Published in final edited form as:

J Exp Psychol Hum Percept Perform. 2020 November ; 46(11): 1344–1367. doi:10.1037/xhp0000860.

The Unique Effects of Relatively Recent Conflict on Cognitive Control

Jackson S. Colvett, **Lindsay M. Nobles**, **Julie M. Bugg**

Washington University in St Louis

Abstract

In tasks like Stroop, it is well documented that cognitive control is affected by experiences with past conflict on two timescales. The "immediate" timescale is evidenced by congruency sequence effects while the "long" timescale is evidenced by list-wide proportion congruence effects. What remains underspecified is whether relatively recent experiences with conflict (i.e., recent timescale of a few preceding trials) also uniquely affect control and how experiences on different timescales are weighted. We conducted three pre-registered experiments using a novel Stroop paradigm designed to isolate the effects of the recent timescale and measured cognitive control via diagnostic items. In Experiment 1, we manipulated the level of conflict experienced in the recent timescale within mostly congruent and mostly incongruent lists. Controlling for conflict experiences in the long and immediate timescales, we found that conflict in the recent timescale affected cognitive control and did so similarly across list types. In Experiment 2 we found a boundary condition for the effects of recent conflict-- when the recent timescale was preceded by 50% congruent trials, conflict in the recent timescale did not affect cognitive control. Experiment 3 systematically replicated the findings of Experiment 1 and demonstrated that conflict in the recent timescale affected cognitive control even after a long unfilled delay between recent conflict and subsequent diagnostic trials. These novel findings expand understanding of how conflict experiences in the recent timescale affect cognitive control and highlight the need to expand theories of cognitive control to incorporate the recent timescale and its interaction with other timescales.

Keywords

Conflict Accumulation; Timescales; Cognitive Control; List-Wide Proportion Congruence; Congruency Sequence Effects; Diagnostic Items

Cognitive control refers to a set of processes that bias attention toward goal-relevant information and away from goal-irrelevant information. Prior experiences resolving conflicts between competing responses affect cognitive control (e.g., whether a focused scope of attention is engaged whereby processing of goal-relevant information is increased and/or

Correspondence concerning this article should be addressed to Jackson S. Colvett, Department of Psychological and Brain Sciences, Washington University in St. Louis, Campus Box 1125, St. Louis, MO 63130. jcolvett@wustl.edu.

Portions of this data were presented at the 61st annual meeting of the Psychonomic Society in Montreal, Canada.

Pre-registrations and data for Experiments 1 & 2 are available on OSF at https://osf.io/zhpb4/.
Pre-registration and data for Experiment 3 is available on OSF at https://osf.io/mkn5p/.

goal-irrelevant information is decreased, or a relaxed scope of attention is engaged). Consider driving a car on the highway. A primary goal is to maintain an appropriate speed and distance from the cars ahead of you in your lane. Erratic drivers in adjacent lanes might distract you from this goal and elicit conflict that heightens your focus on goal-relevant information, demonstrating the effects of prior conflict on the immediate timescale. On a longer timescale, your focus while driving might be influenced by all prior experiences since getting on the highway. For example, if the highway has been mostly busy with many erratic drivers (or mostly empty), it would likely induce generally focused (or relaxed) attention. But what if you suddenly encounter a handful of erratic drivers in a stretch of highway that had been relatively empty? What effect will this experience on the "recent" timescale have on cognitive control? Will the control system maintain a relaxed scope of attention (consistent with the long timescale) or will the recent conflict lead to a heightening of control? If it does, will the heightening be above and beyond that caused by the last car that was driving erratically beside you (immediate timescale)?

Prior research on cognitive control, including computational models, has focused primarily on the effects of conflict on the immediate and long timescales. One such model is the influential conflict monitoring account that proposed conflict monitoring as a mechanism by which experiences with conflict lead to recruitment of cognitive control (Botvinick, Braver, Brach, Carter, & Cohen, 2001). According to this model, some control adjustments occur in response to conflict experiences[1] on the previous trial. Individuals are less susceptible to conflict after experiencing an incongruent (i.e., conflicting) trial than after experiencing a congruent (i.e., non-conflicting) trial, a pattern referred to as the congruency sequence effect (Gratton, Coles, & Donchin, 1992; for reviews, Egner 2007; Duthoo, Abrahamse, Braem, Boehler, & Notebaert, 2014). While congruency sequence effects can be driven by lower-level processes[2] (Mayr, Awh, & Laurey, 2003; see Weissman, Hawks, & Egner, 2015, for retrieval of previous control parameters), they have also been observed in the absence of known confounds (e.g., Weissman, Jiang, & Egner, 2014) supporting the view that conflict experiences on the immediately preceding trial affect cognitive control. A signature of congruency sequence effects is that they are not sustained across "long" intervals between stimuli on the order of 3000 ms or more (Egner, Ely, & Grinband, 2010; Duthoo, Abrahamse, Braem, & Notebaert 2014). This signature exemplifies the transient nature of control adjustments based on the "immediate" timescale.

In contrast, the list-wide proportion congruence effect provides an example of how control is affected by conflict experiences in the long timescale that accumulate across dozens (e.g., Bugg, Diede, Cohen-Shikora, Selmeczy, 2015) or hundreds of trials (i.e., a block or list; Logan & Zbrodoff, 1979). The list-wide proportion congruence (PC; what percentage of experienced trials are congruent) effect is the pattern whereby congruency effects are

[1]The phrase "conflict experiences" is used here to refer to experiences with *either* conflicting (i.e., incongruent) or non-conflicting (i.e., congruent) trials. This phrase is used rather than simply "conflict", as it is also the case that the absence of conflict is a signal for control adjustments (e.g., Schlaghecken & Martini, 2012).

[2]There are many lower-level processes that affect cognitive control (see Braem et al., 2019 for definitions and discussion). In the current context, this term could refer to episodic retrieval of repeated stimulus and response features, contingency learning (learning of responses that are highly associated with specific stimuli), or item-specific control (modulating attention based on the likelihood that a specific item [stimulus feature] is conflicting). Rather than detail these processes each time, we hereafter use the term "lower-level process(es)" as an umbrella for these processes unless a more specific term is needed.

smaller in mostly incongruent (MI) lists than mostly congruent (MC) lists (see Bugg, 2014; Bugg & Chanani, 2011; Cohen-Shikora, Suh, & Bugg, 2019; Gonthier, Braver, & Bugg, 2016; Hutchison, 2011; Spinelli, Perry, & Lupker, 2019 for evidence of list-wide PC effects when controlling for known confounds; for reviews see Bugg, 2012; Bugg & Crump, 2012). The conflict monitoring account suggests that when higher overall conflict is detected in the list, there is a subsequent increase in cognitive control (Botvinick et al., 2001). In other words, the conflict monitoring model captures adjustments in control based on conflict accumulation over a long timescale and not just the preceding trial (immediate timescale). Evidence showing that list-wide PC effects are observed independent of the congruency sequence effect (Torres-Quesada, Funes, & Lupiáñez, 2013; Torres-Quesada, Milliken, Lupiáñez, & Funes, 2014) suggests that effects of conflict on the long timescale are separable from effects of the immediate timescale. Furthering this view, effects of list-wide PC on control are thought to be sustained rather than transient (De Pisapia & Braver, 2006).

While hundreds of studies have examined effects of the immediate and long timescales, there are several important theoretical gaps in the literature. Two such gaps were of initial interest in the current study. One theoretical gap pertains to the question of whether relatively recent experiences with conflict (i.e., the recent timescale of a few trials before the current trial) influence cognitive control above and beyond the effect of the immediate timescale. That is, do experiences occurring on multiple trials preceding the current trial shape the heightening or relaxation of control beyond the effect of the immediately preceding trial? The conflict monitoring account states that the amount of control on a given trial should be based on “an exponentially weighted average of conflict over multiple preceding trials, rather than only on the immediately preceding trial” (Botvinick et al., 2001; p. 639). Although empirical tests of this notion have almost exclusively focused on the entire long timescale (all preceding trials, as in research on list-wide PC effects), this quote leaves open the possibility that the amount of control may in fact depend on just a few preceding trials.

Only a few prior studies have reported findings that speak to the role of the recent timescale. In a flanker task with nine participants, as the number of preceding compatible trials increased from one to six, reaction time on incompatible trials increased (Durston et al., 2003). However, reaction time on incompatible trials did not significantly decrease as a function of the number of preceding incompatible trials. Other studies with larger samples have found a significant effect of several preceding trials, including multiple incongruent trials. In a Simon task, reaction time declined on trial n as a function of the number of consecutive trials of the same trial type preceding trial n for both congruent and incongruent trials (Horga et al., 2011). In a Stroop task, congruency sequence effects were accentuated by multiple preceding congruent trials and attenuated by multiple preceding incongruent trials (Jiménez & Méndez, 2013, Experiment 1; Jiménez & Méndez, 2014). Quite interestingly, this propagation of the congruency effect across runs occurred independently of whether the list mostly repeated runs of the same trial type or mostly alternated runs (Jiménez & Méndez, 2013). In other words, regardless of whether participants expected the trial type (congruent or incongruent) to repeat (Experiment 2a) or alternate (Experiment 2b), the congruency effect grew larger with repeated congruent trials (though the effect of

multiple incongruent trials was less clear across Experiments 2a/b). Together, these findings suggest that control is adjusted in response to conflict experiences that occur several trials back beyond trial $n - 1$, supporting a role for the recent timescale.

The second theoretical gap concerns how conflict experiences on different timescales interact with each other. Although the aforementioned studies (Durston et al., 2003; Horga et al., 2011; Jiménez & Méndez, 2013; Jiménez & Méndez, 2014) provided evidence for the recent timescale, including its influence in lists where trial types tended to repeat and lists in which they tended to alternate (Jiménez & Méndez, 2013), these studies uniformly used unbiased (i.e., 50% congruent) lists. Therefore, they could not assess whether effects of the recent timescale are affected by the proportion of congruent trials in the long timescale. More specifically, control adjustments based on the recent timescale may depend on whether participants are overall in a relaxed state of control, as in a mostly congruent list, or a heightened state of control as in a mostly incongruent list. Alternatively, adjustments in control based on the recent timescale may be independent of the long timescale (i.e., adjustments may be consistent regardless of whether trials preceding the recent timescale were overall 50% congruent, MC or MI). Either way, understanding the relative contributions of each timescale to cognitive control is important for developing a comprehensive model of control.

Here, too, there is little prior research. Aben and colleagues developed a statistical model that documented the effects of different timescales of conflict accumulation on cognitive control in the flanker task (Aben, Verguts, & Van den Bussche, 2017; see Dey & Bugg, 2020, for replications using Stroop tasks). One key finding from this model was that multiple trials prior to the immediately preceding trial (11 of the preceding 12 trials in a flanker list of 160 trials; Aben et al., 2017; each of the 8 preceding trials within an 8 trial window in color-word Stroop list of 288 trials; Dey & Bugg, 2020) significantly informed the level of cognitive control on trial n controlling for the effect of the other trials. These results lend further support to the view that the recent timescale does play a role in cognitive control adjustments. Most relevant to the second theoretical gap in the literature, another key finding was that there was an interaction such that conflict experiences in the recent timescale (recent trials extending beyond $n - 1$) were weighted less strongly in MI lists than in MC lists. Aben and colleagues interpreted this result to mean that recent experiences with conflict have less of an influence on cognitive control when the long timescale biases individuals to engage proactive control (i.e., sustain a heightened attentional bias across trials; Braver, Gray, & Burgess, 2007) than when the system is relatively relaxed and dealing with conflict via reactive control (i.e., more transiently). However, a limitation of this research is that the statistical models are inherently correlational and there is not yet experimental evidence to support these conclusions.

To take stock, prior research provides suggestive evidence that a recent timescale of conflict, and not just the immediate and long timescales, affects whether the scope of attention on a moment-by-moment basis is relatively focused or relaxed but leaves open several important theoretical questions. In three pre-registered experiments, we aimed to fill existing theoretical gaps and thereby tease apart the contributions of the recent timescale from the other two timescales. Across experiments, we tested three predictions regarding the

effects of conflict in the recent timescale, which we will describe next. Additionally, we implemented two novel design features to address limitations of prior research and these will be described thereafter.

The first prediction is that we would conceptually replicate prior studies by demonstrating an effect of the recent timescale on control (cf. Jiménez & Méndez, 2013; Jiménez & Méndez, 2014). This prediction was tested in Experiments 1, 2, and 3. A second prediction is that there would be an interaction between the recent timescale and the long timescale, as a statistical model found (Aben et al., 2017; Dey & Bugg, 2020). In that model, the recent timescale was weighted more strongly in MC lists than in MI lists. However, this prediction has yet to be tested experimentally. This prediction was tested initially in Experiment 1 and then we conducted an additional test of this prediction in Experiment 3. The third prediction was that control adjustments based on recent conflict may be sustained, and therefore persist over a “long” unfilled delay. This prediction was inspired by our novel observation in Experiment 1 that recent conflict affected performance across several subsequent trials. The rationale behind this prediction is developed more fully immediately prior to Experiment 3.

There were two novel design features that were implemented in all three experiments. These features were critical for testing the above predictions and allowed us to significantly expand upon extant research. The first novel feature was that we used one set of items (stimuli) to induce conflict experiences (including in the recent timescale) and a second set of items to assess the effects of this induction during a subsequent diagnostic phase. The first set of items is hereafter referred to as the Inducer Set, and the second set is referred to as the Diagnostic Set. These sets of items did not share features (i.e., they were different words/colors). This is important because it allowed us to rule out lower-level processes as an alternative explanation to a control-based explanation for any differences we observed in the diagnostic phase (see e.g., Bugg, 2014; Braem et al., 2019). Some of the prior studies reviewed above that examined effects of recent conflict did not include this feature (Durston et al., 2003; Horga et al., 2011) leaving open the possibility that those effects may reflect lower-level processes rather than effects of recent conflict experiences on cognitive control. Exceptions are the studies by Jiménez and Méndez (2013, 2014) which controlled for feature repetitions on the immediate timescale by grouping their four stimulus words/colors into two groups of alternating pairs. Therefore, neither the word nor color repeated on consecutive trials. However, the induction that created the biased runs (e.g., several incongruent or congruent trials) and which was critical for demonstrating recent effects of conflict beyond just the preceding trial did include trials that had stimulus features that overlapped with features on the diagnostic trial. For example, on the $n - 2$ trial, an incongruent trial in the color red may have been presented and the color red could then reappear on the diagnostic trial (trial n). In the present study, the induction trials were fully distinct from the diagnostic trials.

The second novel design feature is that we tested the above predictions by examining the effects of conflict in the recent timescale not merely on a single subsequent trial, as has been the typical approach to evaluating effects of preceding conflict (e.g., trial $n - 1$) on control (trial n) in studies investigating the immediate timescale as well as the recent timescale (Durston et al., 2003; Horga et al., 2011; Jiménez & Méndez, 2013; Jiménez &

Méndez, 2014). Instead, we extended the scope of the measurement to assess differences in control several trials following experiences with conflict. To achieve this goal, the effects of conflict experiences across different timescales were assessed during a diagnostic phase of eight trials that followed the critical manipulations of conflict. This enabled us to directly examine whether a conflict manipulation in the recent timescale produces transient changes in cognitive control limited to a single trial post conflict or, potentially, longer-lasting adjustments similar to the more sustained effects of the long timescale (e.g., De Pisapia & Braver, 2006; Aben et al., 2019). We predicted that the effect of conflict in the recent timescale would extend to the 8-trial diagnostic phase. Moreover, we expected that, if the effect of the recent timescale is unique from the effect of the immediate timescale, the effect of the recent timescale would persist even after excluding the first trial of the diagnostic phase from the analysis of diagnostic phase performance.

To foreshadow our results, consistent with the first prediction, Experiments 1 and 3 demonstrated a unique effect of the recent timescale on cognitive control that could not be explained by the long or immediate timescale, and the effect persisted across several subsequent trials. However, in Experiment 2, we found a boundary condition such that recent conflict did not affect control in the diagnostic phase when trials preceding the recent conflict were unbiased. Contrary to the second prediction, the effect of the recent timescale on control did not differ between MC and MI lists in either experiment that tested this prediction (Experiment 1 and Experiment 3). Consistent with our third prediction, the effect of the recent timescale on control persisted over a “long” unfilled delay that separated the induction and diagnostic phases in Experiment 3. Overall, these patterns indicate that the effects of the recent timescale on control can be dissociated from effects of other timescales, and the recent timescale may produce a qualitatively different adjustment in control (relatively sustained) in comparison to the immediate timescale (i.e., relatively transient; see Duthoo et al., 2014; Egner et al., 2010).

Experiment 1

Experiment 1 investigated the potential effects of the recent timescale on cognitive control using a novel variant of the abbreviated lists paradigm (Bugg et al., 2015; Bugg & Diede, 2017; Cohen-Shikora, Diede, & Bugg, 2018) that enables multiple observations per person per condition as opposed to a single observation per list (e.g., a single MC list). Each list comprised 26 Stroop trials. For expository purposes, consider that there were two phases in each list (see Figure 1): an induction phase (18 trials) which included a four-trial “window” of recent experience (Trials 15 – 18), and a diagnostic phase (8 trials) that immediately followed the induction phase. (Phases were not demarcated from the participants’ perspective.) Note that the entire induction phase represented the long timescale. Though studies implementing list-wide proportion congruence manipulations typically use lists of hundreds of trials (for a review, see Bugg, 2017), list-wide PC effects have been observed in abbreviated lists paradigms comprising inductions as brief as 6 to 10 trials (Bugg et al., 2015; Bugg & Diede, 2018; Cohen-Shikora et al., 2018), including on diagnostic trials mixed within the lists (Cohen-Shikora et al., 2018).

A key manipulation in Experiment 1 was that the induction phase was either MC or MI. In addition and critically, in half of the lists in each PC condition (MC and MI), the four-trial window representing the recent timescale comprised only the infrequent trial type for that condition (i.e., incongruent trials in an MC list; congruent trials in an MI list). This recent window manipulation is hereafter referred to as a "shifted" or "unshifted" window. For example, in a $MC_{SHIFTED}$ list, the last four trials would be 100% incongruent, whereas in an $MC_{UNSHIFTED}$ list, congruent and incongruent trials were distributed throughout the induction phase (including the window) in accordance with the PC of the list (in this example, most of those trials were congruent). Consequently, for the key comparison of $MC_{SHIFTED}$ and $MC_{UNSHIFTED}$ lists (and $MI_{SHIFTED}$ and $MI_{UNSHIFTED}$ lists), the long timescale was equated.

