conflict positivity crossed the 1/6 threshold quite early – when we collected 30 participants ($BF_{10}=0.06$). Again, we decided not to stop and go up to the upper bound of 80 participants.

Results

Six participants who had more than 30% of missed or erroneous responses in a Stroop task were removed from the analysis. Nine more participants were removed because they used only one target evaluation option in more than 90% of trials.

Stroop task performance. Participants responded faster on congruent Stroop trials ($M = 1100$ ms, $SD = 155$) compared to incongruent ($M = 1230$ ms, $SD = 157$), $F(1, 79) = 289.64$, $p < 0.001$, $\eta_p^2 = 0.79$, and made less errors on congruent trials ($M = 4.0\%$, $SD = 0.0$) compared to incongruent ($M = 10.4\%$, $SD = 0.1$), $F(1, 79) = 72.04$, $p < 0.001$, $\eta_p^2 = 0.48$.

Affective responses. In block 1, no significant difference was observed between liking ratings for targets following congruent primes ($M = 4.22$, $SD = 1.22$) compared to incongruent ones ($M = 4.19$, $SD = 1.20$), $F(1, 79) = 0.18$, $p = 0.670$, $\eta_p^2 = 0.00$, $BF_{10} = 0.38$ indicating uncertainty about the hypotheses. An overall model including block indicated no significant effects either: effect of block, $F(8, 624) = 0.49$, $p = 0.703$, $\eta_p^2 = 0.01$, effect of prime congruency, $F(1, 78) = 1.92$, $p = 0.170$, $\eta_p^2 = 0.02$, and effect of block $\times$ prime congruency interaction, $F(8, 624) = 0.33$, $p < 0.832$, $\eta_p^2 = 0.00$. BF_{10} for the conflict positivity in blocks 2-9 was 0.04, providing strong evidence for the null hypothesis. Following the pre-registration, we compared conflict positivity in collapsed blocks 1-3 vs. 4-6 and 1-3 vs. 4-9. However, we did not find any significant effects in these comparisons.

Discussion

In Experiment 3, we observed neither a negative nor positive evaluation of incongruent trials relative to congruent trials despite the longer experimental procedure. To further explore the effect of conflict negativity, we ran additional analysis combining data from all three experiments ($N = 240$). We only used the second part (Stroop task + target evaluation) from Experiment 2, and the first six blocks of Experiment 3.

Combined analysis of three experiments

Affective response to incongruency in block 1. In block 1, participants provided higher liking ratings for targets following congruent primes ($M = 4.47$, $SD = 1.20$) compared to incongruent ones ($M = 4.35$, $SD = 1.10$), $F(1, 239) = 7.20$, $p = 0.008$, $\eta_p^2 = 0.03$, $BF_{10} = 12.31$, indicating strong evidence for the alternative hypothesis (conflict negativity). To facilitate comparability with other studies, we computed *Cohen's d*, which was 0.17 [95% CI -0.00, 0.35].

Affective response to incongruency in blocks 2-6. The collapsed data from blocks 2-6 did not provide any evidence for a difference between liking ratings for targets following congruent primes ($M = 4.31$, $SD = 1.15$) compared to incongruent ones ($M = 4.32$, $SD = 1.16$), $F(1, 239) = 0.51$, $p = 0.474$, $\eta_p^2 < 0.01$. $BF_{10} = 0.13$ indicating strong evidence for the null hypothesis, i.e. against conflict positivity. Additionally, we calculated the Bayes factor for the conflict negativity hypothesis, which was $BF_{10} = 0.04$, again indicating strong evidence for the null. *Cohen's d* (for conflict negativity) = -0.05 [95% CI -0.23, 0.13]. Dividing BF_{10} (conflict positivity) by BF_{10} (conflict negativity) gives us $0.13/0.04 = 3.25$, that is, the conflict positivity hypothesis is 3.25 times more likely in blocks 2-6 compared to the conflict negativity hypothesis. However, both hypotheses are much less likely than the null hypothesis given present data.

Affective response to incongruency by block. The overall model with prime congruency and block indicated a significant effect of block, $F(5, 1190) = 2.73$, $p = 0.041$, $\eta_p^2 = 0.01$, a non-significant effect of prime congruency, $F(1, 238) = 0.82$, $p = 0.367$, $\eta_p^2 = 0.00$, and a significant block $\times$ prime congruency interaction, $F(5, 1190) = 3.00$, $p = 0.028$, $\eta_p^2 = 0.01$. To explore this interaction effect in more detail, we ran a mixed-effects linear regression taking into account between-participants and then between blocks (within-

participants) and between prime types (within participants) difference in average ratings³. The linear effect of block had a negative slope indicating gradual decrease in ratings provided by participants after congruent primes, $B = -0.16$ ($SE = 0.04$), $p < 0.001$. The interaction between the linear effect of block and prime congruency was also significant, $B = 0.08$ ($SE = 0.04$), $p = 0.033$, and had a positive coefficient, indicating a smaller decrease (i.e., more positive responses) for incongruent trials compared to congruent ones. Separate t-tests for six blocks (Bonferroni corrected) provide significant result for conflict negativity only in block 1.

