Conflict-monitoring theory in overtime: Is temporal learning a viable explanation for the congruency sequence effect?

Giacomo Spinelli[1] & Stephen J. Lupker[2]

[1] Università degli Studi di Milano-Bicocca

[2] University of Western Ontario

Running Header: Conflict-monitoring theory in overtime

Corresponding Author:

Giacomo Spinelli

Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca

Piazza dell’Ateneo Nuovo 1, 20126 Milano (MI), Italy

E-mail: giacomospinelli@hotmail.it

Word count: 27278

Abstract

In interference tasks (e.g., Stroop, 1935), congruency effects are larger following a congruent vs. an incongruent trial. This “congruency sequence effect” has been traditionally explained in terms of a conflict-monitoring mechanism that focuses attention toward relevant information when conflict has recently been experienced. More recently, it has been suggested that effects of this sort result from differences in the temporal expectancies formed following congruent trials (fast responding) vs. incongruent trials (slow responding). Evidence supporting this “temporal-learning” account was recently reported for a similar effect, the finding that congruency effects are larger in a mostly-congruent list than in a mostly-incongruent list. That is, consistent with the idea that this “proportion-congruent effect” is based on different temporal expectancies following congruent versus incongruent trials in interference tasks, the proportion-congruent effect was eliminated on normal (i.e., immediate-response) trials when temporal expectancies were equated by requiring a delayed response on the prior trial. In two experiments, we examined whether this delayed-response procedure would have a similar impact on the congruency sequence effect. Consistent with the temporal-learning account (but not inconsistent with conflict-monitoring accounts), the congruency sequence effect on immediate-response trials was eliminated when the previous trial required a delayed response. However, no evidence supporting the temporal-learning account emerged from re-analyses of experiments requiring only immediate responses in which the response latency in the previous trial functioned as the temporal-expectancy index. Overall, the present results and analyses do not provide much evidence favoring the temporal-learning account over conflict-monitoring accounts of the congruency sequence effect.

Keywords: congruency sequence effect; proportion-congruent effect; temporal learning; conflict adaptation; conflict monitoring

Public Significance Statement

The "congruency sequence effect" is an effect that has traditionally been interpreted as evidence for people's ability to adapt their attention based on whether they recently had to deal with distraction. The present research examines an alternative interpretation according to which that effect depends on the point in time at which people expect to be able to make a response to a stimulus. Although this alternative interpretation gained support from one of the lines of research that we pursued, it gained no support from other lines of research.

Conflict-monitoring theory in overtime: Is temporal learning a viable explanation for the congruency sequence effect?

In recent years, a question that has been generating considerable research interest is the question of whether attention to task-relevant versus task-irrelevant dimensions of a stimulus can be adaptively controlled in response to the expectation, based on previous experimental trials, that a conflict between those dimensions will arise. A cognitive control system that involves such a conflict-adaptation mechanism would be able to adjust to the expectation of conflict by biasing attention towards the task-relevant dimension and/or away from the task-irrelevant dimension when conflict is expected, while not doing so when conflict is unexpected (Botvinick et al., 2001). Typical examples of situations in which attentional control would likely be invoked are situations in which conflict has been recently experienced, either once or on multiple occasions.

This issue has typically been examined experimentally using interference tasks such as the Stroop (1935) task. In these tasks, participants are required to respond to one dimension of a bidimensional stimulus (e.g., the ink color of a word) while ignoring the other dimension (e.g., the identity of the word). The task-relevant and task-irrelevant dimensions can be congruent (i.e., not conflicting) with each other (e.g., the word RED in red color) or incongruent (i.e., conflicting) with each other (e.g., the word RED in blue color). The typical result is a “congruency effect”, with faster (and often more accurate) responses to congruent than incongruent items (for a review, see MacLeod, 1991).

Of interest here is the fact that previous experience with an incongruent trial typically modulates the magnitude of this effect. For example, the congruency effect is larger following a congruent trial than following an incongruent trial (Gratton et al., 1992; for a review, see Egner, 2014). This effect, known as the “congruency sequence effect”, has traditionally been interpreted as the result of the action of a conflict-adaptation process (Botvinick et al., 2001). According to this interpretation, a control mechanism exists that monitors conflict and adapts attention between task-relevant and task-irrelevant dimensions of the stimulus accordingly. Experiencing conflict between those dimensions during an incongruent trial would cause the conflict-monitoring system to emit a signal indicating the need for more focused attention to the task-relevant dimension (e.g., the color in the Stroop task) on the subsequent trial (this signal, however, will not be used to resolve conflict on the current trial, although other control accounts do allow for such a possibility, e.g., Ridderinkhof, 2002). This additional attention would help reduce the impact of conflict on that subsequent trial, with the result being a smaller impact of incongruity if the following trial is an incongruent trial, leading to smaller congruency effects on those trials. Conversely, experiencing little or no conflict during a congruent trial would induce relaxed attention on the subsequent trial because there is less reason to tighten control. The result would be increased interference from the task-irrelevant dimension if the subsequent trial is an incongruent one, and, hence, a larger congruency effect on those trials.

Further, this conflict-adaptation process is assumed to be cumulative: Experiencing conflict over several trials (e.g., in a sequence of mainly incongruent trials) would induce additional focused attention to the task-relevant dimension, leading to a minimization of interference from the task-irrelevant dimension. In contrast, attention would be increasingly relaxed when little or no conflict is experienced over several trials (e.g., in a sequence of mainly congruent trials), leading to increased interference when conflict does arise (Aben et al., 2017; Jiménez & Méndez, 2013; 2014; see also Colvett et al., 2020, for evidence from a design minimizing confounds such as those discussed below).

Thus characterized, conflict adaptation would explain not only the congruency sequence effect, but also a related effect known as the Proportion-Congruent (PC) effect (Botvinick et al., 2001). This effect refers to the finding that the congruency effect is typically larger in a list in which congruent items are frequent (i.e., a Mostly Congruent (MC) list) than in a list in which congruent items are infrequent (i.e., a Mostly Incongruent (MI) list; e.g., Logan & Zbrodoff, 1979; for a review, see Bugg & Crump, 2012). The conflict-adaptation interpretation of the PC effect proposes that because conflict is frequent in an MI list, attentional control would often be tightened. As a result, interference would be minimized. In contrast, because conflict is only rarely experienced in an MC list, attentional control would often be relaxed. As a result, interference would be much larger when conflict does arise. Essentially, the conflict-adaptation process that produces the congruency sequence effect, a local (trial-level) effect, is assumed to also be mainly responsible for producing the PC effect, a global (list-level) effect. A reasonable question, however, is whether conflict adaptation is, in fact, the correct explanation for either the congruency sequence effect or the PC effect. (note 1)

The Role of Binding and Contingency Learning in the Congruency Sequence Effect and the PC Effect

In recent years, a growing body of research has cast doubt on conflict-monitoring theory and conflict-adaptation mechanisms in general as a valid explanation for the congruency sequence effect and the PC effect (e.g., Algom & Chajut, 2019; Schmidt, 2013a, 2019; Schmidt et al., 2015). The reason for this doubt is the realization that the paradigms used to examine those effects typically contain one or more confounds that allow alternative interpretations of those effects, interpretations not based on conflict-adaptation processes.

One such confound is represented by the fact that, in many paradigms used to examine the congruency sequence effect, congruent-congruent (c-C) sequences and incongruent-incongruent (i-I) sequences often involve either complete repetitions of both task-relevant and task-irrelevant stimulus features (e.g., in i-I sequences, RED in blue may be followed by another RED in blue) or a complete change of both features (e.g., in i-I sequences, RED in blue may be followed by GREEN in yellow). In contrast, congruent-incongruent (c-I) sequences and incongruent-congruent (i-C) sequences often involve partial repetitions in which one feature (e.g., the word) repeats while the other (e.g., the color) changes (e.g., in c-I sequences, RED in red may be followed by RED in yellow). What makes this observation relevant is the finding that partial repetitions typically elicit slower responses than do complete repetitions or complete changes. This finding has been interpreted as being the result of a binding process whereby, on partial repetitions, the repeated stimulus feature encourages retrieval of both features from the previous trial as those features had been at least momentarily bound together (e.g., RED in yellow following RED in red would induce retrieval of both the word RED from the previous trial and the red color that the word RED was bound to), a process that would generate additional interference (e.g., both the word RED and the retrieved red color would interfere with naming the yellow color on the current trial; Hommel et al., 2004). The implication is that the typical finding in the congruency sequence effect that c-C sequences are faster than i-C sequences and i-I sequences are faster than c-I sequences may simply reflect a binding process that would slow down the i-C and c-I sequences because of the frequent presence of partial repetitions in those sequences. That is, the congruency sequence effect itself may be entirely the result of a binding confound (Mayr et al., 2003).