The effects of the induction, including the effects of the recent timescale, were assessed during the subsequent 8-trial diagnostic phase. This meant that, unlike prior studies investigating the transient effect of the recent timescale on just the immediately following trial (trial *n*), the current study assessed whether effects of the recent timescale might be sustained beyond that trial to multiple following trials. As noted earlier, the diagnostic phase was unbiased, and the diagnostic trials were novel words/colors not used to create the PC bias in the induction phase. The combination of these two features enabled us to rule out explanations of performance on the diagnostic trials related to lower-level processes and instead attribute differences in Stroop effects between conditions to induced cognitive control.

The first two predictions of our research (see Introduction) were tested in this experiment: 1) an effect of the recent timescale on control would be found as evidenced by performance in the diagnostic phase, and 2) there would be an interaction between the recent and long timescales. An account that includes a recent timescale of conflict experience that interacts with the other timescales would predict Stroop effects following an $MC_{SHIFTED}$ induction to be attenuated in comparison to $MC_{UNSHIFTED}$ and Stroop effects following an $MI_{SHIFTED}$ induction to be exacerbated in comparison to $MI_{UNSHIFTED}$. Statistically, this would manifest as a significant three-way interaction between trial type, PC, and recent window manipulation. Additionally, according to prior modeling results (Aben et al., 2017; Dey & Bugg, 2020), the effect of the recent timescale should be more pronounced in the MC condition than in the MI condition. To confirm that any differences between shifted and unshifted conditions (that is, between lists that had the same overall PC during induction [i.e., the long timescale] but differing experiences in the recent timescale) were driven by differences in the recent timescale and not the immediate timescale, we pre-registered an analysis to test for the three-way interaction after removing the trial immediately following the induction (i.e., excluding the first trial of the diagnostic phase from the analysis). If the adjustments following recent conflict reflect more than adjustments based on the immediate timescale, the anticipated differences in performance during the diagnostic phase should be observed even after removing the first trial of the diagnostic phase from the analysis. If the recent timescale has no influence on cognitive control beyond the long and immediate timescales, no differences should be observed between shifted and unshifted conditions (and therefore no three-way interaction). Hypotheses, method, and data for all experiments were pre-registered and are available on OSF at https://osf.io/zhpb4/ and https://osf.io/mkn5p/.

Method

Participants.—Sixty-one Washington University undergraduates (32 female, Age M = 18.49, SD = 0.64) participated for course credit. Our pre-registered sample size for this and all subsequent studies was to obtain data for at least 60 participants. We referenced prior literature to inform this decision. There are no direct design parallels in the literature and therefore we refer to two studies that incorporated primary features of our design: a manipulation of conflict in the recent timescale and assessment of diagnostic items in an abbreviated lists paradigm. Jiménez and Méndez, (2013, Experiment 1) manipulated conflict in the recent timescale by incorporating runs of entirely congruent or entirely incongruent trials and assessing effects on a subsequent diagnostic trial. The effect size for the critical interaction of context and trial type (i.e., the progressive sequential congruency effect) was η_p^2 = .17. Using GPower 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007), we calculated that to have .80 power to detect a comparably sized effect in the current study (alpha = .05), 12 participants are required. Cohen-Shikora et al. (2018, Experiment 2) assessed performance on diagnostic items intermixed with inducer items in an abbreviated lists paradigm and found an effect of list-wide PC on diagnostic items. The interaction between PC and trial type was η_p^2 = .06. To have .80 power to detect a comparably sized effect in the current study, 33 participants are required. To be conservative, we aimed to collect data from at least 60 participants.

All participants were native English speakers with normal or corrected vision and color vision. No participants were excluded.

Design and Stimuli.—We adapted an abbreviated-lists design (Bugg et al., 2015) using 26-trial lists presenting congruent trials comprising a word and color that matched (e.g., RED in red ink) and incongruent trials comprising a word and color that mismatched (e.g., RED in blue ink). As shown in Figure 1, lists began with an 18-trial induction phase that was MC or MI and ended with an 8-trial, unbiased diagnostic phase. The long timescale was conceptualized as the conflict experiences during the entire induction phase. The purpose of the induction phase was to present trials that induce relatively relaxed (i.e., MC list) or focused (i.e., MI list) control (see Bugg et al., 2015; Bugg & Diede, 2017; Cohen-Shikora et al., 2018); the effects of the induction were assessed during the diagnostic phase, which was equivalent across conditions. The key manipulation was the experience-defying manipulation of conflict experiences in the four-trial window (i.e., in the recent timescale) at the end of the induction phase, which preceded assessment of participants' cognitive control during the diagnostic phase. In $MC_{SHIFTED}$ lists, this window comprised four incongruent trials; in $MI_{SHIFTED}$ lists, it comprised four congruent trials. In unshifted lists, this window was simply a continuation of the PC in the first part of the induction. Accordingly, in $MC_{UNSHIFTED}$ lists, trials in the window were 70.9% congruent on average; in $MI_{UNSHIFTED}$ lists, trials in the window were 27.8% congruent on average.[3]

One set of stimuli (RED, BLUE, PURPLE, and WHITE in red, blue, purple, or white) served as the Induction Set and was presented during the induction phase according to the

[3]Across all $MC_{UNSHIFTED}$ lists, 34 of the 48 total trials in the window section were congruent. Across all $MI_{UNSHIFTED}$ lists, 13 congruent out of 48 total trials in the window section were congruent.

PC of the list. A second set of stimuli (words GREEN and YELLOW in green and yellow) served as the Diagnostic Set and was presented during the diagnostic phase. A key feature of stimuli in the diagnostic phase is that they were always unbiased. One concern in the current design was that participants might become aware that the words/colors green and yellow always appeared at the end of the list and this could inadvertently affect their cognitive control. To alleviate this concern, two preventive measures were taken: First, a congruent and an incongruent inducer set trial were randomly selected from the pre-window section (Trials 1–14) of the induction phase and interchanged with a congruent and incongruent diagnostic set trial from Trials 3–8 of the diagnostic phase. Induction Set trials transplanted into the diagnostic phase were excluded from the analysis of diagnostic phase performance (and vice versa). Second, filler lists were included in which 13 trials from the Induction Set and 13 trials from the Diagnostic Set were randomly intermixed throughout the list. These lists were unbiased and excluded from analysis.

Experiment 1 used 56 lists that were presented in random order: 12 lists for each of the following: $MC_{SHIFTED}$, $MC_{UNSHIFTED}$, $MI_{SHIFTED}$, and $MI_{UNSHIFTED}$, plus eight filler lists. As each of the 56 lists comprised 26 trials, there were 1,456 trials across the entire experiment. The order of the trials within lists was pseudorandom. Each color was equally represented for both the Induction and Diagnostic Sets. For incongruent trials in the Induction Set, there was an equal number of each distractor, such that for an incongruent trial with the color red, for example, the distractor word was equally likely to be PURPLE, BLUE, or WHITE. The order of trials within lists was fixed to establish the manipulation.

Procedure.—First, participants gave informed consent and completed a brief demographic survey. Participants were instructed to name the color as quickly as possible without sacrificing accuracy. Participants then performed a practice incongruent trial to ensure they understood that the goal of the task was to name the color and not read the word. Next, participants began the first list of the color-word Stroop task. For each trial, a word stimulus was presented centrally on screen in 24-point Arial font. The word remained on screen until the voice key was triggered after which an experimenter coded what response was emitted by the participant. Trials on which the voice key was triggered by irrelevant speech (e.g., “um”) or extraneous noise (e.g., cough), or on which the speech was imperceptible or unintelligible, were coded as scratch trials and excluded. A blank screen was then presented for 500 ms. Trials within each list were presented continuously (i.e., there was no break between phases within a list). In between each list, participants had an opportunity to rest and verbally told the experimenter when to continue. After completing all lists, participants were debriefed. The entire procedure lasted ~ 1 hour.

Results

In the current and subsequent experiment, we used an alpha of .05 for all analyses. In addition, analyses of RT and error rate excluded trials with RTs less than 200 ms or greater than 3000 ms (< 1% of trials were removed; cf. Bugg et al., 2015), and analyses of RT also excluded error trials. To facilitate interpretation, particularly of null effects, Bayesian analyses are reported in the form of BF_{01} values where a value between 1 and 3 means anecdotal evidence for the null hypothesis and a value between 3 and 10 means

substantial evidence for the null hypothesis (Wagenmakers, Wetzels, Borsboom, & Van Der Maas, 2011). The induction and diagnostic trials were analyzed separately (cf. Bugg, 2014); analyses of the diagnostic trials inform conclusions about control (see Braem et al., 2019). For each trial type and dependent variable, a 2 × 2 × 2 repeated measures ANOVA was performed with factors of trial type (congruent or incongruent), PC (MC or MI), and recent window manipulation (shifted or unshifted). All reaction times report milliseconds (ms). See Table 1 for descriptive statistics. Only theoretically relevant inferential statistics are reported; for comprehensive analyses see the supplementary materials.

Reaction Time

Induction items.—To assess Stroop performance on biased (i.e., MC or MI) trials preceding the diagnostic phase, Induction Set trials in the induction phase were analyzed. Recall that the induction phase comprised 14 trials pre-window plus the 4-trial window. There was a main effect of trial type, $F(1, 60) = 467.08$, $p < .001$, $\eta_p^2 = .89$, $BF_{01} < .001$, such that responses for congruent trials ($M = 599$, $SE = 11$) were faster than incongruent trials ($M = 702$, $SE = 11$). The interaction between trial type and PC, $F(1, 60) = 286.28$, $p < .001$, $\eta_p^2 = .83$, $BF_{01} < .001$, was significant, such that the Stroop effect ($Incongruent_{RT} - Congruent_{RT}$) was larger in MC than MI inductions (i.e., there was a list-wide PC effect). In addition, there was a significant three-way interaction between trial type, PC, and window, $F(1, 60) = 70.74$, $p < .001$, $\eta_p^2 = .54$, $BF_{01} < .001$. Follow up 2×2 repeated-measure ANOVAs with factors of trial type and window were performed separately for the MC and MI conditions. The Stroop effect was significantly larger in $MC_{SHIFTED}$ ($M = 163$, $SE = 7$) than $MC_{UNSHIFTED}$ lists ($M = 113$, $SE = 5$), $F(1, 60) = 107.91$, $p < .001$, $\eta_p^2 = .64$, $BF_{01} < .001$, whereas there was no difference between the Stroop effect in $MI_{SHIFTED}$ ($M = 64$, $SE = 6$) and $MI_{UNSHIFTED}$ lists ($M = 70$, $SE = 4$), $F(1, 60) = 2.43$, $p = .124$, $\eta_p^2 = .04$, $BF_{01} = 3.596$.

Diagnostic items.—More critically, effects of the induction on Stroop performance independent of known confounds were examined by analyzing the diagnostic phase. There was a main effect of trial type, $F(1, 60) = 298.82$, $p < .001$, $\eta_p^2 = .83$, $BF_{01} < .001$, such that responses to congruent trials ($M = 620$, $SE = 11$) were faster than incongruent trials ($M = 697$, $SE = 12$). The interaction between trial type and PC was significant, $F(1, 60) = 20.16$, $p < .001$, $\eta_p^2 = .25$, $BF_{01} = .199$, such that the Stroop effect was smaller in MC lists than MI lists. This effect was qualified by a significant three-way interaction between trial type, PC, and window, $F(1, 60) = 18.72$, $p < .001$, $\eta_p^2 = .24$, $BF_{01} = .074$ (see Figure 2). Follow up 2×2 repeated-measure ANOVAs with factors of trial type and window were performed separately for the MC and MI conditions. Demonstrating an effect of the recent window manipulation, $MC_{SHIFTED}$ lists ($M = 62$, $SE = 5$) had an attenuated Stroop effect compared to $MC_{UNSHIFTED}$ lists ($M = 78$, $SE = 6$), $F(1, 60) = 8.94$, $p = .004$, $\eta_p^2 = .13$, $BF_{01} = .804$, whereas $MI_{SHIFTED}$ lists ($M = 93$, $SE = 6$) had a larger Stroop effect than $MI_{UNSHIFTED}$ lists ($M = 76$, $SE = 5$), $F(1, 60) = 8.95$, $p = .004$, $\eta_p^2 = .13$, $BF_{01} = .594$.

To assess whether results for diagnostic items were driven by a congruency sequence effect based on just the immediately preceding trial of the induction phase, we re-analyzed performance in the diagnostic phase after excluding the first trial immediately following the

induction (i.e., trial 19) as specified in our pre-registration. The results converged with the above findings: a significant main effect of trial type, $F(1, 60) = 300.90$, $p < .001$, $\eta_p^2 = .83$, $BF_{01} <. 001$, such that responses to congruent trials ($M = 617$, $SE = 11$) were faster than incongruent trials ($M = 696$, $SE = 12$); a significant interaction between trial type and PC, $F(1, 60) = 14.42$, $p < .001$, $\eta_p^2 = .19$, $BF_{01} = .089$, such that the Stroop effect was smaller in MC lists; and most critically, a significant three-way interaction between trial type, PC, and window, $F(1, 60) = 8.15$, $p = .006$, $\eta_p^2 = .12$, $BF_{01} = .596$. Follow up 2×2 repeated-measure ANOVAs with factors of trial type and window were performed separately for the MC and MI conditions. $MC_{SHIFTED}$ lists ($M = 64$, $SE = 5$) had an attenuated Stroop effect compared to $MC_{UNSHIFTED}$ lists ($M = 76$, $SE = 6$), $F(1, 60) = 4.13$, $p = .047$, $\eta_p^2 = .06$, $BF_{01} = 1.862$, and $MI_{SHIFTED}$ lists ($M = 94$, $SE = 6$) had a larger Stroop effect than $MI_{UNSHIFTED}$ lists ($M = 81$, $SE = 6$), $F(1, 60) = 3.95$, $p = .051$, $\eta_p^2 = .06$, $BF_{01} = 1.741$, although this difference was marginal.

Finally, as an exploratory analysis[4], we compared performance in the first and second half of the diagnostic phase (i.e., comparing Trials 1–4 of the Diagnostic Phase to Trials 5–8 of the Diagnostic Phase) to determine whether adjustments in control persisted or decreased across halves. A four-way repeated measures ANOVA was run with factors of PC (MC or MI), window (Shifted or Unshifted), trial type (Congruent or Incongruent), and Half (First Half or Second Half). There was a four-way interaction $F(1, 60) = 5.96$, $p = .018$, $\eta_p^2 = .09$, $BF_{01} = .842$. The First Half and Second Half conditions were therefore examined in separate three-way repeated measures ANOVAs. In the First Half, mirroring the pattern observed in the overall analysis of the diagnostic phase, there was a significant three-way interaction between PC, window and trial type, $F(1, 60) = 25.75$, $p < .001$, $\eta_p^2 = .30$, $BF_{01} = .001$. In the Second Half, there was no interaction between PC, window and trial type, $F < 1$, $BF_{01} = 6.184$.

Error Rate

Induction items.—There was a main effect of trial type, $F(1, 60) = 64.86$, $p <.001$, $\eta_p^2 = .519$, $BF_{01} < .001$, such that congruent trials ($M = 0.55\%$, $SE = 0.11\%$) were more accurate than incongruent trials ($M = 4.70\%$, $SE = 0.71\%$). The interaction between trial type and PC was significant, $F(1, 60) = 39.15$, $p < .001$, $\eta_p^2 = .30$, $BF_{01} < .001$, such that the Stroop effect was larger in MC than MI inductions. Finally, the three-way interaction between trial type, PC, and window was significant, $F(1, 60) = 12.93$, $p < .001$, $\eta_p^2 = .177$, $BF_{01} = .144$. Follow up 2×2 repeated-measure ANOVAs with factors of trial type and window were performed separately for the MC and MI conditions. $MC_{SHIFTED}$ lists ($M = 7.02\%$, $SE = 1.00\%$) had a larger Stroop effect than $MC_{UNSHIFTED}$ lists ($M = 4.95\%$, $SE = 0.64\%$), $F(1, 60) = 8.42$, $p = .005$, $\eta_p^2 = .12$, $BF_{01} = .883$. $MI_{SHIFTED}$ lists ($M = 1.85\%$, $SE = 0.36\%$) had

[4]This exploratory analysis was suggested by two reviewers. We thank Luis Jiménez and an anonymous reviewer for the suggestion. Because this analysis was not planned, it is important to note several limitations. First, the order of trials in the diagnostic phase was randomized in each list; we did not control whether each half was exactly 50%. The percentage of congruent trials in the first half in each condition was: $MC_{SHIFTED}$: 48%; $MC_{UNSHIFTED}$: 44%; $MI_{SHIFTED}$: 46%; $MI_{UNSHIFTED}$: 52%. Second, because the Induction Set trials that were transplanted into (and not analyzed with) the diagnostic phase were placed in Trials 3–8 of the diagnostic phase, they were more likely to replace trials in the second half. For each condition, the percentage of Diagnostic Set trials replaced in the first and second half respectively was: $MC_{SHIFTED}$: 17%, 33%; $MC_{UNSHIFTED}$: 17%, 33%; $MI_{SHIFTED}$: 21%, 29%; $MI_{UNSHIFTED}$: 15%, 35%. This means there were fewer trials for this analysis in the second half of each list. For descriptive statistics for the first and second halves in both reaction time and error rate, see Supplementary Table 5.

a smaller Stroop effect than $MI_{UNSHIFTED}$ lists (M = 2.82%, SE = 0.43%), $F(1, 60) = 8.39$, $p = .005$, $\eta_p^2 = .12$, $BF_{01} = .842$.

Diagnostic items.—There was a main effect of trial type $F(1, 60) = 57.67$, $p < .001$, $\eta_p^2 = .49$, $BF_{01} < .001$, such that congruent trials (M = 0.59%, SE = 0.17%) were more accurate than incongruent trials (M = 4.01%, SE = 0.63%). There were no interactions between trial type and PC, $F < 1$, $BF_{01} = 5.309$, or between trial type, PC, and window, $F < 1$, $BF_{01} = 5.047$, (see Figure 2).

Although there was no hint of an effect of the recent window manipulation for error rate on diagnostic trials, for completeness we performed the analysis excluding the first trial. This did not appreciably change any of the above patterns.