Stroop task performance by block. In the full model with prime congruency, block and their interaction, all effects were significant, prime congruency, $F(1, 238) = 786.26$, $p < 0.001$, $\eta_p^2 = 0.77$, block, $F(1, 238) = 173.70$, $p < 0.001$, $\eta_p^2 = 0.42$, prime congruency $\times$ block, prime congruency, $F(1, 238) = 4.61$, $p = 0.033$, $\eta_p^2 = 0.02$, indicating that congruency effect changed with blocks. Linear mixed effects model with the same fixed effects and random intercept for each participant showed that coefficient for the interaction was negative, $B = -3.17$ ($SE = 1.83$), $p = 0.083$, indicating trend for decreasing congruency effect with blocks. The same result was observed for the accuracy data: prime congruency, $F(1, 239) = 215.973$, $p < 0.001$, $\eta_p^2 = 0.47$, block, $F(1, 239) = 33.43$, $p < 0.001$, $\eta_p^2 = 0.12$, prime congruency $\times$ block, prime congruency, $F(1, 239) = 15.04$, $p < 0.001$, $\eta_p^2 = 0.06$, indicating that congruency effect changed with blocks. Linear mixed effects model with the same fixed effects and random intercept for each participant showed that coefficient for the interaction was negative, $B = -0.68$ ($SE = 0.19$), $p < 0.001$, indicating decreasing congruency effect in error rates with blocks.

---

³ The model formula was the following: rating $\sim$ prime congruency * block + (1|participant) + (1|participant : prime congruency) + (1|participant : block) mimicking repeated-measures ANOVA as close as possible.

Correlations between Stroop performance and conflict positivity⁴. Additionally, we also performed several correlation analyses between Stroop task performance and conflict positivity. We used average $(RT_{\text{incongruent}} - RT_{\text{congruent}}) / (RT_{\text{congruent}})$ and the difference of arcsine square root transformed error rates between congruent and incongruent trials. As an individual index of conflict positivity, we used average $\text{Rating}_{\text{incongruent}} - \text{average Rating}_{\text{congruent}}$. Following pre-registration, we looked for the one-sided negative correlation in block 1 and one-sided positive correlation in blocks 2-6.

We analyzed, first, the correlation between Stroop performance and conflict positivity in block 1. Spearman rank correlation (one-tailed) was used because of some outliers in the difference scores we used as variables (> 3 SD from the mean in both Stroop effect and affective measures). The correlation was only marginally significant, $\rho = -0.10, p = 0.068$. The correlation was not significant in error rates, $\rho = -0.01, p = 0.455$. No significant correlation was found in collapsed blocks 2-6: RTs, $\rho = -0.07, p = 0.877$, error rates, $\rho = 0.05, p = 0.211$. Trying to assess the influence of initial Stroop experience on subsequent conflict evaluation, we calculated the correlation between the Stroop effect in block 1 and conflict positivity in blocks 2-6. The correlation was also not significant both for RTs, $\rho = 0.02, p = 0.407$, and error rates, $\rho = -0.01, p = 0.613$.

Discussion

Despite the null effects in Experiment 3, the combined analysis of our three datasets supports our overall conclusion that correctly responding to incongruent stimuli can trigger negative affect, but, importantly, it quickly fades away as demonstrated by significant conflict positivity $\times$ block interaction. This finding both replicates Damen et al. (2018) but also supports our hypothesis that experience with the task is an important factor for the

---

⁴ This analysis was pre-registered for the first experiment, but having three similar experiments we decided to run correlation analysis on the combined dataset for greater reliability of the results.

1 conflict resolution effect. Although we did not observe a complete reversal (e.g., a positive
2 evaluation of accurately responding to conflict), we did observe a relative decrease in conflict
3 negativity.

4 To further test whether the above-observed effect of experience on the evaluation of
5 incongruent versus congruent trials is specific to the fact that participants were allowed to
6 respond (and thus resolve the cognitive conflict), we reasoned that this observation that
7 incongruent trials become less negative over time should not be expected when participants
8 are not allowed to respond. In other words, if one can only reduce the aversive value of
9 cognitive conflict by being allowed to resolve it, we should see this aversive value persist
10 under conditions where (explicit) conflict resolution is not possible. We tested this idea in
11 Experiment 4.