Proportion-congruent paradigms also contain a non-conflict confound. This confound is represented by the fact that, in these paradigms, participants can learn contingencies between a word and a motor response (Schmidt & Besner, 2008). For example, in the typical PC paradigm, any word in the MC list most frequently requires a congruent response (e.g., RED would typically require the “red” response because RED frequently occurs in the (congruent) red color). If participants learn these contingencies, responses to the (high-contingency) congruent colors will speed up whereas responses to (low-contingency) incongruent colors will not. As a result, the congruency effect in the MC list will be large. In MI lists, a contingency learning process would, if anything, lead to a reduction in the congruency effect. This reduction would be observed, for example, if the MI list is constructed so that each word appears only in two colors, the infrequent congruent color and a frequent incongruent color, a situation in which use of contingency learning would speed up responses to the incongruent, but high-contingency, color but may not affect responses to the congruent, but low-contingency, color. (note 2)

Since binding and contingency-learning confounds were brought to the attention of the research community, a number of studies have been designed to re-examine the congruency sequence effect and the PC effect in situations in which those confounds were controlled for (Braem et al., 2019). For congruency sequence paradigms, this goal was often accomplished by alternating stimuli belonging to one set (e.g., RED and BLUE and their corresponding colors) with stimuli belonging to another set (e.g., GREEN and YELLOW and their corresponding colors) on successive trials (e.g., Schmidt & De Houwer, 2011; Schmidt & Weissman, 2014). Because the two sets do not overlap, all sequences in this paradigm (c-C, c-I, i-C, and c-I) involve a complete change of the features (i.e., both the word and the color change from one trial to the next), a solution that prevents binding processes from producing a congruency sequence effect.

A somewhat similar solution was developed in the PC paradigm to control for the contingency-learning confound (e.g., Blais & Bunge, 2010; Bugg, 2014; Bugg et al., 2008). Here as well, the stimuli are typically divided into two non-overlapping sets, which we will refer to as the “context” set and the “transfer” set. The transfer items (e.g., RED and BLUE and their corresponding colors) are 50:50 congruent/incongruent and are intermixed in a list with context items (e.g., GREEN and YELLOW and their corresponding colors) that are either mostly congruent (creating an overall MC list) or mostly incongruent (creating an overall MI list). The rationale for this manipulation is that, while a PC effect obtained on the context items might result from learning of word-response contingencies, contingencies that change across lists (e.g., GREEN predicts a “green” response in the MC list and a “yellow” response in the MI list), contingency learning should not produce a PC effect on the transfer items because those items are identical in the two lists.

Essentially, both the alternating-set solution (in the congruency sequence paradigm) and the context/transfer solution (in the PC paradigm) create situations in which non-conflict processes should not be able to produce the congruency sequence effect or the PC effect, respectively. Nevertheless, those effects were obtained in those situations a number of times (for the congruency sequence effect, see, e.g., Kim & Cho, 2014; Schmidt & Weissman, 2014; Weissman et al., 2014; for the PC effect, see, e.g., Bugg, 2014; Bugg & Chanani, 2011; Spinelli & Lupker, 2020, in review), although not always (for the congruency sequence effect, see, e.g., Mayr et al., 2003; Schmidt & De Houwer, 2011; for the PC effect, see, e.g., Blais & Bunge, 2010; Bugg et al., 2008). What these results suggest is that, although binding and contingency learning may contribute to effects traditionally attributed to conflict-induced control, those non-conflict processes likely do not provide the complete explanation for those effects.

Another Non-Conflict Process: Temporal Learning

The fact that the congruency sequence effect and the PC effect emerge in situations in which binding and contingency-learning processes are controlled for raises the question of what process produces those effects in those situations. An obvious answer is that those effects are the result of a process of conflict adaptation, the traditional explanation (Bugg, 2014; Bugg & Chanani, 2011; Hutchison, 2011; Spinelli & Lupker, 2020, in review; Spinelli et al., 2019). More recently, however, Schmidt (2013b, 2017; Schmidt & Weissman, 2016) proposed a different explanation. According to this explanation, in addition to binding and contingency learning, another confound exists in paradigms used to examine the congruency sequence effect and the PC effect that could produce those effects without the necessary involvement of a conflict-adaptation process.

This confound has to do with the fact that participants in speeded tasks are known to form temporal expectancies for the emission of a response, expectancies which influence their performance in the task (for supporting evidence, see, e.g., Lupker et al., 1997; Taylor & Lupker, 2001). Based on these ideas, Schmidt (2013b; Schmidt & Weissman, 2016) proposed that a temporal-learning process, rather than a conflict-adaptation process, is responsible for the congruency sequence effects and the PC effects obtained in situations in which the relevant non-conflict confounds were controlled for.

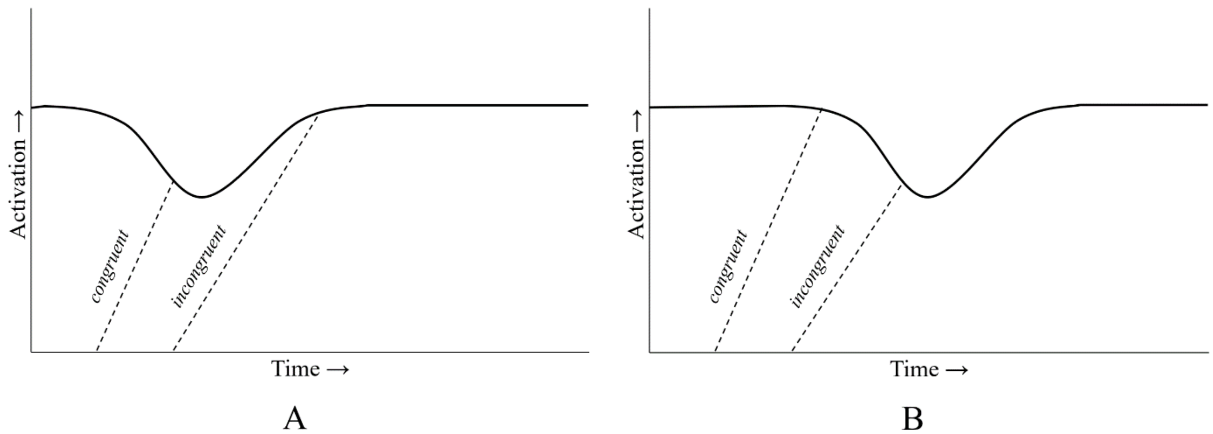
According to this account, participants' response latencies in both congruency sequence and PC manipulations are influenced by temporal expectancies based on the response latencies on the previous trials, particularly the most recent one. (note 3) This temporal expectancy is relatively fast following a congruent trial (or in an MC list where, for any given trial, the previous trials are often congruent) because participants have recently emitted one or more fast responses. Participants will thus anticipate a fast response in those situations. In contrast, the temporal expectancy is relatively slow following an incongruent trial (or in an MI list where, for any given trial, the previous trials are often incongruent) because participants have recently emitted one or more slow responses. Participants will thus anticipate a slow response in those situations.

The crucial assumption of Schmidt's (2013b; Schmidt & Weissman, 2016) particular account, and the one that distinguishes it from other temporal accounts (e.g., the time-criterion account: Lupker et al., 1997), is that the only impact of the process is a speed-up of responding in particular situations. Specifically, emitting a response to an item will be affected (i.e., sped up) by the temporal expectancy if normal processing has not allowed the item to be responded to at the point in time where the expectancy is set but has been processed sufficiently that a response can be emitted at that point. The idea is that a temporal expectancy would cause a temporary drop in the threshold activation that needs to be achieved in order to emit a response around the point in time set by that expectancy. Because the threshold to emit a response is a bit lower than normal around that point in time, responding would be faster than normal (a speed-up) for items for which a response is nearly ready for emission if, in fact, a response had not already been emitted. This situation is represented schematically in Figure 1. In this figure, the response threshold (solid line) drops temporarily around the point in time where the temporal expectancy has been established. Note also that the Figure reflects the idea that response activation accumulates at a faster rate for congruent than incongruent items (dashed lines), because the latter are more difficult to process. If activation for an item crosses the lowered threshold near the temporal expectancy, the response for that item will be emitted earlier than normal. Which item can receive this benefit, however, depends on where the temporal expectancy is set.

When the temporal expectancy occurs early (Figure 1A), such as following a congruent trial or in an MC list, responses to congruent items, but not to incongruent items, will speed up because a response for congruent items will be available around the fast temporal expectancy established in that situation. The result would be a relatively large congruency effect following a congruent trial or in an MC list. In contrast, following an incongruent trial or in an MI list, the drop in the response threshold would occur later in time than it does following a congruent trial or in an MC list. As a result, congruent items will be unaffected by the temporal expectancy because responses to those items have already been emitted by that point in time. Hence, latencies for congruent items following incongruent (vs. congruent) trials, or in an MI list (vs. an MC list) (Figure 1B), will be longer. The situation is different for incongruent items. These items may be processed fast enough to meet the temporal expectancy when that expectancy is slow (e.g., following an incongruent trial or in an MI list) but not when it is fast (e.g., following a congruent trial or in an MC list). Hence, one may observe faster responding to those items in the former situation than in the latter. Although, in practice, slow temporal expectancies tend to have little impact on incongruent items (Kinoshita et al., 2011; Kinoshita & Mozer, 2006; for an explanation, see Schmidt & Weissman, 2016), the overall result of the process of generating expectancies in the way proposed by the temporal-learning account (Schmidt, 2013a) would be a larger congruency effect following a congruent trial or in an MC list. (note 4)

Figure 1

An Illustration of Schmidt's Temporal-Learning Account for Fast vs. Slow Temporal Expectancies



Note. The solid line represents the response threshold. The dip in the threshold occurs at the temporal expectancy, which can be fast (panel A) or slow (panel B). The dashed lines represent the (average) accumulation of activation over time for typical congruent items (fast rate of accumulation) and incongruent items (slow rate of accumulation). The point in time where the dashed line intersects the solid line is when a response will be emitted for that item.