For completeness, we also conducted the exploratory analysis by comparing the first and second half of the diagnostic phases. There was a significant four-way interaction between PC, window, trial type, and half, $F(1, 60) = 5.56$, $p < .022$, $\eta_p^2 = .09$, $BF_{01} = 1.057$. The First Half and Second Half conditions were therefore examined in separate three-way repeated measures ANOVAs. In the First Half, there was no three-way interaction between PC, window and trial type, $F(1, 60) = 1.07$, $p = .306$, $\eta_p^2 = .01$, $BF_{01} = 4.252$. In the Second Half, there was a significant three-way interaction between PC, window and trial type, $F(1, 60) = 4.65$, $p = .035$, $\eta_p^2 = .07$, $BF_{01} = .786$, but the pattern did not follow the predicted effects of recent conflict.

Discussion

Experiment 1 tested two predictions. In support of the first prediction, we found that the recent timescale affected cognitive control in the diagnostic phase. This finding provides converging evidence for the importance of the recent timescale for cognitive control (Aben et al., 2017; Dey & Bugg, 2020; Durston et al., 2003; Horga et al., 2011), evidence that more fully rules out the contribution of lower-level processes (cf. Jiménez & Méndez, 2013; Jiménez & Méndez, 2014). Incongruent windows representing the recent timescale in $MC_{SHIFTED}$ lists attenuated the Stroop effect during the diagnostic phase in comparison to $MC_{UNSHIFTED}$ lists; congruent windows representing the recent timescale in $MI_{SHIFTED}$ lists exacerbated the Stroop effect during the diagnostic phase in comparison to $MI_{UNSHIFTED}$ lists. The divergent Stroop effects across conditions matched in overall PC but with differing windows of experience four trials prior to the diagnostic phase (e.g., comparing $MC_{UNSHIFTED}$ and $MC_{SHIFTED}$ lists) can be uniquely attributed to the manipulation in the recent timescale. The long timescale cannot explain the effect because conflict experiences during the induction (i.e., frequency of congruent and incongruent trials prior to the diagnostic phase) were matched between shifted and unshifted conditions, yet these conditions differed. The immediate timescale also cannot explain the effect because the effect of recent conflict remained after we removed the first trial of the diagnostic phase, the trial on which any adjustments associated with the immediately preceding trial (last trial of the induction) would have been observed.

In contrast to the second prediction, we did not find evidence indicating that recent conflict experiences had a larger effect in the MC condition than in the MI condition. A statistical model anticipated that relatively recent experience should be weighted more strongly than relatively distal experience especially in MC lists (Aben et al., 2017; Dey & Bugg, 2020). Inconsistent with this model, the effect sizes were equivalent when comparing differences in the diagnostic phase between $MC_{SHIFTED}$ and $MC_{UNSHIFTED}$ and between $MI_{SHIFTED}$ and $MI_{UNSHIFTED}$ lists (η_p^2 = .13 in both cases), suggesting a similar effect of recent experience regardless of the PC of the list. Taken together, the findings of Experiment 1 provide evidence of an effect of recent conflict on subsequent cognitive control in MC and MI lists that cannot be accounted for if only experiences in the long or immediate timescale are considered. Further, the findings suggest that the effects of the recent timescale on control are equally potent in MC and MI lists.

The key predictions in Experiment 1 were tested by examining performance during an 8-trial, unbiased diagnostic phase using novel diagnostic items that did not share features with inducer items, thereby enabling us to a) examine cognitive control independent of known confounds, and b) to gauge whether effects of recent conflict experiences may be relatively sustained, here meaning that they were not limited to a single trial post induction. Regarding the first opportunity, the findings suggest that effects of recent conflict affected cognitive control (i.e., extent to which attention was biased toward word versus color information) and not lower-level processes, processes that could potentially account for performance on inducer trials. One also cannot attribute the current findings demonstrating effects of recent conflict on cognitive control to adjustments that may have occurred based on the immediate timescale within the diagnostic phase as effects of such adjustments should be equivalent across lists, given all lists entailed an unbiased diagnostic phase. Regarding the second opportunity, the fact that effects of recent conflict were found when averaging across trials in the diagnostic phase, including after removing the first trial, significantly expands prior findings and raises the very interesting possibility of relatively sustained adjustments in control based on conflict experiences in the recent timescale. Even if the effects of recent conflict are limited to just the first half of the diagnostic phase as the exploratory analysis of reaction time suggested, that still implies a relatively sustained effect given the first half comprised four diagnostic trials. Also, as a reviewer pointed out, the effect may have been limited to the first half if the first four trials (which were ~50% congruent across lists) served as the recent experience for the last four trials. We revisit these interesting theoretical possibilities in Experiment 3.

Finally, we want to comment on a finding that may strike readers as non-intuitive, which concerns the relationship between the Stroop effects observed in the induction and the diagnostic phases. For the MC condition, the $MC_{SHIFTED}$ lists had a larger Stroop effect than the $MC_{UNSHIFTED}$ lists in the induction phase (nominally, this same pattern was present in the MI condition), but a smaller Stroop effect in the diagnostic phase as discussed above. We believe this is attributable to differences in the composition of the induction phase for the shifted vs. unshifted conditions. Let us illustrate with the MC condition. To equate the long timescales across the $MC_{SHIFTED}$ and the $MC_{UNSHIFTED}$ lists while implementing the critical manipulation of recent conflict within the window, the five incongruent trials were differentially distributed within the induction phase across these two conditions. In

the pre-window section of the $MC_{SHIFTED}$ lists, there were 13 congruent trials and one incongruent trial (followed by four incongruent trials in the window). This means that there was a high probability that the single incongruent trial would occur after several consecutive congruent trials, that is after one likely developed a very relaxed scope of attention. Thus, reaction time on that incongruent trial was likely exacerbated while reaction time on the surrounding congruent trials was speeded, creating a large Stroop effect in the induction phase. By contrast, in the $MC_{UNSHIFTED}$ lists, the five incongruent trials were randomly distributed across the entire induction (pre-window and window combined) such that participants encountered incongruent trials more "regularly", rather than primarily after several consecutive congruent trials.

In summary, Experiment 1 provided evidence that the recent timescale of conflict accumulation influences cognitive control independent of other timescales, and furthermore showed that the effects of the recent timescale on control adjustments in the diagnostic phase were comparable for MC and MI lists. In Experiment 2, we aimed to investigate a potential boundary condition for the effect of the recent timescale. In Experiment 1, one could say that the recent timescale strongly "defied experience" in that the window was comprised entirely of the type of trial participants rarely experienced before the window (e.g., a window of incongruent trials in an MC list). As such, in shifted lists, the recent conflict experience differed markedly from the preceding experience (e.g., shift from 93% congruent pre-window to 0% congruent during the window in $MC_{SHIFTED}$ induction). A key question is whether such an extreme shift is necessary for conflict experiences within the recent window to drive subsequent adjustments in attention.

Experiment 2

Experiment 2 aimed to assess the possibility that the effects of the recent timescale depend on the experience within that timescale strongly defying experience in the long timescale. Theoretically, such defiance of experience may be important for the emergence of these effects because it plausibly generates a prediction error (e.g., den Ouden, Kok, & de Lange, 2012; see also Chiu, Jiang, & Egner, 2017 and Jiang, Heller, & Egner, 2014 for models that have incorporated prediction error) and such errors have been shown to be consequential in producing adjustments to cognitive control (Brown & Braver, 2005; Alexander & Brown, 2011). To address this possibility, Experiment 2 manipulated conflict in the recent timescale while reducing the degree to which experience in the window defied the experience preceding the window.

As in Experiment 1, each list was comprised of an induction phase followed by a diagnostic phase. However, in Experiment 2, the induction phase prior to the window (i.e., the 14- trial pre-window) was always unbiased (i.e., 50% congruent). Again, recent conflict was manipulated within the last four trials of the induction phase (i.e., window) such that the window was entirely congruent, entirely incongruent, or unbiased. In Experiment 1, the experience within the window strongly defied previous experience. For example, an entirely incongruent window in Experiment 1 (see Figure 1) represented a shift from a 93% congruent experience pre-window to a 0% congruent experience during the window in $MC_{SHIFTED}$ lists. The experience within the window in Experiment 2 defied

previous experience comparatively weakly. For example, an entirely incongruent window in Experiment 2 (see Figure 3) represented a shift from a 50% congruent experience pre-window to a 0% congruent experience during the window. The diagnostic phase was identical to Experiment 1. Therefore, this experiment again had the opportunity to inform the theoretical question of whether effects of the recent timescale, if observed, on cognitive control are transient or relatively sustained across multiple trials.

If the effect of recent conflict does not depend on strong defiance of the long timescale as in Experiment 1, one would predict that the window manipulation in Experiment 2 should lead to differences in cognitive control during the diagnostic phase, as in Experiment 1. This result would be consistent with our first prediction, regarding the effect of the recent timescale on subsequent control. That is, the Stroop effect should be smallest after an incongruent window, intermediate after an unbiased window, and largest after a congruent window. Statistically, this effect would manifest as an interaction between trial type and window type during the diagnostic phase. Alternatively, if the effects of conflict accumulation in the recent timescale rely on strongly defying previous experience, given that there is objectively weaker defiance in Experiment 2, one would not predict a difference between the three conditions, or the difference may be smaller. Statistically, there would be no interaction between trial type and window type.

Method

Participants.—Sixty-two Washington University undergraduates (43 female, Age M = 20.03, SD = 1.43) participated for course credit. All participants were native English speakers with normal or corrected vision and color vision. One participant was excluded for falling asleep during the task, and one participant was excluded for difficulty using the microphone. Therefore, 60 were included in the reported analysis (42 female, Age M = 20.05, SD = 1.43). The most relevant effect to inform power for Experiment 2 was the three-way interaction observed in Experiment 1. The effect size for the interaction between window, PC, and trial type was η_p^2 = .24. Once again using GPower 3.1 (Faul et al., 2007) we found that to have .80 power to detect a comparably sized effect in the current study (alpha = .05), 9 participants are required. Again, being conservative, we aimed to collect data from at least 60 participants.

Design and Stimuli.—As in Experiment 1, each list was comprised of an induction phase and a diagnostic phase (see Figure 1) and the induction and diagnostic sets were identical to Experiment 1. The pre-window section of the induction phase began with 14 unbiased (50% congruent) trials. In contrast, the four-trial window at the end of the induction phase varied in PC (congruent, incongruent, or unbiased). Contrasting Experiment 1, the long timescale was not matched across conditions (instead, the long timescale was 61.11%, 38.88%, and 50%, respectively, in the congruent, incongruent, and unbiased window conditions). This decision was justified by our finding in Experiment 1 that the long timescale was not the determinant of adjustments in control in the diagnostic phase, and the decision allowed us to manipulate the recent timescale (experience within the window) while holding experience preceding the manipulation constant across all conditions. In Experiment 1, one could not compare effects of congruent and incongruent windows on control because they were in lists

that differed in their pre-window sections ($MI_{SHIFTED}$ and $MC_{SHIFTED}$, respectively). In Experiment 2, one can compare the effects of congruent and incongruent windows because they are preceded by the same initial experience within the long timescale (50% congruent). The diagnostic phase was equivalent to Experiment 1.

There were 44 lists in the experiment including 12 lists for each of the following conditions of window type: congruent window, unbiased window, and incongruent window. Additionally, again there were eight filler lists and the order of the trials within the lists was pseudorandom and fixed to maintain the manipulation. As each of the 44 lists comprised 26 trials, there were 1144 trials across the entire experiment. The order in which the 44 lists were presented was random.

Procedure.—The procedure was identical to Experiment 1 with the exception that there were 44 lists of 26 trials. The entire procedure lasted ~ 45 min.

Results

The RT trim removed < 1% of trials. All analyses used a 2 × 3 repeated measures ANOVA with factors of trial type (congruent or incongruent) and window type (congruent, incongruent, or unbiased). See Table 2 for descriptive statistics. Only theoretically relevant inferential statistics are reported; for comprehensive analyses see the supplementary materials.

Reaction Time

Induction items.—As in Experiment 1, induction analysis included Induction Set trials in the entire induction (14 pre-window trials and the four trials in the window). There was a significant main effect of trial type, $F(1, 59) = 242.68$, $p < .001$, $\eta_p^2 = .80$, $BF_{01} < .001$, such that responses to congruent trials ($M = 614$, $SE = 12$) were faster than incongruent trials ($M = 719$, $SE = 15$). There was a significant main effect of window, $F(2, 118) = 3.48$, $p = .034$, $\eta_p^2 = .06$, $BF_{01} = 23.066$, such that performance during an induction with a congruent window ($M = 662$, $SE = 14$) was significantly faster than during an induction with an unbiased window ($M = 669$, $SE = 14$, $t(59) = 2.49$, $p = .016$) and an induction with an incongruent window ($M = 668$, $SE = 13$, $t(59) = 2.05$, $p = .046$); performance during inductions with unbiased and incongruent windows did not differ, $t(59) = 0.55$, $p = .58$. There was no interaction between window and trial type, $F(2, 118) = 0.13$, $p = .876$, $\eta_p^2 < .01$, $BF_{01} = 17.601$, such that the Stroop effects in the induction did not differ across conditions.

Diagnostic items.—There was a significant main effect of trial type, $F(1, 59) = 99.77$, $p < .001$, $\eta_p^2 = .63$, $BF_{01} < .001$, such that responses to congruent trials ($M = 629$, $SE = 13$) were faster than incongruent trials ($M = 724$, $SE = 17$). There was a significant main effect of window, $F(2, 118) = 3.39$, $p = .037$, $\eta_p^2 = .05$, $BF_{01} = 13.551$. Responses following a congruent window ($M = 671$, $SE = 15$) were significantly faster than responses following an incongruent window ($M = 683$, $SE = 15$), $t(59) = 2.48$, $p = .016$), but responses following a congruent window and an unbiased window ($M = 674$, $SE = 14$) did not differ, $t(59) = 0.64$, $p = .526$. Responses following an incongruent window were marginally faster than responses

following an unbiased window, $t(59) = 1.95$, $p = .056$. There was no interaction between window and trial type, $F < 1$, $BF_{01} = 17.035$ (see Figure 4).

Although there was no hint of an effect of the recent window manipulation in the performance on diagnostic trials, for completeness we report the analysis excluding the first trial in the diagnostic phase. Again, there was a significant main effect of trial type, $F(1, 59) = 100.21$, $p < .001$, $\eta_p^2 = .63$, $BF_{01} < .001$, such that responses to congruent trials ($M = 624$, $SE = 12$) were more accurate than incongruent trials ($M = 721$, $SE = 17$). The main effect of window was no longer significant, $F(2, 118) = 1.93$, $p = .15$, $\eta_p^2 = .032$, $BF_{01} = 20.396$, and there was still no interaction between window and trial type, $F(2, 118) = 1.38$, $p = .256$, $\eta_p^2 = .02$, $BF_{01} = 12.374$.

For consistency with Experiment 1, we compared the first and second half[5] of the diagnostic phase in an exploratory analysis. A three-way repeated measures ANOVA was run with factors of Window Type (Congruent Window, Incongruent Window, and Unbiased Window), Trial Type (Congruent or Incongruent), and Half (First Half and Second Half). There was no three-way interaction between Window Type, Trial Type, and Half, $F < 1$, $BF_{01} = 17.174$.

Error Rate

Induction items.—There was a significant main effect of trial type, $F(1, 59) = 74.51$, $p < .001$, $\eta_p^2 = .56$, $BF_{01} < .001$, such that responses to congruent trials ($M = 0.57\%$, $SE = 0.12\%$) were more accurate than incongruent trials ($M = 4.01\%$, $SE = 0.45\%$). There was no main effect of window, $F < 1$, $BF_{01} = 28.983$. There was a significant interaction of window and trial type, $F(2, 118) = 4.30$, $p = .016$, $\eta_p^2 = .07$, $BF_{01} = 2.394$, such that the Stroop effect was smaller in the congruent window condition ($M = 2.80\%$, $SE = 0.41\%$) than the incongruent window ($M = 3.47\%$, $SE = 0.51\%$) and unbiased window ($M = 4.03\%$, $SE = 0.48\%$) conditions.

Diagnostic items.—There was a significant main effect of trial type, $F(1, 59) = 47.96$, $p < .001$, $\eta_p^2 = .45$, $BF_{01} < .001$, such that responses to congruent trials ($M = 0.51\%$, $SE = 0.16\%$) were more accurate than incongruent trials ($M = 4.18\%$, $SE = 0.65\%$). There was a significant main effect of window, $F(2, 118) = 6.41$, $p = .002$, $\eta_p^2 = .10$, $BF_{01} = 2.075$, such that performance following a congruent window ($M = 3.03\%$, $SE = 0.60\%$) was less accurate than an unbiased window ($M = 2.09\%$, $SE = 0.51\%$), $t(59) = 2.91$, $p = .005$, or an incongruent window ($M = 1.92\%$, $SE = 0.46\%$), $t(59) = 3.02$, $p = .004$. Accuracy did not differ following an induction with an unbiased window compared to an incongruent window, $t(59) = 0.57$, $p = .572$. However, there was a significant interaction between window and trial type, $F(2, 118) = 5.06$, $p = .008$, $\eta_p^2 = .08$, $BF_{01} = 1.154$ (see Figure 4). The Stroop effect was smallest following an induction with an incongruent window ($M = 2.93\%$, $SE = 0.59\%$), followed by a congruent window, ($M = 3.2\%$, $SE = 0.72\%$) and the unbiased window ($M = 3.53\%$, $SE = 0.63\%$).

[5]In Experiment 2, a limitation of this analysis is that estimates of performance in the second half are based on fewer trials. For each condition, the percentage of Diagnostic Set trials replaced in the first and second half respectively was: Congruent Window: 17%, 33%; Incongruent Window: 17%, 33%, Unbiased Window: 17%, 33%. For descriptive statistics for the first and second halves in both reaction time and error rate, see supplementary Table 6.

In the diagnostic phase, excluding the first trial, there was a significant main effect of trial type, $F(1, 59) = 43.44$, $p < .001$, $\eta_p^2 = .42$, $BF_{01} < .001$, such that responses to congruent trials ($M = 0.46\%$, $SE = 0.16\%$) were more accurate than incongruent trials ($M = 4.13\%$, $SE = 0.68\%$). There was a significant main effect of window, $F(2, 118) = 3.30$, $p = .040$, $\eta_p^2 = .05$, $BF_{01} = 7.573$; performance following an induction with a congruent window ($M = 2.80\%$, $SE = 0.60\%$) was significantly less accurate than performance following an unbiased window ($M = 2.15\%$, $SE = 0.55\%$), $t(59) = 2.14$, $p = .036$, and an incongruent window ($M = 1.95\%$, $SE = 0.49\%$), $t(59) = 2.17$, $p = .034$. Performance did not differ between unbiased and incongruent conditions, $t(59) = 0.50$, $p = .623$. The interaction between window and trial type was no longer significant, $F(2,118) = 2.14$, $p = .123$, $\eta_p^2 = .04$, $BF_{01} = 5.583$.