12 Experiment 4 also allowed us to rule out another interpretation of the findings of
13 Experiments 1-3. Namely, one may argue that weak and disappearing negative affect by
14 incongruency in a conflict task with overt responses could also be due to longer prime-target
15 stimulus onset asynchrony (SOA) and the difference in response times in congruent and
16 incongruent trials (e.g., Fritz & Dreisbach, 2015). To address this, we ensured that SOAs
17 were similar to the response times in Experiment 3, as well as the length of that experiment (9
18 blocks of 32 trials).

19 Experiment 4

20 The pre-registration protocol for Experiment 4 can be found on the OSF page:

21 <https://osf.io/ma3pc>.

22 Method

23 **Participants.** We determined the sample size using sequential Bayesian hypothesis
24 testing by increasing sample size until a decisive Bayes Factor (BF) larger than 6 (or 1/6) was
25 obtained for the effect of congruency on target evaluation (conflict positivity) in main

CORRECT RESPONSES & CONFLICT EVALUATION

experimental blocks (aggregated). For financial reasons, the upper limit was set to 80 participants (eligible for analysis). Eighty-eight participants (47 women, min age = 18, max = 35, $M = 26.8$, $SD = 5.3$) took part in the experiment in exchange for a small fee (£ 2). The experiment adhered to the general ethical protocol of the ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University.

Stimuli. The main stimulus set was the same as in Experiment 3. We also added four stimuli where RED, GREEN, BLUE, and YELLOW was written in purple ink and four stimuli where PURPLE was written in red, green, blue, and yellow ink.

Procedure. Before the main part of the experiment, participants performed 30 training trials with the Stroop task (different from the main experiment, all possible incongruent trials were allowed in this short practice phase). During training, the response deadline was set to 4000 ms. Corresponding feedback was provided after correct responses, errors, and responses exceeding the response deadline.

After this practice phase, participants were told that from now on they would not have to respond to color-word stimuli, but they should evaluate the stimuli presented right after. They were told, however, that they should still pay attention to the color-words because if they see a purple print color or the word “PURPLE” they should not evaluate the target and should press the space key instead. These catch trials were introduced to ensure that participants would still process both the color and the meaning of the primes (Fritz & Dreisbach, 2013). If participants missed a catch trial by providing a rating for a target (catch trial miss), they received a message “Error. Missed purple color! Press SPACE when you see purple”. If they pressed the space key when no purple color was presented (catch trial false alarm), they received a message “Error. No purple color”. Catch trial feedback messages were presented for 3500 ms. In every block, one congruent and one incongruent trial were replaced by a catch trial.

CORRECT RESPONSES & CONFLICT EVALUATION

The sequence of events is presented in Figure 4. The main difference from Experiment 3 was that participants did not respond to the prime stimulus anymore. To equate the different prime presentation durations with our previous experiments, we presented primes for varying durations. These durations were taken from the actual response times participants exhibited in Experiment 3. That is, for every trial, we randomly chose a response time from Experiment 3 ($M = 1158$ ms, $SD = 292$, Median = 1123, 1st quartile = 940, 3rd quartile = 1347). Moreover, we also took into account prime congruency (RTs are usually longer for incongruent trials) and block (participants respond faster with practice). This way we mimicked SOA from the task when people are allowed to respond to the primes. The rest of the procedure details were the same as in Experiment 3. There were nine blocks of 32 trials each.

Figure 4. The sequence of events in a trial in Experiment 4.

Results

One participant who had more than 30% of errors in a 30-trials Stroop training was removed from the analysis. Six more participants were removed because they used only one target evaluation option in more than 90% of trials. The average error rate in catch trials was 1.3%, $SD = 1.4$, suggesting that participants encoded both meaning and print color of the primes.

Affective response to incongruency in block 1. In block 1, no significant difference was observed between liking ratings for targets following congruent primes ($M = 4.30$, $SD = 1.32$) compared to incongruent ones ($M = 4.23$, $SD = 1.20$), $F(1, 80) = 0.85$, $p = 0.361$, $\eta_p^2 = 0.01$ (see Figure 3, plots D, F), $BF_{10} = 0.78$ indicating uncertainty about the hypotheses.