Essentially, the crucial point here is that a temporal-learning process of the sort described by Schmidt (2013b) will tend to produce a larger congruency effect in situations promoting a fast temporal expectancy (e.g., following a congruent trial or in an MC list) than in situations promoting a slow temporal expectancy (e.g., following an incongruent trial or in an MI list), the typical patterns of the congruency sequence effect and the PC effect. (note 5) If this account is accurate, conflict adaptation would not need to be invoked to explain those effects, even when those effects were obtained while controlling for known confounds such as binding and contingency learning. The yet-to-be settled question, however, is whether humans actually do use a temporal-learning process of the sort Schmidt proposed.

Schmidt and colleagues (Schmidt, 2013b; 2014; 2016; Schmidt et al., 2014; Schmidt & Weissman, 2016) initially attempted to address this question using two approaches. The first approach Schmidt and colleagues used to test Schmidt's (2013b) temporal-learning account, albeit only the temporal-learning account of the PC effect, was based on the idea that temporal learning should not be specific to interference tasks, i.e., tasks where conflict from an irrelevant dimension produces interference. The other approach consisted of attempting to control for temporal learning while re-analyzing the data from confound-minimized studies which reported a congruency sequence effect (Schmidt & Weissman, 2016) or a PC effect (Schmidt, 2013b). As will be discussed later in the manuscript (see the sections "Generalized Linear Mixed-Effects Model (GLMM) Re-Analyses" and "Additional Challenges for the Temporal-Learning Account" in the General Discussion), these approaches either seem to work in a limited number of situations (the approach based on conflict-free tasks: Spinelli et al., 2019) or derive all their support from questionable methodology (the approach based on re-analyses: Cohen-Shikora et al., 2019; Spinelli et al., 2019).

Schmidt (2017), however, has recently reported a third approach aimed at supporting his temporal-learning account. Notably, this approach, not scrutinized so far, does appear to support an important role of temporal learning in the PC effect. Because the present research was an attempt to follow up on Schmidt (2017), this approach is described in detail in the next section.

Controlling for Temporal Learning in the PC Paradigm: Schmidt (2017)

To support his temporal-learning account of the PC effect, Schmidt (2017) developed a new paradigm allowing for control of the temporal-learning process in a PC manipulation. This manipulation was implemented in a prime-probe task, a task in which participants are required to press the key designated for a word (the “probe”) while ignoring another word (the “prime”), that is briefly presented in a larger font than the probe, just before the probe appears (Schmidt & Weissman, 2014; see also Kunde & Wühr, 2006). The prime and the probe typically belong to a restricted category of words (e.g., location words) and they can be congruent with each other (e.g., “Right” as the prime and “$_{\text{Right}}$”, in smaller font, as the probe) or incongruent with each other (e.g., “Left” as the prime and “$_{\text{Right}}$” as the probe). This task not only produces a basic congruency effect (i.e., shorter latencies for congruent than incongruent pairs) but also a robust congruency sequence effect (i.e., a larger congruency effect following a congruent trial than following an incongruent trial) even when no binding confound exists (Schmidt & Weissman, 2014; for a comparison with other tasks, see, e.g., Duthoo et al., 2014; Schmidt & De Houwer, 2011).

In order to control for contingency learning, Schmidt (2017) divided the items in this task into two sets: A set of transfer items (e.g., left and right) presented on even numbered trials which were 50:50 congruent/incongruent in both lists, and a set of context items (e.g., up and down) presented on odd numbered trials which were mainly congruent in the MC list and mainly incongruent in the MI list. That is, although the transfer items were 50:50, the inclusion of context items made, overall, one list an MC list and one list an MI list.

Paralleling similar PC manipulations (e.g., Bugg et al., 2008), different contingencies could be learned in the two lists for the stimuli used for the context set (e.g., in the MC list, a contingency could be learned between the prime “up” and the congruent response “up”; conversely, in the MI list, a contingency could be learned between the prime “up” and the incongruent response “down”). As a result, a PC effect obtained for the context stimuli would be compatible with any of a number of processes (i.e., conflict adaptation, contingency learning, or temporal learning). In contrast, because the stimuli used for the transfer set were identical in the two lists, a PC effect obtained for those stimuli would not be compatible with a contingency-learning process, with the remaining options being conflict adaptation and/or temporal learning.

The new contribution made by Schmidt (2017) was introducing a manipulation intended to disable any impact of temporal learning for the transfer items (i.e., an impact deriving from the immediately preceding trial involving a context item). The rationale was that, if a situation is created in which not only contingency learning but also temporal learning cannot produce a PC effect, the finding that a PC effect is maintained would represent evidence in support of the view that this effect reflects, at least in part, a conflict-adaptation process. In contrast, if the PC effect is completely eliminated in this situation, that result would suggest that the PC effect is entirely produced by contingency-learning and/or temporal-learning processes.

To control for the potential impact of temporal learning affecting responding on the transfer item trials, Schmidt (2017) presented a wait cue requiring participants to momentarily withhold the response to most of the context items, specifically, on most of the *congruent* context item trials in the MC list and on an equivalent number of *incongruent* context item trials in the MI list. These items, termed by Schmidt “inducer” items, defined the congruency proportion of the list because the two lists were otherwise identical. A normal, i.e., immediate, response was required for the other context items (termed by

Schmidt "biased" items, which had a 50:50 congruent/incongruent ratio) and for the transfer items (those appearing on even numbered trials, the "diagnostic" items in Schmidt's terminology).

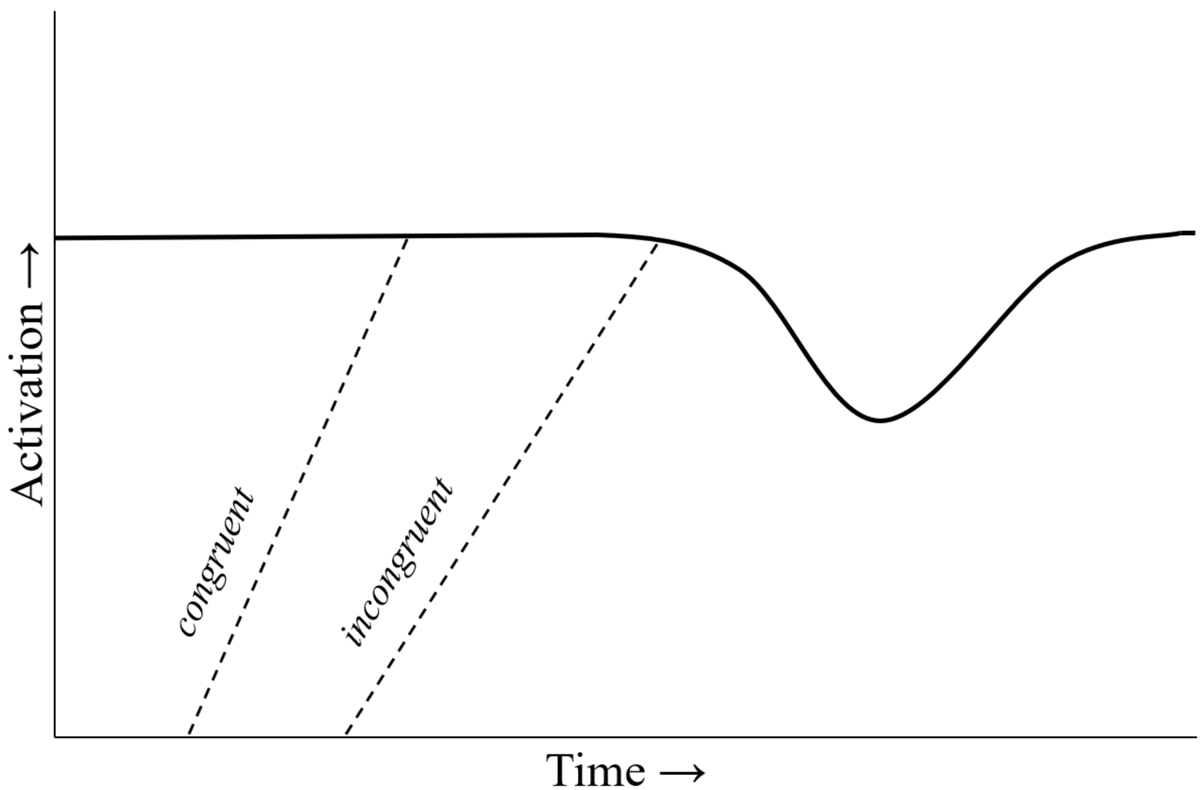
One group of participants (the short-wait group) functioned as a baseline as the requested delay was very short (the wait cue disappeared 200 ms after probe onset, signaling that a response could be made, although virtually no responses could be made faster than 200 ms in this task). As such, an essentially normal temporal-learning process should have been in play for this group. For another group of participants (the long-wait group), the requested delay was substantially longer (the wait cue disappeared 800 ms after probe onset; note that average latencies in this task are typically in the 500-700 ms range).