We compared the first and second half of the diagnostic phase in the same exploratory analysis performed for reaction time. There was no interaction between Window Type, Trial Type, and Half, $F < 1$, $BF_{01} = 11.186$.

Discussion

Experiment 2 tested the prediction that conflict experiences in the recent timescale would affect control as evidenced by performance in the diagnostic phase, here under conditions in which experiences preceding the recent timescale were unbiased in all lists. Although the Stroop effect in reaction time during the diagnostic phase did not differ between the varying window conditions, the recent timescale did affect the Stroop effect in error rate. However, this effect did not survive after controlling for the effect of the immediate timescale (i.e., after excluding the first trial in the diagnostic phase from the analysis). It should also be noted that the frequentist and Bayesian evidence occasionally contradicted each other. Specifically, the evidence was not consistent for the main effect of window in the induction and diagnostic phases for reaction time and the diagnostic phase excluding the first trial for error rate. Even though frequentist analyses found a significant main effect of Window in the diagnostic phase, the Bayesian analyses demonstrated evidence favoring the null. Moreover, there was no effect of the recent timescale selectively in the first half of the diagnostic trials as could have been possible if the effect dissipated over time. All things considered, the results did not favor the first prediction. That is, there was not an effect of recent conflict experiences on cognitive control that was uniquely attributable to the recent timescale.[6]

Recall that Experiments 1 and 2 were similar in the manipulation of the recent timescale (i.e., four-trial windows that were entirely congruent or entirely incongruent at the end of the induction phase) but differed in the pre-window section of the induction. The key difference was that the pre-window section of the induction more strongly defied the window section in Experiment 1 compared to Experiment 2. Put simply, the same recent conflict manipulation that modulated control in Experiment 1 (i.e., presentation of four consecutive congruent or incongruent trials) did not affect cognitive control in Experiment 2. We interpret this result to mean that if the conflict in the recent timescale sufficiently defies previous experience

[6]It might be suggested that differences, if observed in the diagnostic phase, could have been attributable to differences in the long timescale rather than the recent timescale. As acknowledged in the method section, the long timescales were not equated in this experiment and therefore this is possible. Had we observed differences in the diagnostic phase, additional experimentation would have been needed to completely rule out this possibility.

(as set by the induction trials), then conflict information in the recent timescale may be weighted more strongly than distal conflict information. We will revisit this interpretation in the general discussion.

The findings of Experiment 2 demonstrated a boundary condition --when the recent timescale was preceded by 50% congruent trials, conflict in the recent timescale did not affect cognitive control. The findings may seem surprising given the findings of prior studies that manipulated conflict experiences several trials before a diagnostic trial within unbiased lists (Durston et al., 2003; Horga et al., 2011; Jiménez & Méndez, 2013; Jiménez & Méndez, 2014). These studies found that performance on trial *n* did vary as a function of the amount of conflict in the recent timescale. The results in Experiment 2 are not necessarily inconsistent with these studies, however, because of two important methodological differences. First, the diagnostic trials in Experiment 2 were from the Diagnostic Set, and they were therefore a different color and word than any trial experienced in the preceding window. To observe an effect on these diagnostic trials, adjustments in control had to be sufficiently abstract to extend from the induction to new stimulus features in the diagnostic phase. That was not the case in the prior studies where the diagnostic trial (*n*) could comprise the same features as preceding inducer trials in the recent timescale (Durston et al., 2003; Horga et al., 2011) or a subset of those trials (e.g., trial n − 2; Jiménez & Méndez, 2013; 2014). Thus, it is possible that in prior studies effects of the recent timescale were found even though lists were unbiased because these effects may at least in part reflect lower-level processes such as feature repetition.

A second methodological consideration is that Experiment 2 considered a larger diagnostic scope (*n* + 7 trials) than prior studies (trial *n*). To know whether a similar effect was found in Experiment 2 as in the prior studies, one would evaluate evidence for an effect of the recent timescale on the first trial of the diagnostic phase. Some evidence exists for such an effect. Specifically, reaction time was slower overall following the incongruent window in the diagnostic phase and error rate was significantly higher following a congruent window in the diagnostic phase but neither effect remained after removing the first trial of the diagnostic phase, suggesting much of the observed effects in Experiment 2 could be accounted for by congruency sequence effects that were driven by the final trial of the induction.

To take stock, while Experiment 1 found an effect of recent conflict on cognitive control that could not be accounted for by the long and immediate timescales, Experiment 2 demonstrated a boundary condition for this effect, raising the possibility that the recent timescale may not uniquely affect control when it does not strongly defy the long timescale. Although it remains to be determined what constitutes “sufficiently” defying experience, this boundary condition may demonstrate the importance of *prediction error*. Further discussion of this possibility is saved for the general discussion.

Experiment 3

The primary goals of Experiment 3 were twofold. First, we aimed to systematically replicate the novel pattern observed in Experiment 1 whereby conflict experiences in the recent timescale in MC and MI lists uniquely affected cognitive control during a diagnostic

phase. Second, whereas Experiment 1 demonstrated that the effects of the recent timescale could be dissociated from effects of the immediate timescale (by analyzing the diagnostic phase after excluding the first trial), Experiment 3 aimed to inform a related but distinct theoretical question, namely whether the immediate and recent timescales also may produce qualitatively different adjustments in control. This question was inspired by our observation in Experiment 1 that effects of conflict experiences in the recent timescale were observed during a relatively lengthy diagnostic phase comprising multiple trials, raising the possibility that the adjustments based on recent conflict may be relatively sustained. If further evidence for this possibility were found, it would stand in stark contrast to previously documented effects of the immediate timescale on cognitive control that appear to be transient.

Prior studies have established that adjustments based on conflict in the immediate timescale are relatively transient in that they do not affect performance on the subsequent trial if a long amount of time has elapsed between trials. Egner et al. (2010) found that congruency sequence effects were observed for RSIs representing “short” delays between 500 and 2000 ms but not for RSIs representing “long” delays of 2,500 ms or greater (longest was 5000 ms). Duthoo et al. (2014) found a similar result with significant congruency sequence effects for short delays (750 and 1500 ms RSIs) but not long delays (2250 or 3000 ms RSIs). The findings of Experiment 1 raised the interesting possibility that conflict experiences in the recent timescale may produce a more sustained effect, like experiences on the long timescale (e.g., De Pisapia & Braver, 2006; Aben et al., 2019). Indeed, considering just Trial 2 of the diagnostic phase (estimates are even longer for the subsequent trials), on average ~ 2500 ms elapsed between the last trial in the window and assessment of the effects of recent conflict in the diagnostic phase.[7]

To provide a direct test of the possibility that adjustments based on recent conflict sustain across a long delay, we employed the delay (RSI) manipulation that has been empirically shown to eradicate the effects of the immediate timescale on subsequent control at a long delay (Duthoo et al., 2014; Egner et al., 2010). The design was almost identical to Experiment 1 save for a few exceptions (see Method section below), with the most critical being that we manipulated the length of time between the participant’s response to the final trial of the induction phase and the onset of the first trial of the diagnostic phase. The short delay was 1000 ms and the long delay was 4000 ms, consistent with the operational definitions in the prior studies (Duthoo et al., 2014; Egner et al., 2010). If adjustments based on recent conflict are like adjustments based on immediate conflict, then the effects of recent conflict should be relatively transient and observed in the short but not the long delay condition. However, if conflict experiences in the recent timescale produce adjustments that are relatively sustained, then the effects should be observed in both the short and long delay conditions. This would demonstrate that conflict in the recent timescale has a unique operating characteristic that distinguishes it from the immediate timescale.

[7]Assuming average RT on a given trial is ~650 ms (as in Experiment 1) and the average time for the experimenter to code the response is 417 ms (as in Experiment 1), there would be ~ 2500 ms that passed between the last trial of the induction and the second trial of the diagnostic phase. This includes a 417 ms experimenter response, followed by a 500 ms blank screen, followed by the next stimulus, a 417 ms experimenter response and 500 ms blank screen. Note that the time between the last trial of the induction and the first trial of the diagnostic phase in Experiment 1 was ~1000 ms.

One may ask whether the delay manipulation provides any theoretical insights not already addressed by observing adjustments in control during the 8-trial diagnostic phase, as in Experiment 1. The novel insights it affords concern the question of what it means for control to be sustained, a defining feature of proactive control (Braver et al., 2007; DePisapia & Braver, 2006). Thus far, we have demonstrated that conflict experiences in the recent timescale lead to adjustments in control that affect not just the immediately following trial, but multiple subsequent trials with ~ 1000 ms between trials (Experiment 1). What Experiment 3 examines is whether adjustments based on the recent timescale are still observed when there is an *unfilled* delay between the experiences in the recent timescale and the start of the diagnostic phase. In other words, will the focused or relaxed state of cognitive control triggered by experiences in the recent timescale sustain in the sense that it continues to persist across the long delay (and therefore affect the diagnostic phase) even though participants are not responding to trials with goal relevant features that potentially reinforce maintenance of control. Thus, Experiment 3 provides an opportunity to seek converging evidence for the possibility that effects of recent conflict may be sustained, not just in the sense of continuing to operate across trials that may provide some bottom-up support for maintenance of control (as in Experiment 1) but additionally in the sense of continuing to operate across a long unfilled delay without any bottom up support for maintenance of control. Observing the latter would reinforce the view that recent conflict leads to relatively proactive adjustments in control.

Method

Participants.—74 Washington University undergraduates (56 female, Age M = 19.39, SD = 1.23) participated for course credit. All participants were native English speakers with normal or corrected vision and color vision. Our pre-registered inclusion criterion was the completion of at least 7 lists in each of the eight conditions within the 1-hour experiment; we chose this cut-off to alleviate concerns with having too few observations per condition. This criterion was included because we were concerned not all participants would finish all 64 lists in the hour (by comparison, Experiment 1 had 56 lists). Seven participants were excluded for not reaching that mark. In addition, four participants were excluded for falling asleep during the task, one participant was excluded due to a computer error, and one participant was excluded for difficulty using the microphone. Therefore, 61 were included in the reported analysis (46 female, Age M = 19.26, SD = 1.18).

The most relevant effect to inform power for Experiment 3 was the three-way interaction observed in Experiment 1. The effect size for the interaction between window, PC, and trial type was η_p^2 = .24. Once again using GPower 3.1 (Faul et al., 2007), we found that to have .80 power to detect a comparably sized effect in the current study (alpha = .05), 9 participants are required. Again, to be conservative and bearing in mind we were additionally testing another factor (delay) in this experiment, we aimed to collect data from 60 participants.

Design and Stimuli.—As in Experiment 1, each list was comprised of an induction phase and a diagnostic phase (see Figure 5) and the induction and diagnostic item sets were identical to Experiment 1. The trial composition of the pre-window, window, and diagnostic

phases were also identical to Experiment 1. The novel manipulation in Experiment 3 was the amount of time between the end of the induction and the beginning of the diagnostic phase (i.e., the delay). The delay was short (1000 ms) or long (4000 ms). Consequently, the design was a 2 × 2 × 2 × 2 within subjects design with levels of trial type (congruent or incongruent), PC (MC or MI), recent window manipulation (shifted or unshifted), and delay (short or long).

There were eight conditions in the experiment. Each of the $MC_{UNSHIFTED}$, $MC_{SHIFTED}$, $MI_{UNSHIFTED}$, and $MI_{SHIFTED}$ lists from Experiment 1 was presented with a short delay and a long delay. Thus, there were 64 lists in the experiment, eight of each of the eight conditions. As each of the 64 lists comprised 26 trials, there were 1,664 trials across the entire experiment. Again, the order of the trials within the lists was pseudorandom and fixed to maintain the manipulation. However, there were two changes from Experiment 1. First, in all unshifted lists, the induction phase in four lists ended with a congruent trial and in four lists ended with an incongruent trial. Second, in all conditions, the diagnostic phase began with a congruent trial in four lists and began with an incongruent trial in four lists. We made these changes to provide a more equivalent experience across conditions. Third, the different types of lists were presented in “sets” of eight, such that every condition was seen once in each set, unbeknownst to participants, before the next set began. Again, this was done so that if a participant did not complete all 64 lists, they would have an approximately even number in each condition. The order of the lists within each set and the order of the sets within the experiment were randomized for each participant. Finally, there were no filler lists. This change was made in order to prioritize observations in the theoretically relevant conditions.

Procedure.—With a few exceptions, the procedure was identical to Experiment 1. A new ISI procedure was used such that the time between trials (i.e., the blank screen) was 1000 ms minus the amount of time it took the experimenter to code the response. For example, if the experimenter took 417 ms to code the response, the remaining ISI for the trial would be 583 ms. Therefore, the total time between trials in Experiment 3 was consistently 1000 ms. Because a primary goal of Experiment 3 was to assess how long adjustments based on recent conflict are sustained, we employed this change to better control for the time between trials, which was important for implementing the delay manipulation. To implement the delay manipulation (short delays of 1000 ms and long delays of 4000 ms between the end of the induction and beginning of the diagnostic phase), we added no additional delay to the typical ISI in the short delay condition and a 3000 ms delay in the long delay condition between the induction and diagnostic phases, resulting in delays of 1000 ms and 4000 ms, respectively. The delays simply appeared as a continuous blank screen. The entire procedure lasted ~ 1 hour.

Results

The RT trim removed < 1% of trials. All analyses used a 2 × 2 × 2 × 2 repeated measures ANOVA with factors of trial type (congruent or incongruent), proportion congruence (MC or MI), window type (shifted or unshifted), and delay length (long or short). See Tables 3 and 4 for descriptive statistics. Only theoretically relevant inferential statistics are reported; for comprehensive analyses see the supplementary materials.

Reaction Time

Induction items.—As in Experiment 1, induction analysis included the 14 pre-window trials and the four trials in the window. In the induction, congruent trials (M = 618, SE = 8) were responded to more quickly than incongruent trials (M = 728, SE = 9), $F(1, 60)$ = 479.05, $p < .001$, $\eta_p^2 = .89$, $BF_{01} < .001$. Additionally, there was a significant interaction between trial type and PC, such that the Stroop effect was larger in MC lists (M = 141, SE = 7) than MI lists (M = 79, SE = 6) $F(1, 60) = 239.43$, $p < .001$, $\eta_p^2 = .80$, $BF_{01} < .001$. There was also a significant three-way interaction between trial type, PC, and window, $F(1, 60) = 12.43$, $p = < .001$, $\eta_p^2 = .17$, $BF_{01} = .330$. Follow up 2×2 repeated-measure ANOVAs with factors of trial type and window were performed separately for the MC and MI conditions. The Stroop effect was significantly larger in $MC_{SHIFTED}$ (M = 155, SE = 7) than $MC_{UNSHIFTED}$ lists (M = 126, SE = 7), $F(1, 60) = 34.90$, $p < .001$, $\eta_p^2 = .37$, $BF_{01} = .001$, whereas the Stroop effect did not differ between $MI_{SHIFTED}$ (M = 82, SE = 6) and $MI_{UNSHIFTED}$ (M = 76, SE = 6) lists, $F(1, 60) = 2.43$, $p = .124$, $\eta_p^2 = .04$, $BF_{01} = 4.623$. There was no four-way interaction between trial type, PC, window, and delay length, $F < 1$, $BF_{01} = 4.459$.

Diagnostic items.—The 2 × 2 × 2 × 2 ANOVA revealed a significant main effect of trial type, such that congruent trials (M = 632, SE = 10) were responded to more quickly than incongruent trials (M = 710, SE = 11) $F(1, 60) = 245.79$, $p < .001$, $\eta_p^2 = .80$, $BF_{01} < .001$. The interaction between trial type and PC was significant, such that the Stroop effect was smaller in the MC lists (M = 69, SE = 7) than the MI lists (M = 88, SE = 8), $F(1, 60) = 20.93$, $p < .001$, $\eta_p^2 = .26$, $BF_{01} = .007$. More importantly, there was a significant three-way interaction between trial type, PC, and window, $F(1, 60) = 17.74$, $p < .001$, $\eta_p^2 = .23$, $BF_{01} = .015$. Follow up 2×2 repeated-measure ANOVAs with factors of trial type and window were performed separately for the MC and MI conditions. In MC lists, there was a significant interaction between window and trial type, such that the Stroop effect was attenuated on diagnostic items in the $MC_{SHIFTED}$ (M = 62, SE = 7) lists compared to the $MC_{UNSHIFTED}$ lists (M = 76, SE = 6), $F(1, 60) = 7.70$, $p = .007$, $\eta_p^2 = .11$, $BF_{01} = .911$. In MI lists, there was a significant interaction between window and trial type, such that the Stroop effect was larger in the $MI_{SHIFTED}$ (M = 99, SE = 8) lists compared to the $MI_{UNSHIFTED}$ lists (M = 77, SE = 7), $F(1, 60) = 13.29$, $p < .001$, $\eta_p^2 = .18$, $BF_{01} = .532$. Critically, the four-way interaction between trial type, PC, window, and delay length was not significant, $F < 1$, $BF_{01} = 3.505$, implying that the three-way Trial Type × PC × Window interaction did not differ between long and short delay conditions (see Figure 6).

For completeness, we examined whether the effects of recent conflict were evident on diagnostic trials separately in the short and long delay conditions. We ran 2 × 2 × 2 ANOVAs for the short and long conditions, with factors of trial type, PC, and window. In the short delay condition, there was a significant three-way interaction of trial type, PC, and window, $F(1, 60) = 6.26$, $p = .015$, $\eta_p^2 = .10$, $BF_{01} = .927$. The interaction was also evidenced in the long delay condition, $F(1, 60) = 12.52$, $p < .001$, $\eta_p^2 = .17$, $BF_{01} = .048$.

As in previous experiments, the diagnostic phase analyses were repeated after excluding the first trial of the diagnostic phase. In the full 2 × 2 × 2 × 2 ANOVA, the three-way interaction

between trial type, PC, and window remained significant $F(1, 60) = 16.05$, $p < .001$, $\eta_p^2 = .21$, $BF_{01} = .028$ and there was still no four-way interaction $F(1, 60) = 1.45$, $p = .234$, $\eta_p^2 = .02$, $BF_{01} = 3.648$. Looking at the long and short delay conditions separately, the three-way interaction between trial type, PC, and window remained significant in the short delay, $F(1, 60) = 5.51$, $p = .022$, $\eta_p^2 = .08$, $BF_{01} = 2.188$, and long delay, $F(1, 60) = 11.94$, $p = 001$, $\eta_p^2 = .17$, $BF_{01} = .058$, conditions.