Affective response to incongruency in blocks 1-9. The overall model including block indicated a marginally significant effect of block, $F(8, 640) = 1.98$, $p = 0.089$, $\eta_p^2 = 0.02$, a significant effect of prime congruency, $F(1, 80) = 4.30$, $p = 0.041$, $\eta_p^2 = 0.05$, and no significant interaction between block and prime congruency, $F(8, 640) = 1.79$, $p = 0.114$, $\eta_p^2 = 0.02$. BF_{10} for the effect of prime congruency in blocks 1-9 was 4.87, indicating substantial evidence for the alternative hypothesis. A linear mixed-effects model taking into account between-participants and then between blocks (within-participants) and between prime types (within participants) difference in average ratings provided the same result except for significant linear effect of block, $B = -0.14$ ($SE = 0.07$), $p = 0.037$. This model also showed a significant interaction effect between prime type and quadratic effect of block, indicating smaller conflict negativity in the beginning and in the end of the experiment⁵, $B = 0.15$ ($SE = 0.06$), $p = 0.017$.

We addressed the hypothesized evolution of conflict negativity by contrasting collapsed data from blocks 1-3, 4-6, and 7-9. The model indicated a significant effect of prime congruency, $F(1, 80) = 4.31$, $p = 0.041$, $\eta_p^2 = 0.05$, a non-significant effect of the part of the experiment, $F(1, 160) = 2.46$, $p = 0.101$, $\eta_p^2 = 0.03$, and a significant interaction between those two factors, $F(1, 160) = 3.06$, $p = 0.050$, $\eta_p^2 = 0.04$. A linear mixed-effects model taking into account between-participants and then between blocks (within-participants) and between prime types (within participants) difference in average ratings indicated non-significant conflict negativity in blocks 1-3, $B = -0.17$ ($SE = 0.12$), $p = 0.137$, significantly

---

⁵ This model was not pre-registered.

CORRECT RESPONSES & CONFLICT EVALUATION

stronger conflict negativity in blocks 4-6, $B = -0.14$ ($SE = 0.05$), $p = 0.008$, and no significant difference between blocks 1-3 and blocks 7-9, $B = -0.04$ ($SE = 0.05$), $p = 0.480$, supporting the conclusion on the non-linear dynamics of conflict negativity.

Affective response to incongruency with and without responses (Experiments 1-3 vs. Experiment 4). To formally compare conflict negativity with and without responses, we ran mixed ANOVA for all experiments with factors: prime congruency (C vs IC), block (1-6), and response (Experiments 1-3 vs. Experiment 4). We used only six first blocks from experiments 3 and 4 to be able to run a single analysis with the data from Experiments 1 and 2. The modeling output is presented in Table 1. Most importantly, the model showed significant effect of response $\times$ prime congruency interaction, indicating difference in conflict negativity in Experiment 4 compared to Experiments 1-3 and also response $\times$ prime congruency $\times$ block interaction, indicating difference in the evolution of the conflict negativity in Experiment 4.

Table 1

Mixed-ANOVA results for prime congruency, block, and response.

	<i>df 1</i>	<i>df 2</i>	<i>F</i>	<i>p-value</i>	η_p^2
prime congruency	1	318	5.87	0.016	0.02
block	5	1590	5.07	0.001	0.02
response	1	318	1.67	0.197	0.01
prime congruency $\times$ block	5	1590	1.00	0.401	0.00
prime congruency $\times$ response	1	318	11.82	0.000	0.04
block $\times$ response	5	1590	0.50	0.697	0.00
prime congruency $\times$ block $\times$ response	5	1590	4.43	0.003	0.01

Discussion

In Experiment 4, we used a design that did not allow overt responses to the primes. Interestingly, the observed pattern of results was clearly different from what we found in Experiments 1-3. Specifically, participants evaluated neutral targets following congruent primes as more pleasant compared to targets following incongruent primes, but this effect did not fade away throughout the nine blocks. If anything, the negative evaluation of incongruent trials became stronger in the middle of the experiment, suggesting that the responses in Experiment 1-3 (and not SOAs) were crucial in bringing about the decreasing negative evaluation of incongruent trials. It is important to note that the conflict negativity effect was not significant at the beginning of the experiment. However, $BF_{10} = 0.78$ for the effect in block 1 indicated uncertainty, that is that data was too noisy to detect the effect. Our dataset with 9 blocks allowed to make reliable inference on the conflict negativity and its dynamics, however, the noisiness of the online data should be taken into account in future replications of the conflict positivity/negativity effects.

Our novel approach – matching SOAs to response times in previous experiments – allowed us to conclude that conflict negativity was not determined by the differences in SOA in congruent and incongruent trials. Also, we observed conflict negativity despite the SOAs being longer than 1000 ms. In Fritz & Dreisbach (Fritz & Dreisbach, 2015), conflict negativity was observed with SOAs of 200 ms and 400 ms, but not when SOA was 800 ms. This way, our results are incompatible with the findings by Fritz and Dreisbach (2015). However, our findings are consistent with those of Damen et al. (2018, Experiment 3A) who found a negative evaluation of incongruent trials even when the SOA was 2000 ms. Given that we replicated the procedure of Damen et al. (2018), we can also attribute our results to the difference in target evaluation procedure. While participants in Fritz & Dreisbach provided a binary evaluation of the targets (positive / negative), our study used a 1-9 liking

scale, which is potentially more under conscious control, thus it is not affected by automatic affective counter-regulation proposed by Fritz & Dreisbach (2015).