The reason that the time to respond was only controlled for the list-defining context items is that, in PC manipulations that distinguish transfer and context items (e.g., Bugg et al., 2008), the context items not only determine the congruency proportion in each list, but also the response speed (or rhythm) of the list. That is, normally, the presence of numerous congruent context items in the MC list inevitably makes the temporal expectancy in that list relatively fast overall (a fast rhythm). Conversely, the presence of numerous incongruent context items in the MI list inevitably makes the temporal expectancy in that list relatively slow overall (a slow rhythm). However, Schmidt reasoned that if responses to those context items are delayed in a way that makes responses to those items have essentially the same latencies in both lists (i.e., participants are forced to emit a response to congruent context items (in the MC list) and incongruent context items (in the MI list) at roughly the same point in time, a situation that would only occur when a longer delay is imposed, e.g., 800 ms), the temporal expectancy across lists would essentially be equated (i.e., the rhythm would be the same) even though a PC manipulation is also in place.

With a slow temporal expectancy being created by a long wait, temporal learning should produce no bias to respond rapidly to congruent or incongruent items requiring an immediate response (i.e., the transfer items) in either list. This situation is represented in Figure 2. In this figure, the response threshold drops so late that neither congruent nor incongruent items accumulate activation at a slow enough rate to make that temporal expectancy relevant. As a result, neither type of item would draw any benefit from that slow temporal expectancy. Thus, compared to the normal situation in which only immediate responses are required (or the situation in which the wait time is so short that it does not affect normal responding), the speed-up for congruent trials in an MC list (a situation that normally creates a fast temporal expectancy, see Figure 1A) will be lost, and so will be the (potentially smaller) speed-up for incongruent trials in an MI list (a situation that normally creates a slow temporal expectancy, although not as slow as the temporal expectancy created by the presence of delayed-response trials in the list, see Figure 1B).

Figure 2

An Illustration of Schmidt's Temporal-Learning Account for a Very Slow Temporal Expectancy (e.g., in Schmidt's (2017) long-wait condition)



Note. The solid line represents the response threshold. The dip in the threshold occurs at the temporal expectancy, which is very slow in this case. The dashed lines represent the (average) accumulation of activation over time for typical congruent items (fast rate of accumulation) and incongruent items (slow rate of accumulation). The point in time where the dashed line intersects the solid line is when a response will be emitted for that item.

The implication of this line of reasoning is that, in the long-wait condition used by Schmidt (2017), it would be impossible for a temporal-learning mechanism to produce a PC effect. Thus, obtaining a PC effect on the transfer items (i.e., items for which contingency learning is not possible due to the 50:50 congruency proportion) in this situation could only be due to a conflict-adaptation process. The data showed no PC effect on the transfer items.

Specifically, in Schmidt's (2017) Experiment 1, whereas participants in the short-wait group were, as expected, faster at producing a "delayed" response to congruent inducer items (in the MC list, the items that normally create the fast temporal expectancy in that list) than to incongruent inducer items (in the MI list, the items that normally create the slow temporal expectancy in that list), participants in the long-wait group took about the same time to produce a delayed response to congruent inducer items (in the MC list) and incongruent inducer items (in the MI list). According to Schmidt, this result suggests that his manipulation was effective at equating temporal expectancies following inducer items across lists in the long-wait group. More importantly, in the latencies (but not in the error rates), transfer items following inducer items showed a PC effect in the short-wait group, with this effect disappearing in the long-wait group, as the congruency effect for those transfer items was the same size in the MC and MI lists. Although a more complicated pattern emerged in Schmidt's Experiment 2, an experiment that controlled for the time spent processing conflict on long-wait trials, in that experiment, too, no regular PC effect was observed for transfer items following inducer items in the long-wait group. These results, overall, would seem consistent with the idea that a temporal-learning process of adjustment to those differences has an important role in the PC effect. The main question these results pose, however, is whether these data are also *in*consistent with conflict-monitoring explanations for the PC effect (Botvinick et al., 2001).

The basic idea behind Schmidt's (2017) manipulation was that the original version of the conflict-monitoring theory (Botvinick et al., 2001) would not predict the pattern of results he obtained. The reason is that, in both short-wait and long-wait conditions, conflict is more frequent in the MI list than in the MC list, and thus, a mechanism of adaptation to conflict frequency should lead to more focused attention to relevant information in the former list than in the latter. Schmidt argued that there is no component of conflict-monitoring theory that would suggest that requesting a withholding of a response on the items that determine the congruency proportion of the list would prevent such a control mechanism from working in the way that it normally does. Thus, according to this interpretation of Botvinick et al.'s (2001) conflict-monitoring theory, a PC effect should have emerged in both the short-wait and the long-wait group.

Naturally, however, additional assumptions could be added to Botvinick et al.'s (2001) conflict-monitoring theory so as to explain the elimination of the PC effect in the long-wait group. For example, as suggested by Schmidt (2017), it could be assumed that the delayed-response manipulation would essentially prevent conflict from having any real impact on the long-wait trials. That is, the presentation of the long wait cue would allow sufficient time for the participant to easily determine the identity of the probe even on incongruent trials, making the long-wait incongruent context items no more conflicting than the long-wait congruent context items. As a result, in the long-wait group, the two lists would not differ in conflict frequency anymore, leading to the elimination of the PC effect on transfer items.

In any case, while explanations of this sort are certainly possible, they would reflect, according to Schmidt (2017), a significant departure from the characteristics of the conflict-adaptation process implemented in the original conflict-monitoring model. Therefore, he argued that temporal learning would appear to offer a more straightforward and parsimonious explanation of his findings, an explanation that is directly compatible with a more general model of learning (Schmidt, 2013b; Schmidt et al., 2016). These considerations led Schmidt (2017) to argue that we should call "time-out" on conflict-monitoring theory as an account of the PC effect.

The Present Research

In the present research, we aimed to further explore Schmidt's (2017) delayed-response procedure. Specifically, we were interested in whether this procedure would have an impact not only on the PC effect but also on the congruency sequence effect. The reason that this idea is plausible is that, in its current form, Schmidt's temporal-learning account assumes that the PC effect and the congruency sequence effect arise from one and the same mechanism (see footnote 3). According to this view, when examining the PC effect, the congruency effect is larger in an MC list than in an MI list because on any given trial, the preceding trials typically lead to the formation of fast vs. slow temporal expectancies, respectively. Similarly, when examining the congruency sequence effect, the congruency effect is larger following a congruent trial than following an incongruent trial because the previous trial leads to the formation of fast vs. slow temporal expectancies, respectively. That is, both effects depend on the different temporal expectancies that previous trials, particularly the most recent one, create. The implication is that, if a manipulation is used that eliminates the temporal expectancy differences, both of those effects should be eliminated as well. Using a delayed-response procedure, Schmidt (2017) demonstrated this result for the PC effect. We aimed to examine whether that procedure would have a similar impact on the congruency sequence effect.

To this end, we applied Schmidt's procedure in an experiment that was similar to Schmidt's Experiment 1 in most ways except that proportion congruency was not manipulated (i.e., there was a single list where all items (both context and transfer) were 50:50 congruent/incongruent), as is standard in congruency sequence paradigms. In two experiments, context items, presented on odd trials, required either a delayed or an immediate response (with the duration of the delay being a between-subject factor), whereas transfer items, presented on even trials, always required an immediate response. The question of interest was whether the congruency sequence effect evaluated on transfer items would be altered by the fact that a (long) delayed, rather than a normal (i.e., immediate) response was requested for the context item presented on the previous trial.

The temporal-learning account predicts that the congruency sequence effect on transfer items would be reduced, if not eliminated, when a long-delay response was requested for the context item presented on the previous trial. The reason is that in this situation neither congruent nor incongruent items would benefit from the very slow temporal expectancy that is created (see Figure 2), unlike in the normal situation (i.e., when the previous trial requires an immediate response) in which congruent items benefit from the fast temporal expectancy created by congruent items (see Figure 1A) whereas incongruent items benefit, potentially, from the slow temporal expectancy created by incongruent items (see Figure 1B).

Following Schmidt's (2017) reasoning, the conflict-monitoring model, at least in its initial version (Botvinick et al., 2001), does not seem to make the same prediction. The reason is that, in that model, the conflict being monitored appears to be the conflict that is created whenever an (incongruent) stimulus causes (typically, two) conflicting response units to become simultaneously active. On context incongruent items, that type of conflict would presumably still be experienced when a (short or long) delay is required for the response to those items. Therefore, it could be presumed that conflict adaptation would occur even following long-delay context items. As a result, unless additional assumptions are added to that conflict-monitoring model (e.g., the assumption that wait cues would mitigate the impact of any conflict on long-delay incongruent context items), the prediction of the model appears to be for a regular congruency sequence effect on transfer items following long-delay context items.