As in Experiments 1 and 2, we performed the exploratory ANOVA to see if the effects differed between the First Half and the Second Half[8] of the diagnostic phase. Here, we ran a five-way repeated measures ANOVA with factors of PC (MC or MI), Window (Shifted or Unshifted), Delay (Long or Short), Trial Type (Congruent or Incongruent), and Half (First Half and Second Half). There was no five-way interaction $F < 1$, $BF_{01} = 12.592$ nor was there a four-way interaction with Half, PC, Window and Trial Type, $F(1, 60) = 1.82$, $p < .182$, $\eta_p^2 = .03$, $BF_{01} = 3.120$.

To explore these effects further, we compared performance in the First Half and Second Half conditions in separate four-way repeated measures ANOVAs. In the First Half, there was a significant three-way interaction between PC, Window and Trial Type, $F(1, 60) = 5.56$, $p = .022$, $\eta_p^2 = .09$, $BF_{01} = 1.227$, but there was no four-way interaction between Delay, PC, Window, and Trial Type ($F < 1$, $BF_{01} = 5.113$). The three-way interaction between PC, Window and Trial Type, $F(1, 60) = 10.18$, $p = .002$, $\eta_p^2 = .15$, $BF_{01} = .018$, was also significant in the Second Half. Again, there was no four-way interaction between Delay, PC, Window, and Trial Type ($F < 1$, $BF_{01} = 5.721$).

Finally, for greater comparability to the analyses performed in Experiment 1, we also compared the halves in the short and long delay conditions separately. In the short delay condition, there was no significant four-way interaction between Window, PC, Trial Type, and Half, $F < 1$, $BF_{01} = 4.270$. In the long delay condition, there was no significant four-way interaction between Window, PC, Trial Type, and Half, $F < 1$, $BF_{01} = 3.179$.

Error Rate

Induction items.—In the induction, congruent trials ($M = 0.21\%$, $SE = 0.08\%$) were responded to more accurately than incongruent trials ($M = 4.66\%$, $SE = 0.64\%$), $F(1, 60) = 70.35$, $p < .001$, $\eta_p^2 = .54$, $BF_{01} < .001$. Additionally, there was a significant interaction between trial type and PC, such that the Stroop effect was larger in MC conditions ($M = 7.48\%$, $SE = 0.96\%$) than MI conditions ($M = 3.79\%$, $SE = 0.48\%$) $F(1, 60) = 90.95$, $p < .001$, $\eta_p^2 = .60$, $BF_{01} < .001$. There was no interaction between trial type, PC, and window, $F < 1$, $BF_{01} = 5.130$. Finally, there was no four-way interaction between trial type, PC, window, and delay, $F < 1$, $BF_{01} = 2.300$.

[8]The same limitations noted for the exploratory analyses in Experiment 1 hold true for Experiment 3. The percentages of congruent trials in the first half of the diagnostic phase in each condition were as follows: $MC_{SHIFTED}$: 47%; $MC_{UNSHIFTED}$: 44%; $MI_{SHIFTED}$: 44%; $MI_{UNSHIFTED}$: 53%. For each condition, the percentage of Diagnostic Set trials replaced in the first and second half respectively was: $MC_{SHIFTED}$: 16%, 34%; $MC_{UNSHIFTED}$: 19%, 31%; $MI_{SHIFTED}$: 13%, 38%; $MI_{UNSHIFTED}$: 16%, 34%. Thus, analysis of the second half was based on fewer trials. For descriptive statistics for the first and second halves in both reaction time and error rate, see supplementary Table 7.

Diagnostic items.—The 2 × 2 × 2 × 2 ANOVA revealed a significant effect of trial type, such that congruent trials (M = 0.26%, SE = 0.13%) were responded to more accurately than incongruent trials (M = 3.80%, SE = 0.74%), $F(1, 60) = 39.50$, $p < .001$, $\eta_p^2 = .40$, $BF_{01} < .001$. There was no interaction between trial type and PC, $F < 1$, $BF_{01} = 8.797$. Additionally, the three-way interaction between trial type, PC, and window, $F(1, 60) = 1.20$, $p = .278$, $\eta_p^2 = .02$, $BF_{01} = 5.973$, and the four-way interaction between trial type, PC, window, and delay length, $F < 1$, $BF_{01} = 4.473$ were not significant.

For completeness, we ran separate 2 × 2 × 2 ANOVAs for the long and short conditions, with factors of trial type, PC, and window, as well as the 2 × 2 × 2 × 2 ANOVA after removing the first trial of the diagnostic phase. No results meaningfully changed in error rate.

We again performed the exploratory five-way repeated measures ANOVA with factors of PC (MC or MI), Window (Shifted or Unshifted), Delay (Long or Short), Trial Type (Congruent or Incongruent), and Half (First Half and Second Half). There was no five-way interaction between PC, Window, Delay, Trial Type, and Half, $F(1, 60) = 2.30$, $p = .135$, $\eta_p^2 = .04$, $BF_{01} = 1.347$. There was no four-way interaction between PC, Window, Delay, and Trial Type $F < 1$, $BF_{01} = 8.259$.

To explore these effects further, the First Half and Second Half conditions were examined in separate four-way repeated measures ANOVAs. In the First Half, there was no four-way interaction between Delay, PC, Window, and Trial Type ($F < 1$, $BF_{01} = 4.672$), nor was there a significant three-way interaction between PC, Window and Trial Type, $F < 1$, $BF_{01} = 6.659$. In the Second Half, there was no four-way interaction between Delay, PC, Window, and Trial Type ($F(1, 60) = 2.59$, $p = .113$, $\eta_p^2 = .04$, $BF_{01} = 5.131$), nor was there was a significant three-way interaction between PC, Window and Trial Type, $F < 1$, $BF_{01} = 5.258$.

For greater comparability to the analysis performed in Experiment 1, we compared the halves in the short and long delay conditions separately. In the short delay condition, there was no significant four-way interaction between Window, PC, Trial Type, and Half, $F < 1$, $BF_{01} = 2.423$. In the long delay condition, there was no significant four-way interaction between Window, PC, Trial Type, and Half, $F < 1$, $BF_{01} = 3.584$.

Discussion

There were two key findings in Experiment 3. First, Experiment 3 systematically replicated the primary finding in Experiment 1 of differing Stroop effects on diagnostic trials for lists that were matched on the long timescale but differed in the recent timescale. Stroop effects were smaller following an $MC_{SHIFTED}$ induction compared to an $MC_{UNSHIFTED}$ induction and larger following an $MI_{SHIFTED}$ induction compared to an $MI_{UNSHIFTED}$ induction. Additionally, these patterns held after controlling for adjustments on the immediate timescale. These findings support our first prediction and reinforce our conclusion that conflict experiences in the recent timescale uniquely affect cognitive control. As in Experiment 1, we did not find a stronger effect of the recent timescale in the MC lists

compared to the MI lists, if anything, it leaned in the opposite direction (see General Discussion for further discussion).

The second key finding was a novel finding demonstrating that the effects of recent conflict on cognitive control in the diagnostic phase were evidenced not only in the short delay condition, which approximated Experiment 1, but additionally in the long delay condition. In other words, even when 4000 ms elapsed between the end of the induction and the presentation of the first trial in the diagnostic phase, control adjustments produced by the recent timescale affected performance in the diagnostic phase. This result is strikingly different from the adjustments produced by the immediate timescale, as evidenced by the effects of delay length on the congruency sequence effect (e.g., Duthoo et al., 2014; Egner et al., 2010). In those studies, the effects of previous conflict were not significant for delays of 3000 ms or longer. The only exception was Experiment 2 in Duthoo et al. (2014) wherein the researchers biased participants to use proactive control rather than reactive control by disproportionately presenting long delay trials; in that case a congruency sequence effect was observed for a 3000 ms delay. Experiment 3 therefore supports our third prediction and provides initial experimental evidence that conflict in the recent and immediate timescales have distinct effects, such that the recent timescale triggers more sustained (proactive) adjustments than that of the immediate timescale.

Importantly, the significant interaction between window, PC, and trial type remained after removing the first trial of the diagnostic phase. This result provides further evidence that the recent timescale effects are distinct from immediate timescale effects and that the adjustments following recent conflict experiences are sustained rather than transient. Consider the length of the delay between the participant's response to the last trial of the induction and the second trial in the long delay condition: after a response is produced for the final trial of the induction, the experimenter coding and RSI for that trial would occur (which together was 1000 ms), the delay would occur (3000 ms), and the participant would respond to the first trial of the diagnostic phase (the average response length was 677 ms). Then, the experimenter coding and RSI would occur for the first trial (1000 ms), and then the onset of the second trial would occur. The time between the offset of the last trial of the induction phase and the onset of the second trial of the diagnostic phase was thus 5677 ms on average. This is a substantial interval over which to still see an effect of previous conflict, and it would be unprecedented if the effect was driven by the immediate timescale (Egner et al., 2010; Duthoo et al., 2014). Furthermore, consider that the analysis of diagnostic trial performance minus the first trial represents an aggregation of Trials 2 through 8. The length of each subsequent trial following Trial 2 is an average of 1677 ms after the preceding trial. By Trial 8, an average of 15,739 ms had elapsed since the induction. The fact that the exploratory analysis indicated that the effects of the recent timescale were evident in both the first and second halves of the diagnostic phase, regardless of the delay length, also suggests a clearly sustained effect.

A novel theoretical implication of the Experiment 3 findings regards our use of the unfilled gap in the long delay condition. The recent conflict manipulation still affected performance in the diagnostic phase in the long delay condition even though the 4,000 ms delay was unfilled. This suggests that the control adjustments (i.e., relatively focused or relaxed)

persisted across this delay in the notable absence of bottom-up support, that is, when participants were not continuously responding to trials with goal-relevant features (as occurs during the 8-trial diagnostic phase). Consistent with the arguments set out in Duthoo et al. (2014), this aligns with the notion that some form of proactive control was active during the delay. In short, this persistence across the long delay demonstrates that effects of recent conflict produce adjustments that are also sustained in the sense that they continue to operate without any bottom up support for maintenance of control. All things considered, the findings of Experiment 3 provide striking evidence that the effect of recent conflict is sustained over many seconds and intervening trials, supporting the view that the recent and immediate timescales produce qualitatively different adjustments affecting subsequent performance.

General Discussion

While most previous research investigating the effects of conflict experiences on cognitive control has examined effects of the previous trial (immediate timescale) or effects of the entire block (long timescale), we aimed to examine the effects of relatively recent conflict experience (recent timescale). To that end, two overarching questions were addressed: what evidence exists for a recent timescale on its own or in interaction with other timescales, and are effects of the recent timescale relatively transient or sustained? First, we summarize the evidence for each of these questions, and then we discuss theoretical implications, limitations, and future directions.

The first overarching goal was to further understand whether conflict experiences in the recent timescale affect cognitive control (cf. Durston et al., 2003; Horga et al., 2011; Jiménez & Méndez, 2013; Jiménez & Méndez, 2014) by using an experimental approach involving diagnostic items and to examine whether the effect of conflict experiences in the recent timescale depends on other timescales (cf. Aben et al., 2017; Dey & Bugg, 2020). Controlling for conflict experiences in the long and immediate timescales, Experiment 1 found a unique effect of the recent timescale on cognitive control (consistent with our first prediction), as indicated by performance on the diagnostic items. In MC lists, presenting a run of four incongruent trials at the end of the induction led to an attenuated Stroop effect in the diagnostic phase; in MI lists, presenting four congruent trials at the end of the induction led to a larger Stroop effect in the diagnostic phase. Experiment 3 systematically replicated these findings, further supporting the first prediction. In contrast to Experiments 1 and 3, Experiment 2 used an unbiased pre-window section of the induction before the recent conflict manipulation. In this case, the recent timescale did not affect cognitive control in the diagnostic phase.

In contrast to the findings from statistical models examining timescales of control (Aben et al., 2017; Dey & Bugg, 2020) and our second prediction regarding the interaction of the recent timescale with other timescales, the effect of the recent timescale was not stronger in MC lists than MI lists in the current study. In Experiment 1, the effect of recent conflict was comparable for MC and MI lists while in Experiment 3, the effect size was nominally *smaller* in MC than MI lists. Overall, it can be concluded that there is evidence for a unique effect of conflict experiences in the recent timescale on cognitive control that

cannot be accounted for by the immediate or long timescale. In addition, while effects of recent conflict may be similar for MC and MI lists, that is not to say that the long timescale is irrelevant. The extent to which conflict experiences in the recent timescale affect cognitive control may depend on the degree to which such experiences differ from (i.e., defy) preceding conflict experience and not exclusively on the degree of conflict within the recent timescale (as suggested by the pattern of results in Experiments 1 and 3 compared to Experiment 2).

The second overarching goal was to understand whether effects of the recent timescale are relatively transient or sustained. In comparison to previous studies that examined recent conflict experiences and not just the immediately preceding trial (Durston et al., 2003; Horga et al., 2011; Jiménez & Méndez, 2013; Jiménez & Méndez, 2014), the current experiments lengthened the diagnostic scope beyond a single subsequent trial to a phase of eight trials. In Experiments 1 and 3, the recent conflict manipulation affected performance in the 8-trial diagnostic phase and the effect remained significant after removing the first trial of the diagnostic phase indicating that the effect of the recent timescale was not merely an effect of the immediate timescale. Additionally providing evidence of a sustained effect were the exploratory analyses examining adjustments in control within each half of the diagnostic phase, although they produced somewhat different patterns across these experiments (i.e., in Experiment 1, a “short delay” scenario, the predicted effects of recent conflict were observed in the first half but not the second half, whereas in Experiment 3, they were observed in both halves regardless of the delay possibly because the inclusion of the long delay condition may have biased participants to adopt a more proactive approach across all lists compared to the lists in Experiment 1; cf. Duthoo et al., 2014, where disproportionately presenting long delay trials biased adoption of proactive control). Overall, these findings provided initial evidence showing that the effect of recent conflict leads to relatively sustained adjustments in control.

Experiment 3 further tested this possibility by contrasting effects of recent conflict across two conditions: a short delay condition in which 1000 ms elapsed between the induction (last trial of window) and diagnostic phase, and a long delay condition in which 4000 ms elapsed. Critically, the effects of recent conflict were observed for both conditions—in fact the effects in the long delay condition were just as strong as and nominally stronger than the effects in the short delay condition—and both survived the analysis that removed the first trial of the diagnostic phase in both conditions. The effects of recent conflict in the 4000 ms condition supported our third prediction and provided strong and converging evidence indicating that cognitive control adjustments following recent conflict are sustained, in this case in the sense that they persist even when there is no bottom up support to maintain the control settings (i.e., in the unfilled delay). These findings stand in striking contrast to prior findings demonstrating transient effects of the immediate timescale on cognitive control (Egner et al., 2010; c.f. Duthoo et al., 2014). In those studies, effects of the immediate timescale did not survive unfilled delays longer than 2000 ms, suggesting such effects are transient.

Theoretical Implications

All three experiments in the current study manipulated the experience of conflict in the recent timescale via the same four trial window of varying conflict experiences. One key question is how to interpret the pattern whereby the Stroop effect in the diagnostic phase depended on the recent timescale in Experiments 1 and 3 but not Experiment 2. One way to interpret this is that effects of the recent timescale were found when the comparison lists (e.g., $MC_{UNSHIFTED}$ vs. $MC_{SHIFTED}$ lists in Experiments 1 and 3) differed in their conflict experiences early in the list (pre-window) but not when the comparison lists were equated in the pre-window (lists of Experiment 2). In other words, it might be suggested that pre-window differences account for what we have been referring to as effects of recent conflict (i.e., variation in the diagnostic phase performance). While this could explain the differences between Experiments 1 and 3 and Experiment 2, it is inconsistent with the overall patterns in Experiments 1 and 3. If performance in the diagnostic phase in these experiments was driven by conflict experiences in the pre-window, then $MC_{SHIFTED}$ lists (which had a stronger MC bias in the pre-window) should have produced bigger Stroop effects than $MC_{UNSHIFTED}$ lists, and $MI_{SHIFTED}$ lists (which had a smaller MI bias in the pre-window) should have produced smaller Stroop effects than $MI_{UNSHIFTED}$ lists. This is opposite to what we found.

Another possible interpretation, as alluded to earlier, is that the effects of the recent timescale depend on the degree to which conflict experiences deviate from preceding conflict experience. For example, the deviation may affect whether a prediction error occurs (e.g., den Ouden, Kok, & de Lange, 2012; see also Chiu, Jiang, & Egner, 2017 and Jiang, Heller, & Egner, 2014 for models that have incorporated prediction error), and consequently whether information in the recent timescale is weighted more heavily than the long timescale (overall accumulation of conflict experiences; see Supplementary Materials for an exploratory analysis examining a “degree of defiance” interpretation). An alternative interpretation draws on the concept of *learning rate*, or the degree to which new information is weighted when updating attentional settings (Behrens, Woolrich, Walton, & Rushworth, 2007). Some models (e.g., conflict monitoring account; see Botvinick et al., 2001) assume a fixed learning rate whereas other models assume a variable learning rate (Jiang, Heller, & Egner, 2014). Models with a fixed learning rate do not accommodate the present findings because the recent timescale did not have the same effect on subsequent trials regardless of conflict experiences preceding the recent timescale, as a fixed learning rate model anticipates. Effects of the recent timescale were observed only in two of the three experiments and depended on the experience preceding the recent timescale.

Turning to a model with a variable learning rate, the volatility model incorporates flexible changes in learning rate as a function of volatility (i.e., the likelihood that conflict is relatively consistent or fluctuates over the course of several trials; Jiang et al., 2014). When volatility is high, the learning rate is also high and accordingly, trials occurring more recently are weighted more strongly when informing whether attention should be heightened or relaxed. When volatility is low, the learning rate is low and a larger window of (preceding) trials (not just recent trials) is used to inform cognitive control adjustments.