General discussion

To test the idea that one needs conflict task experience to feel conflict resolution positivity, we started from one of the recent studies showing conflict negativity, even after a successful response, in remarkably short experiments (Damen et al., 2018). Specifically, we conducted four experiments trying to understand the influence of practice on the affective evaluation of conflict, by prolonging these experiments. In Experiments 1-3, we observed a small and short-lived negative evaluation of incongruent trials when participants could respond to conflict stimuli. Specifically, we replicated the observation by Damen et al., showing a negative evaluation in the first block of 32 trials, but this effect was weak and disappeared soon, in line with the idea that the positive evaluation of conflict resolution only develops later. This is also consistent with new analyses of the data by Schouppe et al. (2015) and Ivanchei et al. (2019) that we presented in the introduction. To further ensure that the observed pattern was specific to the responses participants were allowed to give (and not just prime duration), we also ran a fourth experiment, using matched prime durations but without a response. In this condition, we observed a clear negative evaluation of incongruent trials supporting the interpretation that the reduced negative affect in Experiment 1-3 was caused by responding to the stimuli. This way we can also rule out other possible explanations related to stimulus material we used, e.g. targets identity, target repetition, or evaluation task.

Despite detecting a decrease in conflict negativity and contrary to our expectations, we did not observe conflict positivity. That is, we did not observe significantly more positive affect triggered by incongruent trials compared to congruent trials. We even found substantial evidence for the null hypothesis – against both conflict positivity and conflict negativity hypotheses after the first block. Possibly, the initial negative experience of the actual conflict

CORRECT RESPONSES & CONFLICT EVALUATION

was never completely counteracted by its resolution. Importantly, there were several differences between our study and previous studies demonstrating also a positive evaluation of incongruent trials (Ivanchei et al., 2019; Schouppe et al., 2015). In our experiments, the average RT for Stroop trials was 1161 ms, whereas in Schouppe et al. (Schouppe et al., 2015), it was 575 ms (Flanker task) and 771 ms (Stroop task). In Ivanchei et al. (Ivanchei et al., 2019), it was 591 ms. So, the prime-target SOA was much longer in the present studies (however, compatibility effects were comparable, see [Appendix B](#)). The longer prime task RTs were due to mouse responses in our experiment, which we used because we tried to keep the experimental conditions as close as possible to Damen et al. (2018).

Another procedural difference between current studies and previous studies showing conflict resolution is the measure of affect. We used the affect misattribution procedure, presenting neutral targets after primes, while Schouppe et al. (2015) and Ivanchei et al. (2019) used an affective priming procedure. In the affective priming procedure, participants are presented with positively or negatively valenced targets (words or pictures) after the primes, that should be categorized as positive or negative as fast as possible. There, the positive evaluation of incongruent trials was demonstrated by a relatively faster categorization of positive targets following accurate responses to incongruent primes compared to congruent primes (the opposite pattern was observed for negative targets). Some studies suggest that affect misattribution procedure may be more prone to demand characteristics because it strongly depends on the prime-target relationship awareness (Cummins et al., 2019). Additionally, it may be easier for participants to come up with the hypothesis linking incongruency to negative affect rather than the “conflict resolution” hypothesis. This might have amplified the overall negative evaluation of incongruent trials in the affect misattribution procedure.

Finally, original studies demonstrating the conflict positivity effect (Schouppe et al., 2015; Ivanchei et al., 2019) were run offline, i.e. in the presence of the experimenter. It could affect participants' achievement motivation, which in turn could increase their task engagement in the task and affective evaluation of task events, e.g. errors, conflicts, and conflict resolutions.

One of the possible alternative accounts to our results may be an overall decrease in liking over time, due to fatigue or boredom, that may mask the difference between C and IC stimuli. Indeed, we can see a significant decrease in ratings over time in Experiment 1 and 2 (see also combined analysis and Figure 3E). The same linear trend was observed in Experiment 4, although it was rather curvilinear – average ratings slightly increased at the end of the experiment (see Figure 3F). On the basis of Experiment 4, we can counter this argument. In Experiment 4, the overall ratings (for congruent and incongruent trials) also decrease, potentially reflecting a time-on-task effect. However, in Experiment 4 (without responses) the difference between congruent and incongruent trials actually increases. This suggests that an overall decrease in liking does not have to result in decreased conflict negativity. We propose that is the initial (positive experience) with incongruent trials that decreases the conflict negativity effect.