On the other hand, a more recent version of the conflict-monitoring model (Yeung et al., 2011; see also Yeung et al., 2004) seems to make a somewhat different prediction, one that is essentially the same as that made by the temporal-learning account. The reason is that, in that version of the model, the claim was explicitly made that the monitor “does not distinguish between conflict caused by incongruent stimulus features and conflict caused by other sources of processing variability and noise… such as trial-to-trial fluctuations in attentional focus, noise in stimulus processing, and idiosyncratically varying response biases” (p. 317). In other words, any source of conflict would be monitored, not just conflict deriving from incongruent task-irrelevant information. Crucially, because this “conflict” would always result in a lengthening of RTs, RTs, not congruency, would be the most appropriate proxy for conflict. Therefore, a manipulation that equates RTs for congruent and incongruent stimuli, like what the long delay in Schmidt’s (2017) procedure appears to do, would essentially equate the conflict associated with those long-delay congruent and incongruent items, causing those items to produce a similar signal for adaptation. With a similar signal for adaptation being produced by long-delay context items regardless of their congruency, the prediction for the subsequent transfer items would then be that no congruency sequence effect should be observed for those items – essentially the same pattern as that predicted by the temporal-learning account.

This version of the conflict-monitoring model, however, is not without problems (Grinband et al., 2011a, 2011b; see also Algom & Chajut, 2019). Perhaps the most compelling problem is that, by assuming what essentially is a one-to-one correspondence between the explanandum (i.e., RTs) and the explanans (i.e., conflict), this version appears difficult to falsify. For example, whenever an increase in RTs is observed, even when it would not be expected from the theory (e.g., for incongruent stimuli following other incongruent stimuli), the argument could always be made that a high amount of an unspecified “conflict” was experienced in that situation for whatever reason. Indeed, the argument could be made that this “conflict” is experienced even in tasks in which there is no irrelevant dimension that can produce interference at all, such as in a simple signal detection task (Grinband et al., 2011a, 2011b) or the “conflict-free” tasks, mentioned above, that Schmidt (e.g., Schmidt, 2013b) designed specifically to eliminate conflict (i.e., interference from task-irrelevant information) from the picture when examining temporal-learning processes. Because, therefore, we do not see this version of the conflict-monitoring model as being particularly useful, in the following, we will mainly refer to the original version (Botvinick et al., 2001), the version that Schmidt (2017) also referred to, as the control-based account to contrast with the temporal-learning account of the congruency sequence effect.

What seems to be clear, in any case, is that a delayed-response procedure, by itself, would be unlikely to ever offer strong support for time-based accounts and against control-based accounts of the congruency sequence effect, although control-based accounts would likely require adjustments or a rethinking to explain any findings deriving from that procedure. Therefore, as the main goal of the present research was to evaluate the idea that temporal learning really plays a crucial role in the congruency sequence effect, and whether it really is this process that Schmidt's (2017) delayed-response procedure may have an impact on, we undertook an additional analysis.

Recent research from our lab suggests that in the picture-word interference task, temporal learning, as defined by Schmidt, may not play a critical role in either the PC effect or the congruency sequence effect (Spinelli et al., 2019; see also Cohen-Shikora et al., 2019). Based on these considerations, we also used a statistical technique developed to determine whether temporal expectancies can produce effects that otherwise appear to be effects of adaptation to conflict in a normal situation (i.e., a situation in which responses are not delayed) by re-analyzing the data from the short-wait groups in our experiments, a situation close to normal because, as noted, the delay imposed on delayed-response trials was negligible in the short-wait group (as it was in Schmidt's (2017) experiments as well). Further, we also re-analyzed the (no-wait) data Schmidt and Weissman (2016) used to make the case for an important role of temporal learning in the congruency sequence effect, data which also come from a prime-probe task but represent a completely normal situation because no delayed-response procedure was used at all.

Experiment 1

In this experiment, we first examined the question of whether Schmidt's (2017) delayed-response procedure has a similar impact on the congruency sequence effect as it had on the PC effect in Schmidt's experiments. Because a difference in temporal expectancies following congruent vs. incongruent trials may be crucial for the congruency sequence effect to emerge on the current (immediate-response) trial, eliminating that difference by requesting a long delay for the response on the previous trial should also eliminate the congruency sequence effect.

In this experiment, we adapted Schmidt's (2017) paradigm so that the set of items termed context items (e.g., up and down) were 50:50 congruent/incongruent and required either an immediate or a delayed (long or short) response; those items alternated with another set of items, the transfer items (e.g., left and right), which were also 50:50 congruent/incongruent but always required an immediate response. The two sets were not permitted to cross (e.g., the horizontal prime "left" was used for the horizontal probes LEFT and RIGHT, not for the vertical probes UP and DOWN), and congruency proportion was not manipulated, as typically occurs in examinations of trial-by-trial effects in interference tasks (e.g., Gratton et al., 1992; see also Egner, 2014). As a result, different from Schmidt (2017), contingency learning was impossible for all of the items in the task.

What was similar to Schmidt's experiments, however, was the fact that there were two blocks of trials and the congruency of delayed responses on the context items was blocked: All delayed-response trials were congruent in one block (called the "congruent-wait list", semi-paralleling Schmidt's MC list) and all delayed-response trials were incongruent in the other block (called the "incongruent-wait list", semi-paralleling Schmidt's MI list; the other type of context items in the two blocks always required an immediate response).

The reason that we maintained this aspect of Schmidt's procedure is that, although not discussed by Schmidt, there is a possibility that participants in situations of that sort may learn associations between congruency and response delay and use those associations in a strategic fashion. For example, a congruent-wait list where all delayed-response trials are congruent may bias participants to anticipate that a delayed response would be required whenever a congruent trial is presented, even if the item presented on that trial is a transfer item, a type of item that requires an immediate response. The use of this response strategy on the immediate-response congruent trials would result in participants momentarily withholding their response to those trials until it is determined that an immediate response is actually required. In contrast, because incongruent trials never require a delayed response in that list, participants would not feel a similar need to withhold their response when an incongruent item is presented. Ultimately, this strategy would result in a reduction of the congruency effect in that situation over and above any reduction caused by temporal learning per se. To the extent that response strategies of this sort may have had an influence in Schmidt's experiments, we decided to allow those strategies in the present Experiment 1 as well. In Experiment 2, the manipulation was changed slightly in order to negate the use of this type of strategy.

The rest of the procedural details were also identical to Schmidt's (2017) Experiment 1, including the use of a control short-wait group (in which the duration of the wait cue on delayed-response trials was so short that, in virtually all cases, participants did not have to actually withhold a response) and a long-wait group (in which the duration of the wait cue on delayed-response trials was long enough that, in most cases, participants had to withhold a response for some time because they were able to determine the appropriate response well before the wait cue disappeared).

Method

Participants

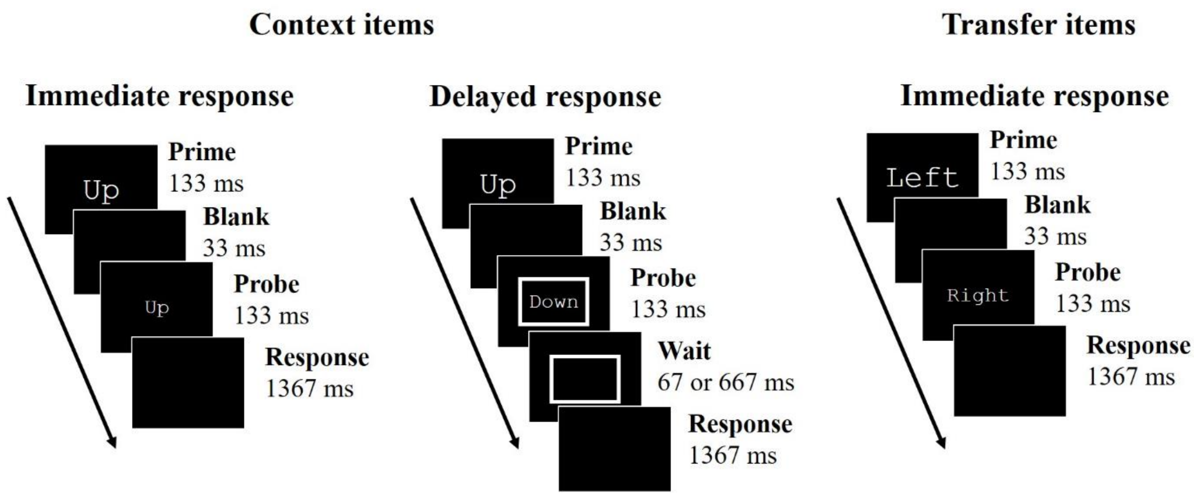
An a priori power analysis was performed using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to calculate the sample size needed to have a power of .80 for obtaining an effect using Schmidt's (2017) delayed-response procedure on the congruency sequence effect which would have a similar magnitude as the effect that that procedure was found to have on the PC effect in Schmidt's experiments. Based on an effect size of η_p^2 = .14, the average effect size of the crucial interaction in Schmidt's (2017) Experiments 1 (η_p^2 = .12) and 2 (η_p^2 = .16; i.e., the interaction reflecting significantly different PC effects in the short-wait vs. long-wait group for transfer items in the latencies), we determined that a minimum sample of 54 participants would be needed. Fifty-five students (20 males) at the University of Western Ontario (age 17–36 years) participated for course credit. All participants were native English speakers and had normal or corrected-to-normal vision.