Interpreting the current results from the volatility model's perspective, Experiments 1 and 3 were considerably volatile on a list level. In an $MC_{SHIFTED}$ list, for example, participants shifted from an MC bias in the pre-window to 0% congruent in the window, to unbiased in the diagnostic phase. Experiments 1 and 3 were also volatile on an experimental level, as MC and MI lists were randomly intermixed. Experiment 2 was less volatile on a list-level. For example, in a list with a 0% congruent (i.e., 100% incongruent) window, participants shifted from an unbiased pre-window, to 0% congruent in the window, to unbiased in the diagnostic phase. Experiment 2 may also have been less volatile on an experimental level (most lists were unbiased or weakly biased). If learning rate increases with increased volatility and a higher learning rate leads to a stronger weighting of more recent trials (Jiang, Beck, Heller, & Egner, 2015), then Experiments 1 and 3 should have been more likely to yield an effect of recent conflict, as was observed.

Another key question that emerges from the current study is the significance of our finding that effects of recent conflict on cognitive control were not larger in MC lists than MI lists, as previously evidenced in a statistical model (Aben et al., 2017; see also Dey & Bugg, 2020). In Experiment 1, the effect of recent conflict was equivalent for MC and MI lists (η_p^2 = .13 and .13, respectively) while in Experiment 3, the effect size was nominally though not statistically smaller in MC than MI lists (η_p^2 = .11 and .18, respectively)[9]. This was true for both delay conditions as well. The effect of recent conflict was nominally though not statistically smaller for MC than MI lists in the short delay condition (η_p^2 = .04 and .07, respectively) and in the long delay condition (η_p^2 = .07 and .15, respectively)[10].

Given the multitude of design differences between the current study and the statistical modeling studies (e.g., experimental approach vs. correlational approach; abbreviated vs. long lists; recent timescale conceptualization; presence of diagnostic trials [not included in Aben et al., but included in Dey & Bugg, 2020]; placement of diagnostic trials [intermixed with induction in statistical modeling studies but presented separately in current]), it is not possible to pinpoint the cause of the difference. However, it is notable that the volatility model can also accommodate this finding. From a volatility perspective, our MC and MI lists were equivalently volatile and resided in the same experimental context, and thus should have produced equal effects of recent conflict. Possibly, in the statistical modeling studies that used long MC and MI lists with randomly distributed trials (as compared to the current approach of controlling for conflict experiences in the different timescales), there were volatility differences between these lists that led to differences in the degree to which recent conflict influenced cognitive control. Interestingly, Aben et al. (2017) did find that in

[9]Following the suggestion of an anonymous reviewer, whom we thank, we assessed whether the effect of recent conflict was similar for MC and MI conditions in Experiments 1 and 3 by comparing the magnitude of the difference between shifted and unshifted lists for each condition. The three-way interaction did not address this question, as the interaction was a cross-over. Therefore, to compare the magnitudes, we reversed the sign of ($MC_{SHIFTED} - MC_{UNSHIFTED}$) and compared that to ($MI_{SHIFTED} - MI_{UNSHIFTED}$), since the direction of the difference is irrelevant to this question. In Experiment 1, there was not a difference between the (reversed) MC (M = 15, SE = 5) and MI (M = 17, SE = 6) conditions, $t(60)$ =0.15, p =.876, d =.02, as expected given the identical effect sizes. In Experiment 3, though the effect sizes were nominally different, there was again no difference between the (reversed) MC (M = 13, SE = 5) and MI (M = 22, SE = 6) conditions, $t(60)$ = 1.22, p =.227, d =.15.

[10]As in Footnote 9, we compared the magnitude of the difference between shifted and unshifted lists for the MC and MI conditions in each delay condition following the same procedure. In the short conditions, there was no difference between the (reversed) MC (M = 11, SE = 7) and MI (M = 16, SE = 8) conditions, $t(60)$ = 0.60, p =.550, d =.07. In the long delay condition, there was no difference between the (reversed) MC (M = 16, SE = 8) and MI (M = 27, SE = 8) conditions, $t(60)$ = 1.08, p =.286, d =.14.

a volatile but 50% condition in which the PC of the list varied every 20 trials between 80% and 20% congruent, recent conflict experiences strongly influenced performance. Possibly MC lists in their study were also more volatile than MI lists. Additional research is needed to further understand the roles that PC and volatility play in how much the recent timescale is weighted.

Turning to a different pattern that also has theoretical implications, one may be surprised that no difference was observed between $MC_{UNSHIFTED}$ and $MI_{UNSHIFTED}$ lists in the diagnostic phase of Experiments 1 and 3 considering that the typical list-wide PC pattern is a larger Stroop effect in an MC list than an MI list. This raises the question as to whether our findings regarding the unique effects of the recent timescale may reflect sensitivity to change (the shift within the window representing the recent timescale) under conditions where participants have not had enough time to acquire the list-wide PC (i.e., information about the long timescale)[11]. Several observations counter this notion. First, $MC_{UNSHIFTED}$ and $MI_{UNSHIFTED}$ lists produced markedly different Stroop effects in the induction phase (113 ms vs. 70 ms, respectively, in Experiment 1; 126 ms vs. 76 ms in Experiment 3), indicating a list-wide PC effect. Second, prior studies have demonstrated list-wide PC effects on diagnostic trials following inductions that were even shorter than the present 18-trial induction. Cohen-Shikora et al. (2018) showed larger Stroop effects for diagnostic trials in MC lists compared to MI lists following a six-trial induction. Notably, in that study diagnostic trials were intermixed with inducer trials as is the dominant approach in list-wide PC studies (including those comprising typical block lengths of ~100 trials; see Bugg, 2014), rather than presented in a separate phase at the end of the list. That brings us to the third observation. We conducted an exploratory analysis on the two diagnostic set items that were transplanted into the pre-window section of the induction phase (i.e., and thus intermixed with inducer trials as in a typical list-wide PC design) and we found that the Stroop effect for those trials was modulated by the PC of the pre-window.[12] In Experiment 1, this modulation followed the typical direction of a list-wide PC effect but the difference between $MC_{UNSHIFTED}$ and $MI_{UNSHIFTED}$ lists was not significant, possibly due to too few observations. However, in Experiment 3, there was a significant list-wide PC effect for the unshifted conditions with a larger Stroop effect for the $MC_{UNSHIFTED}$ than the $MI_{UNSHIFTED}$ lists. Critically, this analysis suggests that the induction was sufficiently long

[11]We thank Luis Jiménez for raising this point.

[12]This exploratory analysis assessed performance on Diagnostic Set trials that were integrated into the pre-window section of the induction for Experiments 1 and 3 to see if "typical" list-wide PC effects would be observed for these trials, most notably in unshifted lists. Note that there were only two trials per list, which should be taken into consideration when interpreting the results. The pre-window section of the induction differed in average PC across conditions ($MC_{SHIFTED}$ = 93%; $MC_{UNSHIFTED}$ = 72%; $MI_{UNSHIFTED}$ = 28%; $MI_{SHIFTED}$ = 7%). In Experiment 1, we ran a 2 × 2 × 2 repeated measures ANOVA with factors of Window, PC, and Trial Type. There was a significant three way interaction between PC, window, and trial type, $F(1, 60) = 4.72$, $p = .034$, $\eta_p^2 = 07$, $BF_{01} = .528$. For the unshifted lists, although the Stroop effect was larger for the Diagnostic Set trials in the MC lists ($MC_{UNSHIFTED}$: $M = 74$, $SE = 7$) than in the MI lists ($MI_{UNSHIFTED}$: $M = 63$, $SE = 7$), mirroring the list-wide PC pattern, the PC × Trial Type interaction was not significant, $F(1, 60) = 1.34$, $p = .251$, $\eta_p^2 = .02$, $BF_{01} = 3.376$. For the shifted lists, the PC × Trial Type interaction was significant, $F(1, 60) = 16.87$, $p < .001$, $\eta_p^2 = .22$, $BF_{01} = .005$, such that the Stroop effect was significantly larger in the MC lists ($MC_{SHIFTED}$: $M = 88$, $SE = 8$) than in the MI lists ($MI_{SHIFTED}$: $M = 44$, $SE = 7$). In Experiment 3, we ran a 2 × 2 × 2 × 2 repeated measures ANOVA with factors of PC, window, delay, and trial type, and there was a non-significant interaction between PC, window, and trial type, $F < 1$, $BF_{01} = 6.042$. This was because equivalent list-wide PC effects were found in the unshifted and shifted conditions as indicated by follow up contrasts. In the unshifted conditions, there was a significant list-wide PC × Trial Type interaction, $F(1, 60) = 10.23$, $p = .002$, $\eta_p^2 = 15$, $BF_{01} = .463$, such that the Stroop effect was larger for the Diagnostic Set trials in the MC lists ($MC_{UNSHIFTED}$: $M = 74$, $SE = 11$) than in the MI lists ($MI_{UNSHIFTED}$: $M = 48$, $SE = 11$). There was also a significant interaction in the shifted conditions, $F(1, 60) = 10.06$, $p = .002$, $\eta_p^2 = 14$, $BF_{01} = .027$, such that the Stroop effect was larger in the MC lists ($MC_{SHIFTED}$: $M = 81$, $SE = 11$) than in the MI lists ($MI_{SHIFTED}$: $M = 46$, $SE = 11$).

and powerful to produce a list-wide PC effect for both inducer items and diagnostic items in the induction phase, including in unshifted lists, suggesting participants did have enough experience to learn and adjust attention based on the list-wide PC in each list.

A provocative theoretical idea that emerges from the current findings is the possibility that list-wide PC effects are not a reflection of the (entire) long timescale but instead reflect the accumulation of recent conflict experiences. Previous work has demonstrated that list-wide PC effects are not just an accumulation of the effects of the immediate timescale (i.e., an accumulation of congruency-sequence effects; Torres-Quesada et al., 2013; Torres-Quesada et al., 2014; see also Meier & Kane, 2013). However, to our knowledge, no study to date has directly assessed whether list-wide PC effects may be an accumulation of adjustments based on the recent timescale. The effects of recent conflict on control in a subsequent diagnostic phase (differing Stroop effects) for lists matched in PC, as evidenced in Experiments 1 and 3, underscore this possibility. Indeed, it may be more plausible that list-wide PC effects are an accumulation of adjustments based on recent as opposed to immediate conflict experiences given the novel pattern that we observed in Experiment 3. Effects of the recent timescale, unlike the transient effects of the immediate timescale (e.g., Egner et al., 2010; Duthoo et al., 2014) were found to be sustained. List-wide PC effects, like effects of the recent timescale, are also thought to reflect sustained adjustments in control (e.g., De Pisapia & Braver, 2006; Aben et al., 2019).

In line with this possibility is a noteworthy pattern in Experiment 1, which was systematically replicated in Experiment 3, showing a “reversed” list-wide PC effect. That is, in the diagnostic phase, the Stroop effect was smaller for $MC_{SHIFTED}$ lists which were overall MC but ended on four incongruent trials than entirely MI lists (i.e., $MI_{UNSHIFTED}$); similarly, the Stroop effect was smaller for $MI_{SHIFTED}$ lists which were overall MI but ended in four congruent trials than entirely MC lists (i.e., $MC_{UNSHIFTED}$). In other words, a recent conflict experience with four trials drove subsequent adjustments in attention, and not the overall experience within the list. These patterns are quite unexpected when considering the list-wide PC literature, but they may speak to the relative strength of conflict in the recent timescale. Given the profound theoretical implications of challenging the prevailing view of list-wide PC effects, future research should seek converging evidence through modeling and experimentation to determine whether list-wide PC effects are driven by several preceding trials rather than whole lists. Novel procedures will be needed to fully tackle this theoretical possibility as the typical list-wide PC manipulation conflates conflict experiences in the long and recent timescales (e.g., both are likely MC in MC lists and MI in MI lists).

Limitations and Future Directions

Two major methodological differences should be considered when comparing the results of this study to previous research: use of the abbreviated lists paradigm and the conceptualization of recent conflict. We used an abbreviated-lists paradigm that enabled us to experimentally control for the long and recent timescales in each list, while obtaining multiple observations per individual in each condition. In contrast, the statistical modeling studies that provided initial evidence for asymmetric effects of the recent timescale in MC and MI lists (Aben et al., 2017; Dey & Bugg, 2020) were based on longer, randomly

generated lists (lists ranged from 160 to 480 trials). It is possible that effects of the long timescale (or the recent timescale) differ when effects of the long timescale are based on a much longer history of trials. Future research is needed to evaluate this possibility.

The current study also differed from prior research in its conceptualization of recent conflict experience as runs of four consecutive congruent or incongruent trials. We chose four trials because prior behavioral studies had investigated four previous trials as one of several conceptualizations of recent conflict (in addition to e.g., two, three previous trials; Durston et al., 2003; Horga et al., 2011; Jiménez & Méndez, 2013; Jiménez & Méndez, 2014) and prior modeling efforts found an effect of conflict at least four trials prior to the current trial (Aben et al., 2017; Dey & Bugg 2020). It might be argued, however, that manipulating the four trials to be entirely congruent (or entirely incongruent) represents an extreme case of a recent conflict experience. For example, in the lists that were statistically modeled in Aben et al. (2017; Dey & Bugg, 2020), it is possible that some runs of four consecutive congruent or incongruent trials occurred prior to trial n (the trial on which performance was predicted). More likely, given the random distribution of trials, the runs were a mix of congruent and incongruent trials. It will be valuable for future research to examine the effects of absolute window size (i.e., different window sizes for the manipulation of recent conflict in lists of the same length) and relative window size (i.e., a given size such as four trials presented in lists of differing lengths) as possible moderators of the effects of recent conflict on cognitive control.

Another direction for future research will be to better understand the consequences of assessing cognitive control in a “separate” diagnostic phase that follows the induction, as in the current experiments, versus intermixing diagnostic trials within the induction phase of the lists (see e.g., Bugg, 2014). We elected to use a separate phase because we wanted to experimentally control experiences within the diagnostic phase across lists, examine a diagnostic phase that was relatively extended (8 trials), and assess whether effects of recent conflict were transient or sustained. However, a finding in Experiment 1 that was also observed in Experiment 3 leads to the question of whether these two approaches to assessing control via diagnostic items may be capturing somewhat different effects. In particular, there was no difference in the Stroop effect between $MC_{UNSHIFTED}$ and $MI_{UNSHIFTED}$ lists in the diagnostic phase. This stands in contrast to prior findings showing that Stroop effects were larger in MC lists than MI lists on diagnostic items that were intermixed with inducer items (e.g., Bugg, 2014; Bugg & Chanani, 2011; Gonthier et al., 2016; see also Hutchison, 2011). It also stands in contrast to our exploratory analyses (see Footnote 12) where we found that Stroop effects on the two unbiased diagnostic items that were intermixed with inducer items in the pre-window section of the induction phase were larger in MC lists than MI lists, consistent with typical list-wide PC effects. Future research should consider whether different mechanisms may be contributing to list-wide PC effects depending on the placement of diagnostic items.

Conclusion

Using a systematic experimental approach, the current study demonstrated the unique effects of relatively recent conflict experiences on cognitive control. These effects could neither

be explained by conflict experiences in the long timescale or the immediate timescale. Additionally, a novel component of the current study was that the effects of recent conflict were assessed via unique diagnostic items that were presented in an 8-trial diagnostic phase, enabling us to draw conclusions about effects of recent conflict on cognitive control independent of lower-level processes. The use of an extended diagnostic phase further allowed us to initially surmise from the results of Experiment 1 that the recent timescale may affect control in a sustained fashion. This conclusion was significantly reinforced by the findings of Experiment 3, which showed that effects of the recent timescale on control sustain across a long unfilled delay (4000 ms), unlike effects of the immediate timescale (Egner et al., 2010; Duthoo et al., 2014). Collectively, the current findings suggest that relatively recent experiences do affect the control of attention and that these effects can be distinguished both experimentally and conceptually from the effects of the immediately preceding experience and the overall experience (long timescale). In general, this research points toward a need for theories and models of memory and attention to consider the important influence of relatively recent experiences. Doing so will lead to a better understanding of how previous conflict experiences are weighted and inform adjustments to cognitive control.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks to Eva Jeliazkova, and Janelli Rodriguez for their time and effort thinking about this project and to Janelli for her assistance with data collection. Thanks to Jihyun Suh for consultation on a statistical analysis. Thanks to the Cognitive Control and Aging Lab, Dave Balota, and Todd Braver for their helpful feedback on the project.

JMB was supported by NIH AG057937

References

Aben B, Calderon CB, Van der Cruyssen L, Picksak D, Van den Bussche E, & Verguts T (2019). Context-dependent modulation of cognitive control involves different temporal profiles of fronto-parietal activity. NeuroImage, 189, 755–762, doi: 10.1016/j.neuroimage.2019.02.004 [PubMed: 30735827]

Aben B, Verguts T, & Van den Bussche E (2017). Beyond trial-by-trial adaptation: A quantification of the time scale of cognitive control. Journal of Experimental Psychology: Human Perception and Performance, 43(3), 509, doi:10.1037/xhp0000324 [PubMed: 28080112]

Alexander WH, & Brown JW (2011). Medial prefrontal cortex as an action-outcome predictor. Nature Neuroscience, 14(10), 1338, doi:10.1038/nn.2921 [PubMed: 21926982]

Behrens TE, Woolrich MW, Walton ME, & Rushworth MF (2007). Learning the value of information in an uncertain world. Nature Neuroscience, 10(9), 1214, doi:10.1038/nn1954 [PubMed: 17676057]

Braver TS, Gray JR, & Burgess GC (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. Variation in working memory, 75, 106.

Braem S, Bugg JM, Schmidt JR, Crump MJ, Weissman DH, Notebaert W, & Egner T (2019). Measuring adaptive control in conflict tasks. Trends in Cognitive Sciences, doi: 10.1016/j.tics.2019.07.002.

Brown JW, & Braver TS (2005). Learned predictions of error likelihood in the anterior cingulate cortex. Science, 307(5712), 1118–1121, doi:10.1126/science.1105783 [PubMed: 15718473]

Bugg JM (2017). Context, conflict, and control. In Egner T (Ed.), The Wiley handbook of cognitive control (pp. 79–96). Chichester, UK: Wiley.