Our study suggests that accurate responding to conflict counteracts negative affect triggered by conflict after some experience with a conflict task. While most theories emphasize the supposed aversive nature of cognitive conflict (Botvinick, 2007; Dignath et al., 2020; Dreisbach & Fischer, 2012, 2015; Inzlicht et al., 2015; van Steenbergen, 2015), some have also suggested that the resolution of cognitive conflict could actually bring about a positive experience (Alexander & Brown, 2011; Braem et al., 2012, 2015; Chetverikov & Kristjánsson, 2016; Silvetti et al., 2011, 2018). The present study contributes to the development of these idea, suggesting that one needs some experience with the task to build

expectancies regarding congruent and incongruent trials. Specifically, we suggest that initially, participants evaluate conflict aversively, because incongruent trials are more difficult and lead to errors more often. However, with some practice participants learn that they can deal with incongruency and they are perceived as manageable challenges bringing some reward when successfully met (due to positive prediction errors). Indeed, we found that conflict negativity observed in the first 32 trials disappears later in the experiments. Together with the reanalysis of previous studies this result supports this theoretical suggestion. Our findings also relate to the effort paradox – the fact that we both like and dislike challenging situations despite them being effortful and often contain conflict (Inzlicht et al., 2018). In real-life situations, participants may predict not only the cost of engaging in an effortful situation but also a reward they get if they face this challenge successfully. This way conflict resolution can affect motivation, which may increase the probability of engaging in an effortful task.

Another avenue for exploration is the link between decreasing conflict negativity and control processes. We have shown in the combined analysis of Experiments 1-3 that the congruency effect decreases over time (especially in error rates). Reactive control adjustments can be also considered in this context, e.g. congruency sequence effect (CSE – decreased congruency effect when previous trials were IC). It was shown that the CSE decreases with practice (Mayr & Awh, 2009; van Steenbergen et al., 2015). If, in accordance with the affective signaling hypothesis, the CSE is triggered by negative affect, one would expect diminished CSE when conflict trials don't trigger negative affect anymore. At the same time, the reinforcement learning approach would predict increased CSE with performance-contingent reward (both external and internal), as proposed by Braem et al. (Braem et al., 2012). Although, in our Experiments 1-3, Stroop trials were intervened by evaluation trials, CSE was detectable, $F(1, 237) = 11.2, p = 0.001, \eta_p^2 = 0.05$, and also

CORRECT RESPONSES & CONFLICT EVALUATION

decreased over time, $F(1, 237) = 6.3, p = 0.013, \eta_p^2 = 0.03$. It was significantly different from zero only in the first two blocks. However, not the congruency effect, nor the CSE were significantly correlated with conflict negativity in any block. Therefore, no reliable conclusion can be derived for the conflict negativity – CSE link from our data. However, it can be a promising direction for future research.

In conclusion, our study demonstrates that conflict negativity can be counteracted by accurate responding to conflicts and that this effect requires some experience with the task allowing to build action-outcome dependencies. This result contributes to the resolution of the contradictions in recent literature on the affect-control interaction. It also adds to our understanding of the effects of the reward prediction errors and cognition-emotion interaction in general.

References

- Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, 14(10), 1338–1344. <https://doi.org/10.1038/nn.2921>
- Berger, A., Mitschke, V., Dignath, D., Eder, A. B., & van Steenbergen, H. (2019). *The face of control: Corrugator supercilii tracks aversive conflict signals in the service of adaptive cognitive control*. <https://doi.org/10.31219/osf.io/3qah7>
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychological Bulletin*, 106(2), 265–289. <https://doi.org/10.1037/0033-2909.106.2.265>
- Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, 7(4), 356–366. <https://doi.org/10.3758/CABN.7.4.356>
- Braem, S., Coenen, E., Bombeke, K., van Bochove, M. E., & Notebaert, W. (2015). Open your eyes for prediction errors. *Cognitive, Affective & Behavioral Neuroscience*, 15(2), 374–380. <https://doi.org/10.3758/s13415-014-0333-4>
- Braem, S., King, J. A., Korb, F. M., Krebs, R. M., Notebaert, W., & Egner, T. (2017). The Role of Anterior Cingulate Cortex in the Affective Evaluation of Conflict. *Journal of Cognitive Neuroscience*, 29(1), 137–149. https://doi.org/10.1162/jocn_a_01023
- Braem, S., Verguts, T., Roggeman, C., & Notebaert, W. (2012). Reward modulates adaptations to conflict. *Cognition*, 125(2), 324–332. <https://doi.org/10.1016/j.cognition.2012.07.015>
- Chetverikov, A., & Kristjánsson, Á. (2016). On the joys of perceiving: Affect as feedback for perceptual predictions. *Acta Psychologica*, 169, 1–10. <https://doi.org/10.1016/j.actpsy.2016.05.005>