Materials

The stimuli used for both the primes (the distracter stimuli) and the probes (the target stimuli) were the English words "Left", "Right", "Up" and "Down" presented in white bold Courier New font against a black background. The primes were presented in 20 pt font and the probes were presented in 10 pt font. Horizontal primes (i.e., "Left" and "Right") were only combined with horizontal probes and vertical primes (i.e., "Up" and "Down") were only combined with vertical probes. For both horizontal and vertical stimuli, the congruent prime-probe combination (e.g., Left-$_{\text{Left}}$) and the incongruent prime-probe combination (e.g., Left-$_{\text{Right}}$) were equally probable. One set of stimuli (e.g., the vertical stimuli) functioned as the context items. These items appeared on odd trials only and required an immediate response on half of the trials and a delayed response on the other half of the trials. The other set of stimuli (e.g., the horizontal stimuli) functioned as the transfer items. These items appeared on even trials only and always required an immediate response. The assignment of horizontal and vertical stimuli to the context and the transfer set was counterbalanced across participants. The need for participants to withhold the response on delayed-response trials was signaled with a wait cue. This wait cue was a 5 pixel white outline of a 40 X 80 pixel rectangle presented around the probe. For a representation of the stimuli and the experimental procedure, see Figure 3.

Figure 3

An Illustration of the Procedure for Context and Transfer Items in Experiment 1



Note. Context items appeared on odd numbered trials and required either an immediate response or a delayed response. This example is from the incongruent-wait list, a list where, for the context items, all delayed-response trials were incongruent (and, therefore, all immediate-response trials were congruent). Transfer items (either congruent or incongruent) appeared on even numbered trials and always required an immediate response. On delayed-response trials, the wait group determined how long the wait cue stayed on the screen after the probe disappeared (67 ms for the short wait group, 667 ms for the long wait group).

There were 2 blocks of 200 trials with a 5-second pause between the blocks. The difference between the two blocks was that in one block, the “congruent-wait list”, the congruent context items always required a delayed response and the incongruent context items always required an immediate response. In the other block, the “incongruent-wait” list, the congruent context items always required an immediate response and the incongruent context items always required a delayed response. The transfer items were the same in the two lists. The order in which the congruent-wait list and the incongruent-wait list were presented was counterbalanced across participants. The average frequency of the stimuli in one of the counterbalancing versions of the experiment in the congruent-wait list and the incongruent-wait list is presented in Table 1. The reason that this frequency is an average is that the stimuli were selected at random with replacement, as in Schmidt’s (2017) experiments. As a result, there was not a fixed number of stimuli per condition for every participant.

Table 1

Template for the Average Frequency of Prime-Probe Combinations in Experiment 1

				Context primes			
		Transfer primes		Immediate response		Delayed response	
List type	Probes	Left	Right	Up	Down	Up	Down
Congruent-wait	Left	25	25				
	Right	25	25				
	Up				25	25	
	Down			25			25
Incongruent-wait	Left	25	25				
	Right	25	25				
	Up			25			25
	Down				25	25	

Note. All transfer items required an immediate response.

<u>Procedure</u>

The procedure was identical to that used in Schmidt's (2017) Experiment 1, the experiment that produced the clearest pattern of results (see Figure 3). Each immediate-response trial included the following sequence of events: The prime (the distracter stimulus) presented for 133 ms, a blank screen for 33 ms, the probe (the target stimulus) for 133 ms, and another blank screen serving as a response window, which was presented until a response was made or until 1367 ms elapsed. All stimuli were presented in the center of the screen. Delayed-response trials had a similar structure, except that the white outline of a rectangle, serving as a wait cue, was displayed around the probe and remained on the screen for another 67 ms (200 ms total) in the short-wait condition or for another 667 ms (800 ms total) in the long-wait condition before the response window was presented.

Participants were requested to ignore the prime and to respond to the probe by pressing the corresponding key on a QWERTY keyboard. Specifically, they were instructed to press the F-key with their left middle finger for "left" responses, the G-key with their left index finger for "right" responses, the J-key with their right middle finger for "up" responses, and the N-key with their right index finger for "down" responses (note that these key positions are spatially compatible with the location words). Participants were requested to respond to the probe as quickly and as accurately as possible when no wait cue was presented. When a wait cue was presented, however, they were not supposed to respond until that cue had disappeared. At the end of each trial, the feedback messages "Error!", "Too slow!", and "Too fast!" were displayed in red, bold, 18 pt Courier New font for 1500 ms if the response made was incorrect (i.e., an incorrect button was pressed), if no response was made before the end of the response window, and if a response was made before the wait cue had disappeared, respectively. The trial proceeded immediately to the "too fast" message if a response was made before the response screen had appeared. If a correct response was made during the response screen, a 500-ms blank screen was presented instead. E-prime 2 (Experimental Software Tools, Pittsburgh, PA) software was used to present the stimuli and collect the responses. This research was approved by the Research Ethics Board of the University of Western Ontario (protocol # 108956).

Results

Five participants produced correct responses to less than 70% of the trials and were excluded from the analyses following Schmidt's (2017) criterion for participant exclusion. Fifty participants remained, 27 in the short-wait group and 23 in the long-wait group. Also following Schmidt's treatment of the data, null responses (i.e., too fast or too slow) and incorrect responses were both considered errors. Prior to the analyses, we discarded trials following trials in which an error or a null response was made (9.7% of the observations). Latency analyses were conducted on correct responses only. All RTs are reported from the onset of the probe, i.e., the target stimulus.

Also following Schmidt (2017), for both RTs and error rates, we analyzed the context and transfer items separately. The analysis of the context items, reported in the Supplementary Materials for simplicity, was used to verify that the delayed-response manipulation was effective, that is, that in the long-wait group, delayed responses to congruent items were no faster and no more accurate than delayed responses to incongruent items. The regular finding of faster and more accurate responses to congruent than incongruent items was expected for the other conditions, i.e., for immediate responses in the long-wait group and for both immediate and delayed responses in the short-wait group. Briefly, the results were consistent with our hypotheses: While context items showed a regular congruency effect when an immediate response or a short delay was requested (ranging from 76 to 120 ms in the latencies and from 3.7% to 6.1% in the error rates), this effect was greatly reduced when a long delay was requested (i.e., in the long-wait, delayed-response condition, where the congruency effect went down to 5 ms and 1.6% in the latencies and error rates, respectively). Thus, replicating Schmidt (2017), the delayed-response procedure was effective.

For transfer items, we ran a 2 (Current Congruency: Congruent vs. incongruent, within-subject) X 2 (Previous Congruency: Congruent vs. incongruent, within-subject) X 2 (Previous Response Type: Immediate vs. Delayed, within-subject) X 2 (Wait Group: Short Wait vs. Long Wait, between-subject) ANOVA. This analysis was aimed at examining the main research question, i.e., what impact does a delayed vs. immediate response to congruent vs. incongruent items on the preceding trial have on immediate responses to congruent vs. incongruent items on the subsequent trial.

In addition to traditional null-hypothesis significance testing analyses, to determine whether theoretically relevant non-significant effects (e.g., the interaction between Current Congruency and Previous Congruency) were indeterminate (i.e., potentially real) or null, we also performed Bayes Factor analyses for those effects. These analyses were performed in JASP version 0.14.1 (JASP Team, 2020) by comparing the most complex model containing the effect of interest, but no interactions with that effect (interpreted as the alternative hypothesis H_1), and the equivalent model stripped of that effect (interpreted as the null hypothesis H_0) using the default settings. The result of this comparison was BF_{01}, with $BF_{01} > 1$ suggesting evidence in support of H_0 (i.e., the absence of the effect), and $BF_{01} < 1$ suggesting evidence in support of H_1 (i.e., the presence of the effect) ($BF_{01} = 1$ would suggest equal evidence for the two hypotheses). Jeffreys's (1961) classification scheme (as reported in adjusted form by Lee and Wagenmakers, 2013) was used to help interpret the size of the Bayes Factor.

The mean RTs and error rates for the transfer items are presented in Table 2 with an indication of the list in which each condition appeared. The mean RTs for the transfer items are also depicted in Figure 4. The results of the analyses for these items are reported below. For this and for the following analyses, the raw data and the scripts used for the analyses are publicly available at https://osf.io/ynatu/.