Bugg JM (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(2), 567 doi: 10.1037/a0035032

Bugg JM (2012). Dissociating levels of cognitive control: The case of Stroop interference. Current Directions in Psychological Science, 21(5), 302–309, doi: 10.1177/0963721412453586

Bugg JM, & Chanani S (2011). List-wide control is not entirely elusive: Evidence from picture–word Stroop. Psychonomic Bulletin & Review, 18(5), 930–936, doi:10.3758/s13423-011-0112-y [PubMed: 21638107]

Bugg JM, & Crump MJ (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. Frontiers in Psychology, 3, 367, doi:10.3389/fpsyg.2012.00367 [PubMed: 23060836]

Bugg JM, & Diede NT (2018). The effects of awareness and secondary task demands on Stroop performance in the pre-cued lists paradigm. Acta Psychologica, 189, 26–35, doi: 10.1016/j.actpsy.2016.12.013 [PubMed: 28061943]

Bugg JM, Diede NT, Cohen-Shikora ER, & Selmeczy D (2015). Expectations and experience: Dissociable bases for cognitive control? Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(5), 1349, doi:10.1037/xlm0000106

Botvinick MM, Braver TS, Barch DM, Carter CS, & Cohen JD (2001). Conflict monitoring and cognitive control. Psychological Review, 108(3), 624, doi:10.1037//0033-295X.108.3.624 [PubMed: 11488380]

Chiu YC, Jiang J, & Egner T (2017). The caudate nucleus mediates learning of stimulus–control state associations. Journal of Neuroscience, 37(4), 1028–1038, doi: 10.1523/JNEUROSCI.0778-16.2016 [PubMed: 28123033]

Cohen JD (2017). Cognitive Control: Core Constructions and Current Considerations. In Egner T (Ed.), The Wiley handbook of cognitive control. (pp. 17–41). Chichester, West Sussex, UK: John Wiley & Sons

Cohen-Shikora ER, Diede NT, & Bugg JM (2018). The flexibility of cognitive control: Age equivalence with experience guiding the way. Psychology and Aging, 33(6), 924, doi:10.1037/pag0000280 [PubMed: 30080058]

Crump MJ, & Milliken B (2009). The flexibility of context-specific control: Evidence for context-driven generalization of item-specific control settings. The Quarterly Journal of Experimental Psychology, 62(8), 1523–1532, doi:10.1080/17470210902752096 [PubMed: 19370487]

De Pisapia N, & Braver TS (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. Neurocomputing, 69(10–12), 1322–1326, doi: 10.1016/j.neucom.2005.12.100

Den Ouden HE, Kok P, & De Lange FP (2012). How prediction errors shape perception, attention, and motivation. Frontiers in Psychology, 3, 548, doi:10.3389/fpsyg.2012.00548 [PubMed: 23248610]

Dey A, Bugg JM (2020). The timescale of control: A meta-control property that generalizes across tasks but varies between types of control. Manuscript submitted for publication.

Durston S, Davidson MC, Thomas KM, Worden MS, Tottenham N, Martinez A, … & Casey BJ (2003). Parametric manipulation of conflict and response competition using rapid mixed-trial event-related fMRI. Neuroimage, 20(4), 2135–2141, doi:10.1016/j.neuroimage.2003.08.004 [PubMed: 14683717]

Duthoo W, Abrahamse EL, Braem S, Boehler CN, & Notebaert W (2014). The heterogeneous world of congruency sequence effects: An update. Frontiers in Psychology, 5, 1001, doi:10.3389/fpsyg.2014.01001 [PubMed: 25250005]

Duthoo W, Abrahamse EL, Braem S, & Notebaert W (2014). Going, going, gone? Proactive control prevents the congruency sequence effect from rapid decay. Psychological Research, 78(4), 483–493, doi:10.1007/s00426-013-0498-4 [PubMed: 24077774]

Egner T (2007). Congruency sequence effects and cognitive control. Cognitive, Affective, & Behavioral Neuroscience, 7(4), 380–390, doi:10.3758/CABN.7.4.380

Egner T, Ely S, & Grinband J (2010). Going, going, gone: characterizing the time-course of congruency sequence effects. Frontiers in psychology, 1, 154, doi:10.3389/fpsyg.2010.00154 [PubMed: 21833220]

Faul F, Erdfelder E, Lang AG, & Buchner A (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191, doi: 10.3758/BF03193146 [PubMed: 17695343]

Gonthier C, Braver TS, & Bugg JM (2016). Dissociating proactive and reactive control in the Stroop task. Memory & Cognition, 44(5), 778–788, doi:10.3758/s13421-016-0591-1 [PubMed: 26861210]

Gratton G, Coles MG, & Donchin E (1992). Optimizing the use of information: strategic control of activation of responses. Journal of Experimental Psychology: General, 121(4), 480, doi:10.1037/0096-3445.121.4.480 [PubMed: 1431740]

Horga G, Maia TV, Wang P, Wang Z, Marsh R, & Peterson BS (2011). Adaptation to conflict via context-driven anticipatory signals in the dorsomedial prefrontal cortex. Journal of Neuroscience, 31(45), 16208–16216, doi:10.1523/JNEUROSCI.2783-11.2011 [PubMed: 22072672]

Hutchison KA (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(4), 851, doi:10.1037/a0023437

Jiménez L, & Méndez A (2013). It is not what you expect: Dissociating conflict adaptation from expectancies in a Stroop task. Journal of Experimental Psychology: Human Perception and Performance, 39(1), 271, doi:10.1037/a0027734 [PubMed: 22428671]

Jiménez L, & Méndez A (2014). Even with time, conflict adaptation is not made of expectancies. Frontiers in Psychology, 5, 1042, doi:10.3389/fpsyg.2014.01042 [PubMed: 25278926]

Jiang J, Beck J, Heller K, & Egner T (2015). An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. Nature Communications, 6, 8165, doi:10.1038/ncomms9165

Jiang J, Heller K, & Egner T (2014). Bayesian modeling of flexible cognitive control. Neuroscience & Biobehavioral Reviews, 46, 30–43, doi:10.1016/j.neubiorev.2014.06.001 [PubMed: 24929218]

Logan GD, & Zbrodoff NJ (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. Memory & Cognition, 7(3), 166–174, doi:10.3758/BF03197535

Mayr U, Awh E, & Laurey P (2003). Conflict adaptation effects in the absence of executive control. Nature neuroscience, 6(5), 450–452, doi: 10.1038/nn1051 [PubMed: 12704394]

Meier ME, & Kane MJ (2013). Working memory capacity and Stroop interference: Global versus local indices of executive control. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(3), 748, doi:10.1037/a0029200

Schlaghecken F, & Martini P (2012). Context, not conflict, drives cognitive control. Journal of Experimental Psychology: Human Perception and Performance, 38(2), 272, doi:10.1037/a0025791 [PubMed: 21967273]

Schmidt JR, & 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(3), 514, doi:10.1037/0278-7393.34.3.514

Spinelli G, Perry JR, & Lupker SJ (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture–word interference task. Journal of Experimental Psychology: Human Perception and Performance, 45(8), 995–1014.10.1037/xhp0000656 [PubMed: 31144859]

Torres-Quesada M, Funes MJ, & Lupiáñez J (2013). Dissociating proportion congruent and conflict adaptation effects in a Simon–Stroop procedure. Acta Psychologica, 142(2), 203–210, doi:10.1016/j.actpsy.2012.11.015 [PubMed: 23337083]

Torres-Quesada M, Milliken B, Lupiáñez J, & Funes MJ (2014). Proportion Congruent effects in the absence of Sequential Congruent effects. Psicológica, 35(1), 101–115.

Wagenmakers EJ, Wetzels R, Borsboom D, & Van Der Maas HL (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). Journal of Personality and Social Psychology, 100(3), 426–432, doi:10.1037/a0022790 [PubMed: 21280965]

Weissman DH, Hawks ZW, & Egner T (2015). Different levels of learning interact to shape the congruency sequence effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(4), 566, doi: 10.1037/xlm0000182

Weissman DH, Jiang J, & Egner T (2014). Determinants of congruency sequence effects without learning and memory confounds. Journal of Experimental Psychology: Human Perception and Performance, 40(5), 2022, doi: 10.1037/a0037454 [PubMed: 25089574]

Public Significance Statement:

Previous research has demonstrated that experiences resolving conflict are stored in memory and affect whether our attention is currently focused or relaxed. Many prior studies have shown that our immediately preceding experience as well as the accumulation of all prior experiences affect the control of attention. However, there is a dearth of research regarding how *relatively* recent experiences, like the most recent handful, affect current attention. Our findings suggest that relatively recent experiences do affect the control of attention and that these effects can be distinguished both experimentally and conceptually from the effects of the immediately preceding experience and the overall experience. In general, this research points toward a need for theories and models of memory and attention to consider the important influence of relatively recent experiences on attention.

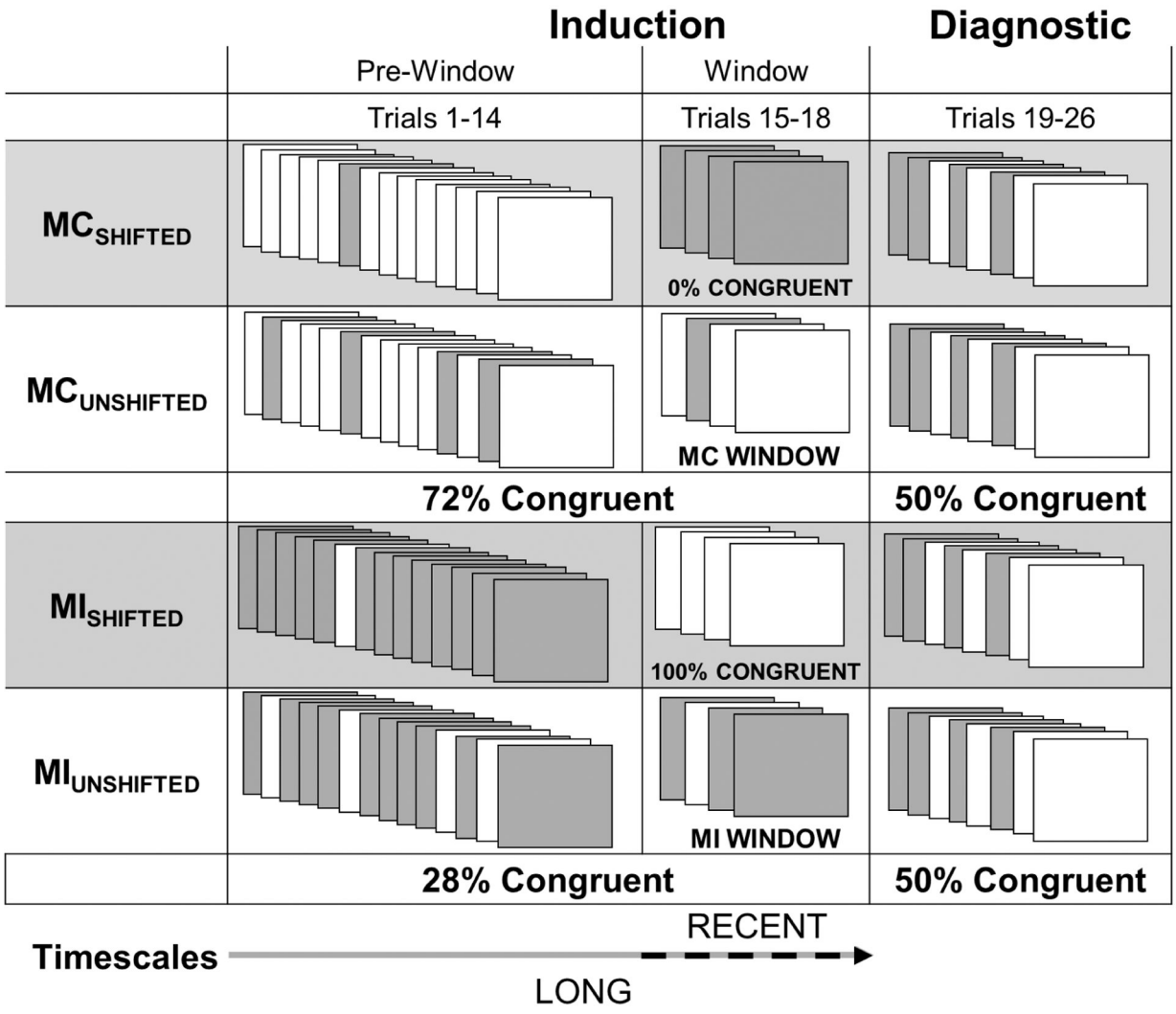

Figure 1.

List composition for Experiment 1. White squares represent congruent trials and gray squares represent incongruent trials. The induction phase refers to both the Pre-Window and Window segments. As illustrated at the bottom of the figure, the entire induction phase constitutes the long timescale, while the window within the induction phase constitutes the recent timescale. The upper two rows represent mostly congruent (MC) lists and the bottom two rows represent mostly incongruent (MI) lists. Within each set of rows, there are shifted and unshifted lists, which refers to the recent timescale manipulation as illustrated in the column labeled "Window". While lists that were contrasted (e.g., $MC_{SHIFTED}$ and $MC_{UNSHIFTED}$) were matched in the long timescale (i.e., equal number of congruent and incongruent trials for unshifted and shifted lists), the recent timescale differed such that the experience-defying trial type was presented on four consecutive trials in shifted lists only (e.g., the induction phase of the $MC_{SHIFTED}$ list ended with four incongruent trials). A diagnostic phase of eight trials (50% congruent) concluded each list.

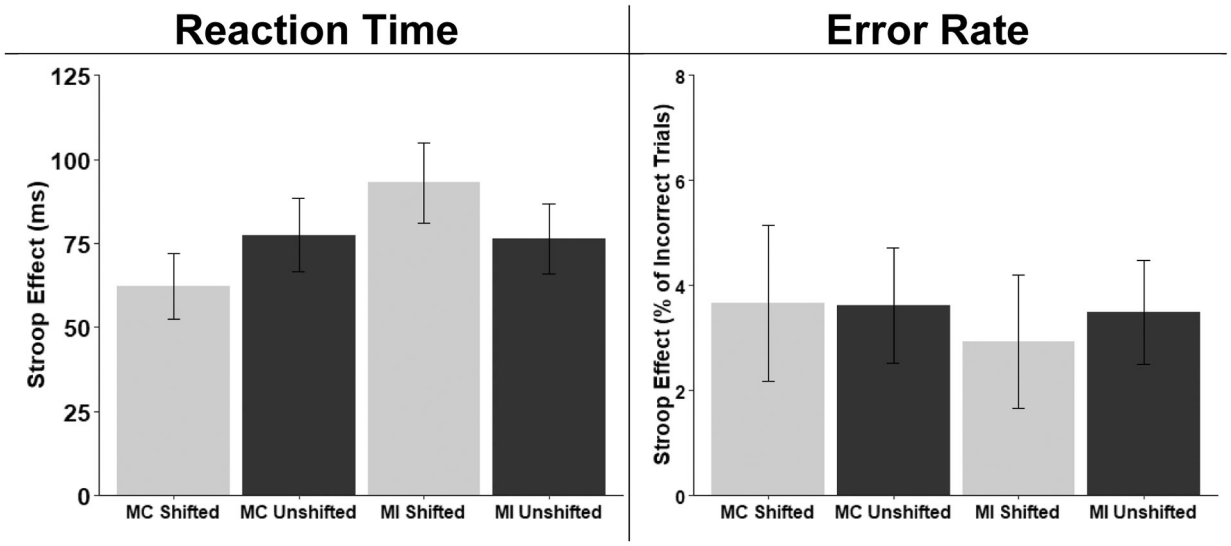

Figure 2.

Reaction time and error rate results for Experiment 1 diagnostic phase trials. Error bars represent a 95% confidence interval. In reaction time, a significant three-way interaction was observed between trial type, PC, and window. The manipulation of the recent window experience attenuated the Stroop effect in MC and exacerbated the Stroop effect in MI lists. In error rate, no three-way interaction between trial type, PC, and window was observed.

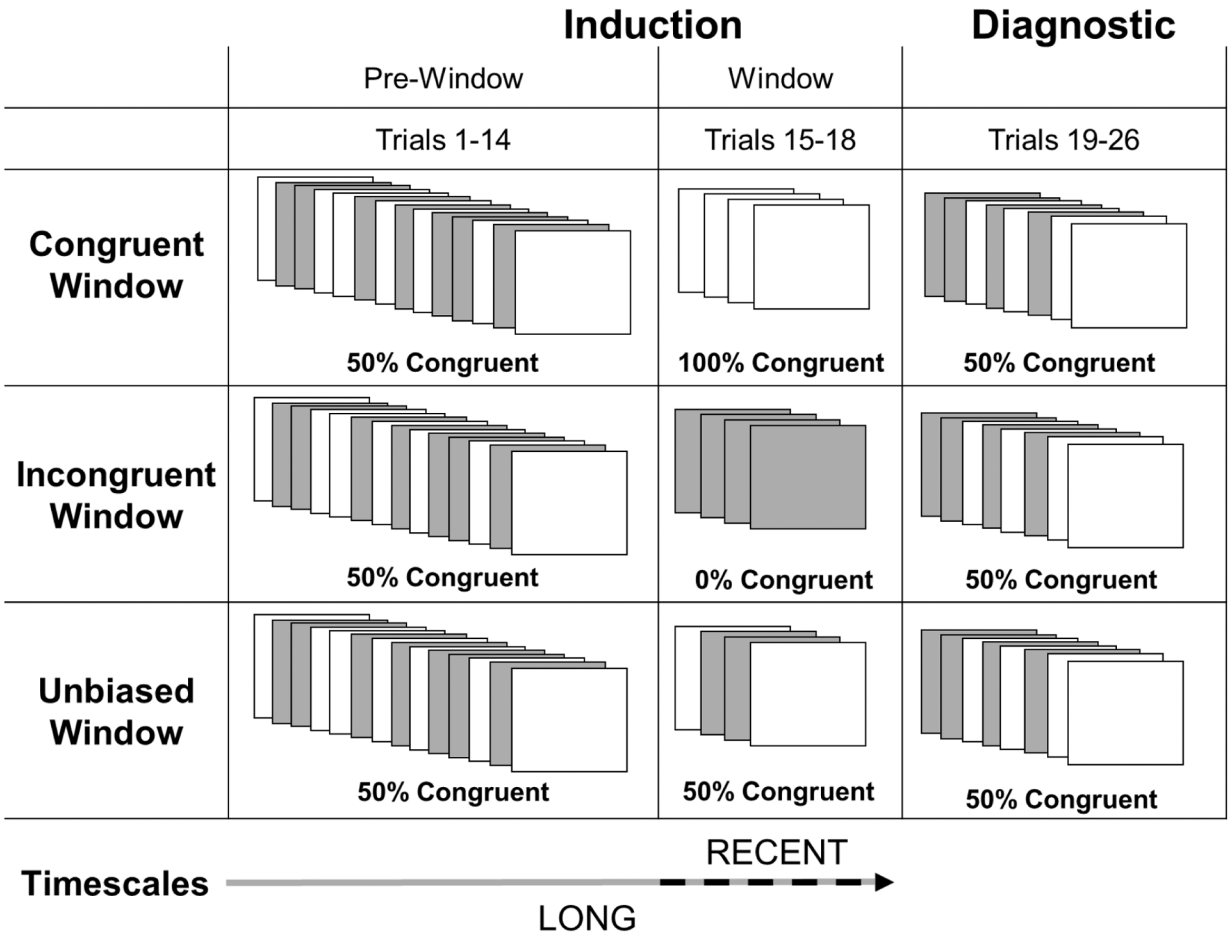

Figure 3.