- 1 Cummins, J., Hussey, I., & Hughes, S. (2019). *The AMPeror's New Clothes: Performance on*
2 *the Affect Misattribution Procedure is Mainly Driven by Awareness of Influence of the*
3 *Primes*. <https://doi.org/10.31234/osf.io/d5zn8>
- 4 Damen, T. G. E., Strick, M., Taris, T. W., & Aarts, H. (2018). When conflict influences
5 liking: The case of the Stroop task. *PLOS ONE*, 13(7), e0199700.
6 <https://doi.org/10.1371/journal.pone.0199700>
- 7 Dignath, D., & Eder, A. B. (2015). Stimulus conflict triggers behavioral avoidance.
8 *Cognitive, Affective, & Behavioral Neuroscience*. [https://doi.org/10.3758/s13415-015-](https://doi.org/10.3758/s13415-015-0355-6)
9 0355-6
- 10 Dignath, D., Eder, A. B., Steinhauser, M., & Kiesel, A. (2020). Conflict monitoring and the
11 affective-signaling hypothesis—An integrative review. *Psychonomic Bulletin &*
12 *Review, (in press)*(1998), 1–68. <https://doi.org/10.3758/s13423-019-01668-9>
- 13 Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *Brain and Cognition*,
14 78(2), 94–98. <https://doi.org/10.1016/j.bandc.2011.12.003>
- 15 Dreisbach, G., & Fischer, R. (2015). Conflicts as Aversive Signals for Control Adaptation.
16 *Current Directions in Psychological Science*, 24(4), 255–260.
17 <https://doi.org/10.1177/0963721415569569>
- 18 Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview.
19 *Cognition and Emotion*, 15(2), 115–141. <https://doi.org/10.1080/0269993004200024>
- 20 Fritz, J., & Dreisbach, G. (2013). Conflicts as aversive signals: Conflict priming increases
21 negative judgments for neutral stimuli. *Cognitive, Affective, & Behavioral*
22 *Neuroscience*, 13(2), 311–317. <https://doi.org/10.3758/s13415-012-0147-1>
- 23 Fritz, J., & Dreisbach, G. (2015). The Time Course of the Aversive Conflict Signal.
24 *Experimental Psychology*, 62(1), 30–39. <https://doi.org/10.1027/1618-3169/a000271>

- 1 Goller, F., Kroiss, A., & Ansorge, U. (2019). Conflict-Elicited Negative Evaluations of
2 Neutral Stimuli: Testing Overt Responses and Stimulus-Frequency Differences as
3 Critical Side Conditions. *Frontiers in Psychology, 10*.
4 <https://doi.org/10.3389/fpsyg.2019.02204>
- 5 Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive
6 control. *Trends in Cognitive Sciences, 19*(3), 126–132.
7 <https://doi.org/10.1016/j.tics.2015.01.004>
- 8 Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly
9 and Valued. *Trends in Cognitive Sciences, 22*(4), 337–349.
10 <https://doi.org/10.1016/j.tics.2018.01.007>
- 11 Ivanchei, I., Begler, A., Iamschinina, P., Filippova, M., Kuvaldina, M., & Chetverikov, A.
12 (2019). A different kind of pain: Affective valence of errors and incongruency.
13 *Cognition and Emotion, 33*(5), 1051–1058.
14 <https://doi.org/10.1080/02699931.2018.1520077>
- 15 Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford University Press.
- 16 Mayr, U., & Awh, E. (2009). The elusive link between conflict and conflict adaptation.
17 *Psychological Research Psychologische Forschung, 73*(6), 794–802.
18 <https://doi.org/10.1007/s00426-008-0191-1>
- 19 Pan, F., Ou, Y., & Zhang, X. (2020). Reward Modulates Affective Priming Effect in
20 Cognitive Conflict Processing: Electrophysiological Evidence. *Frontiers in Human*
21 *Neuroscience, 14*. <https://doi.org/10.3389/fnhum.2020.00059>
- 22 Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes:
23 Affect misattribution as implicit measurement. *Journal of Personality and Social*
24 *Psychology, 89*(3), 277–293. <https://doi.org/10.1037/0022-3514.89.3.277>