Table 2

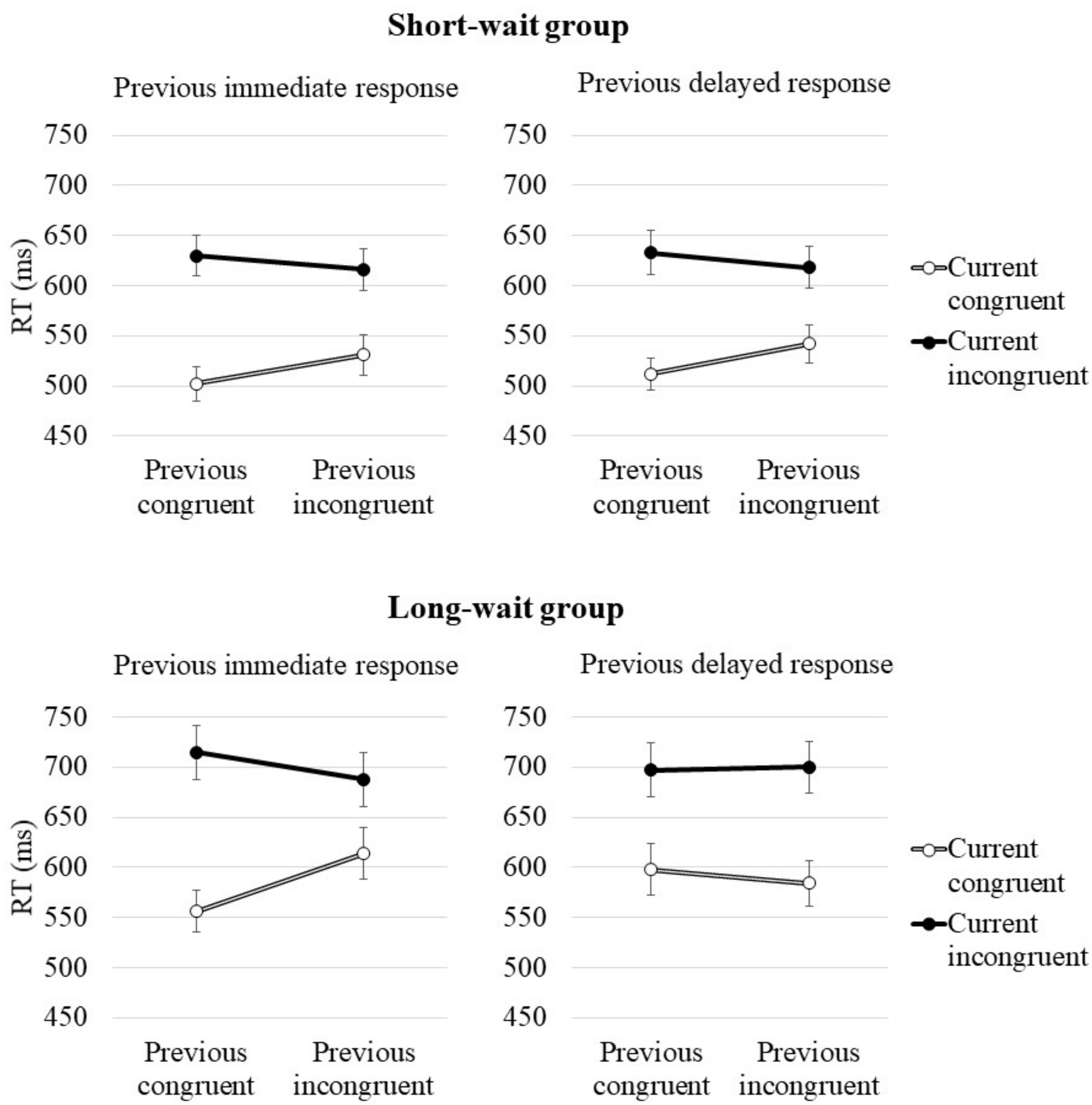
Mean RTs and Error Rates (and Corresponding Standard Errors) for the Transfer Items in Experiment 1

	RTs			Error rates		
Current congruency	Previous immediate response	Previous delayed response	Previous response type effect	Previous immediate response	Previous delayed response	Previous response type effect
<u>Short-wait group</u>						
Previous congruent						
Current congruent	<u>502 (17)</u>	**512 (16)**	10	<u>.050 (.008)</u>	**.044 (.009)**	-.006
Current incongruent	<u>630 (20)</u>	**633 (22)**	3	<u>.130 (.016)</u>	**.096 (.018)**	-.034
Current congruency effect	128	121	-7	.080	.052	-.028
Previous incongruent						
Current congruent	**531 (20)**	<u>542 (19)</u>	11	**.039 (.009)**	<u>.047 (.011)</u>	.008
Current incongruent	**616 (21)**	<u>618 (21)</u>	2	**.101 (.025)**	<u>.088 (.011)</u>	.013
Current congruency effect	85	76	-9	.062	.041	-.021
<u>Long-wait group</u>						
Previous congruent						
Current congruent	<u>556 (21)</u>	**598 (26)**	42	<u>.050 (.014)</u>	**.047 (.017)**	-.003
Current incongruent	<u>715 (27)</u>	**697 (27)**	-18	<u>.145 (.020)</u>	**.092 (.015)**	-.053
Current congruency effect	159	99	-60	.095	.045	-.050
Previous incongruent						
Current congruent	**614 (26)**	<u>584 (23)</u>	-30	**.035 (.009)**	<u>.054 (.018)</u>	.019
Current incongruent	**688 (27)**	<u>700 (26)</u>	12	**.076 (.017)**	<u>.100 (.026)</u>	.014
Current congruency effect	74	116	42	.041	.046	.005

Note. The current congruency effect is the difference between responses to incongruent and congruent items on the current trial. The previous response type effect is the difference between (immediate) responses to the current trial preceded by a delayed-response trial and (immediate) responses to the current trial preceded by an immediate-response trial. Means for trials appearing in the congruent-wait list and the incongruent-wait list are displayed in bold and underlined font, respectively.

Figure 4

Mean RTs (with Error Bars Indicating 1 Standard Error Below and Above the Mean) for the Transfer Items in Experiment 1



RTs

There was a main effect of Current Congruency (congruent faster than incongruent), $F(1, 48) = 304.99$, $MSE = 3749$, $p < .001$, $\eta_p^2 = .864$, and a main effect of Wait Group (short-wait group faster than long-wait group), $F(1, 48) = 6.07$, $MSE = 83368$, $p = .018$, $\eta_p^2 = .112$. Latencies were numerically shorter when the previous trial was congruent than when it was incongruent, however, the main effect of Previous Congruency was not significant, $F(1, 48) = 3.06$, $MSE = 1211$, $p = .087$, $\eta_p^2 = .060$. There was also an interaction between Current Congruency and Previous Congruency, $F(1, 48) = 26.82$, $MSE = 1423$, $p < .001$, $\eta_p^2 = .358$, which indicated a regular congruency sequence effect, with an overall larger congruency effect on the subsequent trial when the previous trial was congruent (125 ms) than when it was incongruent (86 ms). In addition, there was a three-way interaction between Current Congruency, Previous Congruency, and Previous Response Type, $F(1, 48) = 9.52$, $MSE = 1608$, $p = .003$, $\eta_p^2 = .165$. Follow-up analyses for this interaction indicated that while there was a regular congruency sequence effect on trials following immediate-response trials (with a 142-ms congruency effect following congruent trials vs. a 80-ms congruency effect following incongruent trials), $F(1, 48) = 32.53$, $MSE = 1565$, $p < .001$, $\eta_p^2 = .404$, the overall congruency sequence effect was numerically reduced (a 110-ms congruency effect following congruent trials vs. a 94-ms congruency effect following incongruent trials) and nonsignificant on trials following delayed-response trials, $F(1, 48) = 1.75$, $MSE = 1467$, $p = .192$, $\eta_p^2 = .035$. More importantly, a four-way interaction emerged between Current Congruency, Previous Congruency, Previous Response Type, and Wait Group, $F(1, 48) = 10.28$, $MSE = 1608$, $p = .002$, $\eta_p^2 = .176$.

This four-way interaction was explored by analyzing each wait group separately. In the short-wait group, the only significant interaction was that between Current Congruency and Previous Congruency, $F(1, 26) = 23.29$, $MSE = 1129$, $p < .001$, $\eta_p^2 = .473$, indicating a regular congruency sequence effect, with a larger congruency effect on the subsequent trial when the previous trial was congruent (123 ms) than when it was incongruent (80 ms). Notably, in this group, Previous Response Type did not interact with any other factor (all *p*s > .30). In the long-wait group, there was also an interaction between Current Congruency and Previous Congruency, $F(1, 22) = 7.62$, $MSE = 1771$, $p = .011$, $\eta_p^2 = .257$, indicating an overall congruency sequence effect, with a larger congruency effect on the subsequent trial when the previous trial was congruent (128 ms) than when it was incongruent (92 ms). More importantly, the three-way interaction was also significant in this group, $F(1, 22) = 13.22$, $MSE = 2230$, $p = .001$, $\eta_p^2 = .375$.