List composition for Experiment 2. White squares represent congruent trials and gray squares represent incongruent trials. The induction phase refers to both the Pre-Window and Window segments. As illustrated at the bottom of the figure, the entire induction phase constitutes the long timescale, while just the window within the induction constitutes the recent timescale. For all list types (as illustrated in the rows), the Pre-Window section was 14 trials (50% congruent). The recent timescale manipulation is illustrated in the column labeled “Window”. Congruent Window refers to a window of entirely congruent trials, Incongruent Window refers to a window of entirely incongruent trials, Unbiased Window refers to a window with two congruent and two incongruent trials. The diagnostic phase of eight trials (50% congruent) concluded each list.

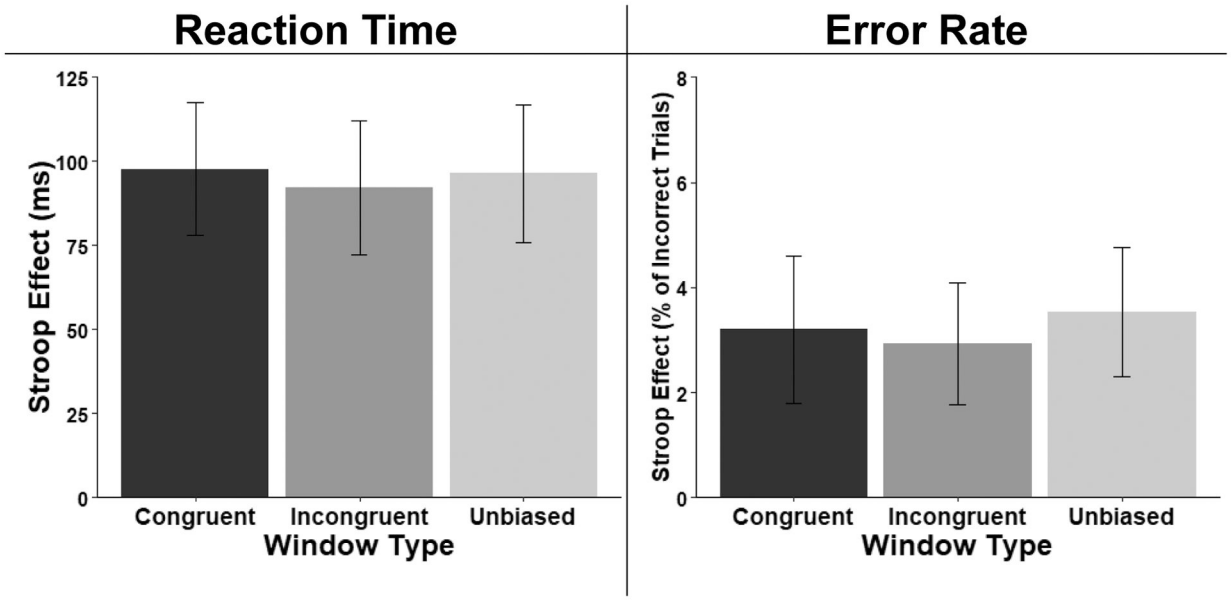

Figure 4.

Reaction time and error rate results for Experiment 2 diagnostic trials. Error bars represent a 95% confidence interval. The axis labels refer to each of the manipulations in the window: Congruent Window refers to a window of entirely congruent trials, Incongruent Window refers to a window of entirely incongruent trials, Unbiased Window refers to a window of two congruent and two incongruent trials. No interaction between trial type and window type was observed in reaction time. A significant interaction between trial type and window type was observed in error rate, such that the Stroop effect was larger following a congruent window. However, the interaction for error rate did not survive removing the first trial of the diagnostic phase.

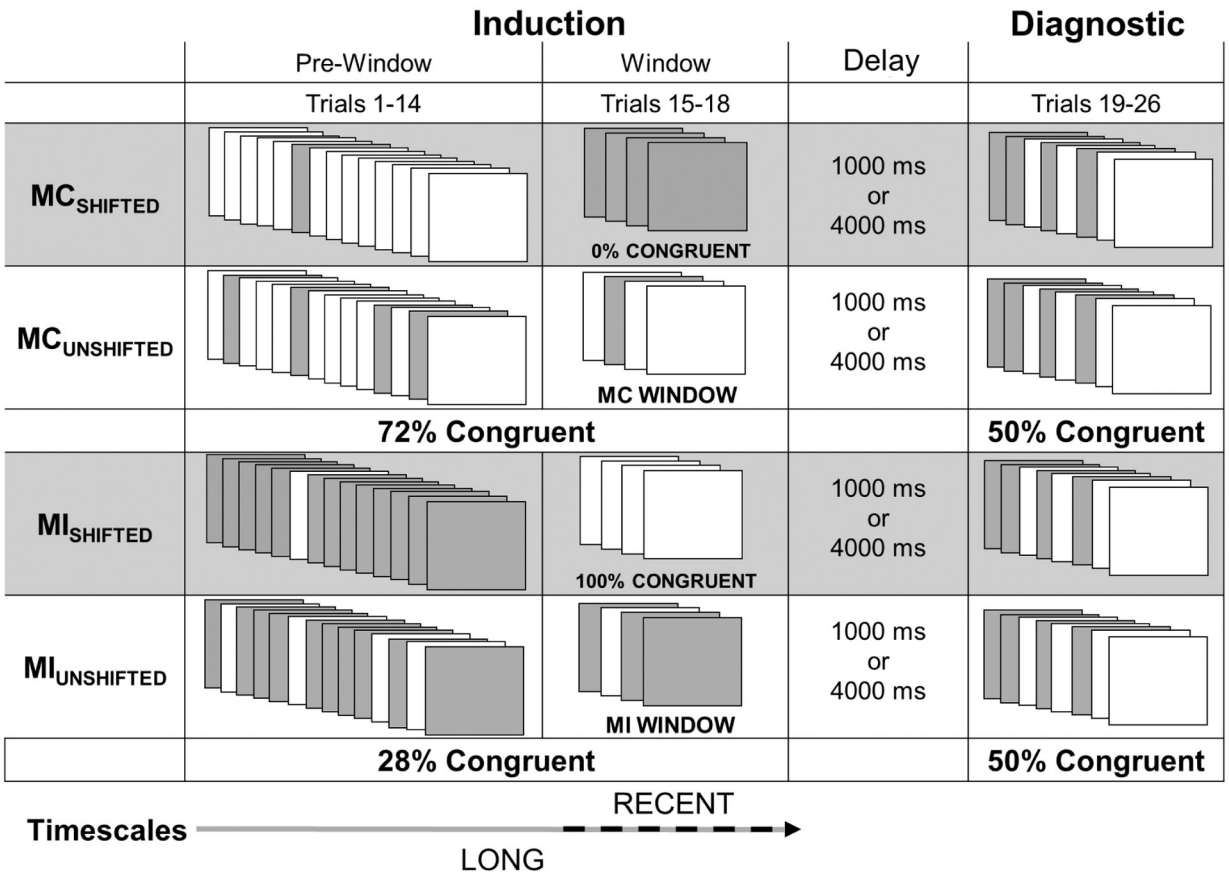

Figure 5.

List composition for Experiment 3. White squares represent congruent trials and gray squares represent incongruent trials. The induction phase refers to both the Pre-Window and Window segments. As illustrated at the bottom of the figure, the entire induction phase constitutes the long timescale, while the window within the induction phase constitutes the recent timescale. The upper two rows represent mostly congruent (MC) lists and the bottom two rows represent mostly incongruent (MI) lists. Within each set of rows, there are shifted and unshifted lists, which refers to the recent timescale manipulation as illustrated in the column labeled “Window”. The design for Experiment 3 was equivalent to Experiment 1, except for the delay between the induction and diagnostic phases. In the short delay condition, the delay was 1000 ms and in the long delay condition the delay was 4000 ms.

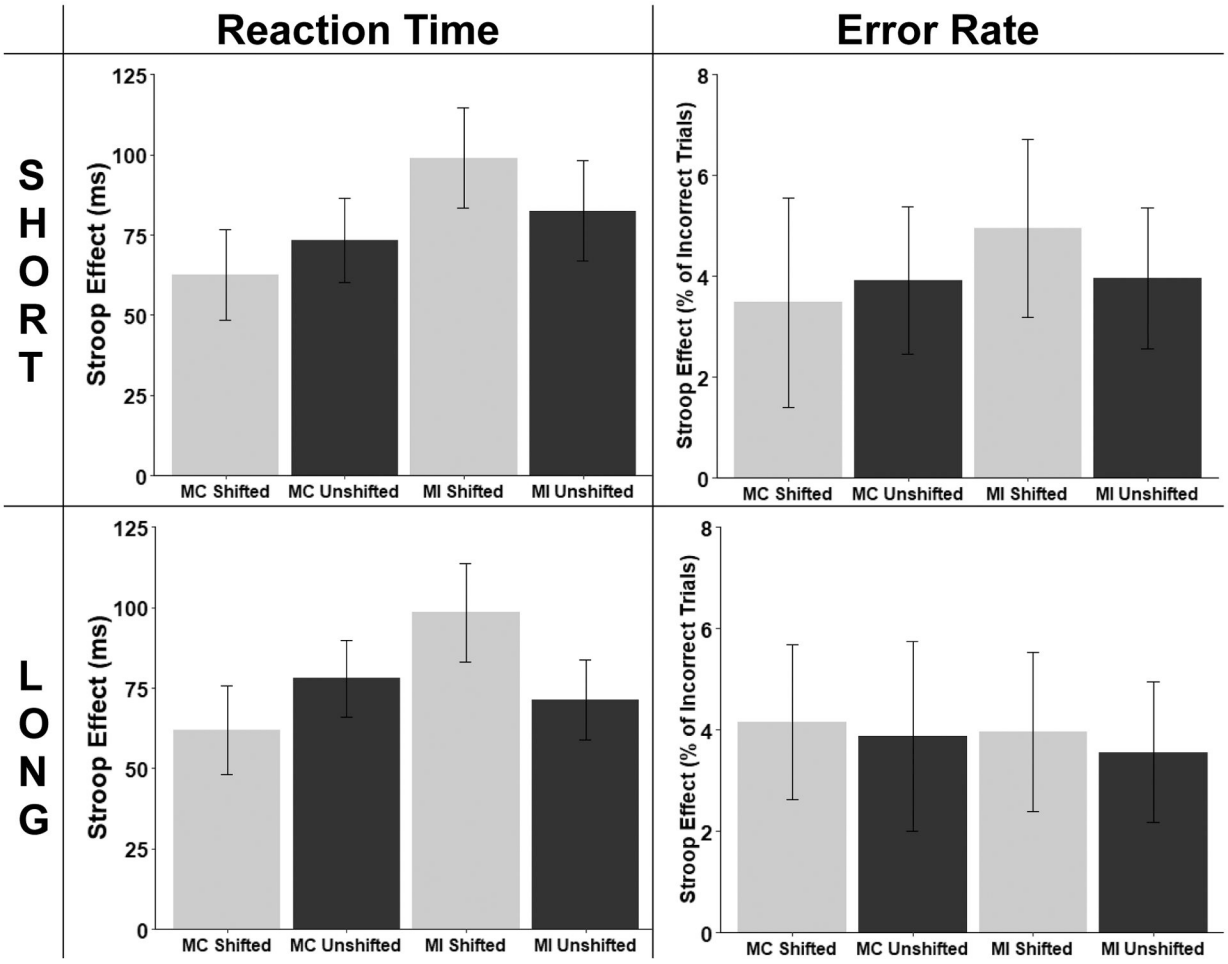

Figure 6.

Reaction time and error rate results for Experiment 3 diagnostic trials. Error bars represent a 95% confidence interval. Both the short and long delay conditions show a significant three-way interaction between trial type, PC, and window in reaction time and no three-way interaction between trial type, PC, and window for error rate in the diagnostic phase.

Table 1

Experiment 1 Reaction Time (ms) and Error Rate with Standard Deviations in Parentheses

Phase	PC	Window	Trial Type	Reaction Time	Error %
Induction	MC	Shifted	Congruent	570 (74)	0.51 (0.62)
			Incongruent	732 (98)	7.52 (8.00)
		Unshifted	Congruent	589 (79)	0.64 (0.80)
			Incongruent	703 (89)	5.59 (5.27)
	MI	Shifted	Congruent	620 (95)	0.56 (0.98)
			Incongruent	684 (90)	2.41 (2.73)
		Unshifted	Congruent	618 (93)	0.48 (1.06)
			Incongruent	688 (90)	3.30 (3.19)
Diagnostic	MC	Shifted	Congruent	636 (97)	0.74 (1.41)
			Incongruent	698 (100)	4.39 (5.58)
		Unshifted	Congruent	615 (85)	0.59 (1.32)
			Incongruent	693 (95)	4.20 (4.54)
	MI	Shifted	Congruent	607 (89)	0.69 (1.49)
			Incongruent	700 (102)	3.62 (5.45)
		Unshifted	Congruent	621 (90)	0.34 (1.11)
			Incongruent	697 (98)	3.82 (4.20)

Note: MC = mostly congruent; MI = mostly incongruent; Shifted = in the last four trials of the induction (the window) only the infrequent trial type (e.g., incongruent) was presented for a given condition (e.g., MC). Unshifted = the last four trials represented a continuation of the ongoing induction (i.e., were MC in an MC list)

Table 2

Experiment 2 Reaction Time (ms) and Error Rate with Standard Deviations in Parentheses

Phase	Window Type	Trial Type	Reaction Time	Error %
Induction	Congruent	Congruent	610 (96)	0.77 (1.08)
		Incongruent	714 (121)	3.57 (3.22)
	Incongruent	Congruent	614 (94)	0.62 (0.92)
		Incongruent	721 (115)	4.08 (3.71)
	Unbiased	Congruent	617 (95)	0.33 (0.63)
		Incongruent	722 (125)	4.37 (3.63)
Diagnostic	Congruent	Congruent	622 (104)	0.62 (1.36)
		Incongruent	720 (139)	5.43 (5.47)
	Incongruent	Congruent	637 (101)	0.50 (1.23)
		Incongruent	730 (138)	3.33 (4.51)
	Unbiased	Congruent	626 (89)	0.40 (1.17)
		Incongruent	723 (137)	3.78 (4.88)

Note: Congruent Window Type = the four trials comprising the window were all congruent. Incongruent Window Type = the four trials comprising the window were all incongruent. Unbiased Window Type = two of the four trials comprising the window were congruent and two were incongruent.

Table 3

Experiment 3 Induction Phase Reaction Time (ms) and Error Rate with Standard Deviations in Parentheses

Phase	Delay	PC	Window	Trial Type	Reaction Time	Error %
Induction	Short	MC	Shifted	Congruent	591 (62)	0.11 (0.41)
				Incongruent	742 (70)	7.93 (7.23)
			Unshifted	Congruent	608 (68)	0.15 (0.40)
				Incongruent	730 (75)	7.16 (6.86)
		MI	Shifted	Congruent	633 (70)	0.44 (1.40)
				Incongruent	716 (71)	3.28 (3.82)
			Unshifted	Congruent	643 (71)	0.11 (0.64)
				Incongruent	718 (73)	3.23 (3.91)
Induction	Long	MC	Shifted	Congruent	588 (62)	0.20 (0.60)
				Incongruent	747 (80)	6.93 (7.14)
			Unshifted	Congruent	610 (64)	0.29 (0.53)
				Incongruent	741 (88)	7.43 (8.60)
		MI	Shifted	Congruent	632 (68)	0.26 (0.89)
				Incongruent	713 (66)	2.69 (3.62)
			Unshifted	Congruent	639 (62)	0.31 (0.96)
				Incongruent	716 (69)	3.30 (3.96)

Note: MC = mostly congruent; MI = mostly incongruent; Shifted = in the last four trials of the induction (the window), only the infrequent trial type (e.g., incongruent) was presented for a given condition (e.g., MC). Unshifted = the last four trials represented a continuation of the ongoing induction (i.e., were MC in an MC list). Long = 4000 ms elapsed between the participant's response to the final trial of the induction phase and the stimulus onset of the first trial of the diagnostic phase. Short = 1000 ms elapsed between the participant's response to the final trial of the induction phase and the stimulus onset of the first trial of the diagnostic phase.

Table 4

Experiment 3 Diagnostic Phase Reaction Time (ms) and Error Rate with Standard Deviations in Parentheses

Phase	Delay	PC	Window	Trial Type	Reaction Time	Error %
Diagnostic	Short	MC	Shifted	Congruent	643 (85)	0.44 (1.51)
				Incongruent	705 (82)	3.92 (8.05)
			Unshifted	Congruent	630 (86)	0.28 (1.29)
				Incongruent	703 (82)	4.20 (5.69)
		MI	Shifted	Congruent	623 (80)	0.23 (4.95)
				Incongruent	722 (94)	5.18 (6.87)
			Unshifted	Congruent	632 (80)	0.28 (1.07)
				Incongruent	715 (100)	4.25 (5.99)
Diagnostic	Long	MC	Shifted	Congruent	645 (79)	0.29 (1.13)
				Incongruent	706 (89)	4.45 (6.16)
			Unshifted	Congruent	629 (85)	0.30 (1.13)
				Incongruent	707 (86)	4.16 (7.22)
		MI	Shifted	Congruent	618 (74)	0.20 (0.87)
				Incongruent	717 (89)	4.16 (6.13)
			Unshifted	Congruent	636 (82)	0.28 (1.07)
				Incongruent	708 (89)	3.84 (5.67)

Note: MC = mostly congruent; MI = mostly incongruent; Shifted = in the last four trials of the induction (the window), only the infrequent trial type (e.g., incongruent) was presented for a given condition (e.g., MC). Unshifted = the last four trials represented a continuation of the ongoing induction (i.e., were MC in an MC list). Long = 4000 ms elapsed between the participant's response to the final trial of the induction phase and the stimulus onset of the first trial of the diagnostic phase. Short = 1000 ms elapsed between the participant's response to the final trial of the induction phase and the stimulus onset of the first trial of the diagnostic phase.