- 1 Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential
- 2 hypothesis testing with Bayes factors: Efficiently testing mean differences.
- 3 *Psychological Methods*, 22(2), 322–339. <https://doi.org/10.1037/met0000061>
- 4 Schouppe, N., Braem, S., De Houwer, J., Silvetti, M., Verguts, T., Ridderinkhof, K. R., &
- 5 Notebaert, W. (2015). No pain, no gain: The affective valence of congruency
- 6 conditions changes following a successful response. *Cognitive, Affective, &*
- 7 *Behavioral Neuroscience*, 15(1), 251–261. [https://doi.org/10.3758/s13415-014-0318-](https://doi.org/10.3758/s13415-014-0318-3)
- 8 3
- 9 Schouppe, N., De Houwer, J., Richard Ridderinkhof, K., Notebaert, W., Ridderinkhof, K. R.,
- 10 & Notebaert, W. (2012). Conflict: Run! Reduced Stroop interference with avoidance
- 11 responses. *The Quarterly Journal of Experimental Psychology*, 65(6), 1052–1058.
- 12 <https://doi.org/10.1080/17470218.2012.685080>
- 13 Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management
- 14 to reward-based decision making: Actors and critics in primate medial frontal cortex.
- 15 *Neuroscience & Biobehavioral Reviews*, 46, 44–57.
- 16 <https://doi.org/10.1016/j.neubiorev.2013.11.003>
- 17 Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and Prediction Error in Medial Frontal
- 18 Cortex: Integrating the Single-Unit and Systems Levels of Analysis. *Frontiers in*
- 19 *Human Neuroscience*, 5(August), 75. <https://doi.org/10.3389/fnhum.2011.00075>
- 20 Silvetti, M., Vassena, E., Abrahamse, E., & Verguts, T. (2018). Dorsal anterior cingulate-
- 21 brainstem ensemble as a reinforcement meta-learner. *PLOS Computational Biology*,
- 22 14(8), e1006370. <https://doi.org/10.1371/journal.pcbi.1006370>
- 23 van Steenbergen, H., Haasnoot, E., Bocanegra, B. R., Berretty, E. W., & Hommel, B. (2015).
- 24 Practice explains abolished behavioural adaptation after human dorsal anterior

CORRECT RESPONSES & CONFLICT EVALUATION

1   cingulate cortex lesions. *Scientific Reports*, 5(1), 9721.

2   <https://doi.org/10.1038/srep09721>

3   van Steenbergen, Henk. (2015). Affective Modulation of Cognitive Control: A Biobehavioral
4   Perspective. In *Handbook of Biobehavioral Approaches to Self-Regulation* (pp. 89–
5   107). Springer New York. https://doi.org/10.1007/978-1-4939-1236-0_7

6   Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive
7   control. *Trends in Cognitive Sciences*, 13(6), 252–257.

8   <https://doi.org/10.1016/j.tics.2009.02.007>

9   Vermeulen, L., Braem, S., & Notebaert, W. (2019). The affective twitches of task switches:
10   Task switch cues are evaluated as negative. *Cognition*, 183, 124–130.

11   <https://doi.org/10.1016/j.cognition.2018.11.002>

12   Vermeulen, L., Wisniewski, D., González-García, C., Hoofs, V., Notebaert, W., & Braem, S.
13   (2019). *Shared Neural Representations of Cognitive Conflict and Negative Affect in*
14   *the Dorsal Anterior Cingulate Cortex*. <https://doi.org/10.1101/824839>

1 Appendix A

2
3 *RTs (left panel) and accuracy (right panel) in a Stroop task in Experiment 2. Blocks with*
4 *negative numbers – Stroop task only, blocks 1-6 – Stroop task + target evaluation. Bars*
5 *indicate 95% within-subject confidence intervals.*

6

CORRECT RESPONSES & CONFLICT EVALUATION

Appendix B

Effects sizes with 95% confidence intervals for the compatibility effects in current experiments, Schouppe et al. (2015), and Ivanchei et al. (2019).

	Accuracy			RT		
	C	IC	Cohen's d	C	IC	Cohen's d
Present Experiments 1-3	3.8%	9.9%	0.98 [0.79; 1.17]	1100 ms	1226 ms	1.83 [1.61; 2.04]
Schouppe et al. (2015, Exp. 2A)	1.2%	5.4%	0.99 [0.32; 1.67]	541 ms	643 ms	2.68 [1.80; 3.57]
Schouppe et al. (2015, Exp. 2B)	3.4%	4.1%	0.34 [-0.16; 0.84]	737 ms	789 ms	1.31 [0.76; 1.86]
Ivanchei et al. (2019)	9.7%	12.4%	0.95 [0.40; 1.49]	581 ms	600 ms	1.78 [1.17; 2.39]

In Experiment 2B (Schouppe et al., 2015), we collapsed response-incongruent and stimulus-incongruent trials.