This three-way interaction in the long-wait group was explored, first, by splitting the data by Previous Response Type. This analysis showed that when the previous trial required an immediate response, Current Congruency and Previous Congruency interacted, $F(1, 22) = 21.95$, $MSE = 1888$, $p < .001$, $\eta_p^2 = .499$. This interaction indicated a regular congruency sequence effect, with a larger congruency effect on the subsequent trial when the previous trial was congruent (in the incongruent-wait list; 159 ms) than when it was incongruent (in the congruent-wait list; 74 ms). In contrast, when the previous trial required a delayed response, Current Congruency and Previous Congruency did not interact, $F(1, 22) = .73$, $MSE = 2113$, $p = .403$, $\eta_p^2 = .032$. The Bayes Factor for the comparison between the model with the interaction and the model without it was $BF_{01} = 2.77$, meaning that the data were 2.77 times more likely to occur under the hypothesis of no interaction than under the hypothesis of an interaction. In Jeffreys's (1961) classification scheme, this value would suggest "anecdotal" evidence for the absence of the interaction. Importantly, however, there was not even a numerical tendency for a congruency sequence effect as the congruency effect following congruent trials (99 ms) was, in fact, smaller than the congruency effect following incongruent trials (116 ms). Thus, a long-delayed response to the previous trial completely eliminated the congruency sequence effect on the subsequent trial.

The three-way interaction between Current Congruency, Previous Congruency, and Previous Response Type found in the long-wait group was also explored from a different perspective, that is, by splitting the data in this group by Previous Congruency. When the previous trial was congruent, Current Congruency and Previous Response Type interacted, $F(1, 22) = 6.95$, $MSE = 2980$, $p = .015$, $\eta_p^2 = .240$. This interaction revealed a Previous Response Type effect selectively for (current) congruent trials: Congruent trials following a delayed-response congruent trial (trials which, given the blocked nature of the design, would only appear in the congruent-wait list) were significantly slower (42 ms) than congruent trials following an immediate-response congruent trial (trials which would only appear in the incongruent-wait list), $t(22) = -2.78$, $p = .011$. The tendency was, if anything, in the opposite direction for incongruent trials, with a delayed response to the previous congruent trial (in the congruent-wait list) leading to a nonsignificant speed-up (18 ms) compared to when the previous congruent trial required an immediate response (in the incongruent-wait list), $t(22) = .99$, $p = .332$, $BF_{01} = 2.95$.

When the previous trial was incongruent, there was also an interaction between Current Congruency and Previous Response Type, $F(1, 22) = 9.26$, $MSE = 1054$, $p = .006$, $\eta_p^2 = .296$. However, this interaction had the opposite pattern compared to the interaction reported above for trials that followed congruent trials: Congruent trials following a delayed-response incongruent trial (in the incongruent-wait list) tended to be faster (30 ms) than congruent trials following an immediate-response incongruent trial (in the congruent-wait list), whereas incongruent trials following a delayed-response incongruent trial (in the incongruent-wait list) were slightly slower (12 ms) than incongruent trials following an immediate-response incongruent trial (in the congruent-wait list). However, neither the 30-ms speed-up for congruent trials nor the 12-ms slow-down for incongruent trials reached significance, $t(22) = 1.63$, $p = .118$, $BF_{01} = 1.46$, and $t(22) = -.61$, $p = .550$, $BF_{01} = 3.87$, respectively. Thus, it would appear that in the long-wait group, the primary reason why the congruency sequence effect was eliminated following a delayed-response trial was a slow-down for congruent trials when the previous (delayed) trial was also congruent.

Error rates

There were main effects of Current Congruency (congruent more accurate than incongruent), $F(1, 48) = 45.55$, $MSE = .007$, $p < .001$, $\eta_p^2 = .487$, and Previous Congruency (overall less accurate responses following congruent trials than following incongruent trials), $F(1, 48) = 5.79$, $MSE = .003$, $p = .020$, $\eta_p^2 = .108$. There was also a numerical tendency for a larger congruency effect on the subsequent trial when the previous trial was congruent (6.7%) than when it was incongruent (4.6%), consistent with the typical pattern of the congruency sequence effect. The interaction between Current Congruency and Previous Congruency, however, was not significant, $F(1, 48) = 3.82$, $MSE = .003$, $p = .057$, $\eta_p^2 = .074$. The Bayes Factor for the comparison between the model with the interaction (but no higher-order interactions involving that interaction) and the matched model without it was $BF_{01} = 2.19$, indicating “anecdotal” evidence for the absence of the interaction.

There was also an interaction between Previous Response Type and Previous Congruency, $F(1, 48) = 4.32$, $MSE = .006$, $p = .043$, $\eta_p^2 = .083$, indicating that responses were less accurate when the previous trial was congruent than when it was incongruent, but only when the previous trial required an immediate rather than a delayed response (in the latter case, the congruency of the previous trial had no effect). There was a numerical tendency for a larger congruency effect on the subsequent trial when the previous trial required an immediate rather than a delayed response, however, the interaction between Current Congruency and Previous Response Type was not significant, $F(1, 48) = 3.76$, $MSE = .004$, $p = .058$, $\eta_p^2 = .073$. No other effect reached significance (all *p*s > .10), including the four-way interaction between Current Congruency, Previous Congruency, Previous Response Type, and Wait Group, $F(1, 48) = .81$, $MSE = .005$, $p = .372$, $\eta_p^2 = .017$, that was significant in the latency data. The Bayes Factor for the comparison between the model with the interaction and the model without it was $BF_{01} = 2.85$, indicating “anecdotal” evidence for the absence of the interaction. This non-significant result parallels Schmidt's (2017) failure to observe significantly different PC effects for transfer items in the short- vs. long-wait group in the error rate data.

Discussion

The purpose of Experiment 1 was to determine whether Schmidt's (2017) delayed-response procedure would have a similar impact on the congruency sequence effect as it did on the PC effect in Schmidt's experiments. Three main findings emerged. First, as expected and consistent with Schmidt's results, we found that the congruency effect was basically gone for delayed responding in the long-wait group. That is, while in the short-wait group there was a congruency effect for both immediate-response context items and delayed-response context items, the congruency effect was eliminated (in the latencies) or greatly reduced (in the error rates) for delayed-response context items in the long-wait group. Hence, there were no temporal differences in this situation which could impact responding on the transfer items.

Second, a congruency sequence effect emerged in most conditions for the transfer items, with a larger congruency effect on the subsequent trial when the previous trial was congruent than when it was incongruent. Notably, similar to Schmidt and Weissman (2014) who also used a prime-probe task with alternating vertical (left/right) and horizontal (up/down) items, this congruency sequence effect was obtained even though there were no repetitions of stimulus features from one trial to the next, a confound known to increase the magnitude of this effect (e.g., Mayr et al., 2003).

Third and most importantly, we found that imposing a long delay for the response to an item not only slowed down responding to that item (unsurprisingly), but it also affected immediate responding to the item appearing on the subsequent trial, albeit in a specific situation. This situation was that in which an immediate-response congruent item was preceded by a delayed-response congruent item requiring a long delay. That is, in the long-wait group, latencies for an immediate response to a congruent item were faster when the previous trial was an immediate-response congruent trial than when it was a delayed-response congruent trial. Notably, the fact that immediate-response congruent trials following a delayed-response congruent trial were not particularly fast produced a smaller-than-normal congruency effect in that situation. That is, the congruency effect on the subsequent trial had the same magnitude regardless of whether the previous delayed-response trial was congruent vs. incongruent. (note 6)

Overall, these results extend to the congruency sequence effect the results obtained by Schmidt (2017) for the PC effect and offer provisional support for the idea proposed by Schmidt and Weissman (2016) that temporal learning, rather than conflict adaptation, may be the source of the congruency sequence effect in situations where other confounds are controlled for. According to this idea, in normal circumstances, the temporal-learning process would produce a congruency sequence effect mainly by speeding up responses to congruent trials following another congruent trial and, hence, inflating the congruency effect in that situation. However, what Schmidt's (2017) delayed-response procedure seems to imply for that situation is that, when a delayed response is required for the previous (congruent) trial, the speed-up for the congruent trials will be lost. The result would thus be a reduced congruency effect in trials following a delayed-response congruent trial and, hence, a reduced congruency sequence effect following delayed-response trials in general. Because in the present experiment the congruency sequence effect following delayed-response trials was eliminated primarily because congruent trials following delayed-response congruent trials were not as fast as they are following immediate-response congruent trials, the present results appear entirely consistent with that account.

A potential problem with this conclusion, however, comes from the fact that in each list of the experiment, as in the lists used in Schmidt's (2017) experiments, participants could learn an association between congruency and response delay. Because response delay for context items was blocked on congruency in Experiment 1, all delayed-response trials were congruent in one list (the congruent-wait list, similar to Schmidt's MC list) whereas all delayed-response trials were incongruent in the other list (the incongruent-wait list, similar to Schmidt's MI list). The implication is that participants in both lists may have used those associations to engage response strategies.

One such response strategy would be to use the congruency of any current trial to anticipate whether a (long) delayed response would be required on that trial. In a list with a strong association between congruent items and delayed responses such as was created in the congruent-wait list in the long-wait group, this strategy would lead participants to hesitate on any congruent trial anticipating that a delayed response to that trial is required even when an immediate response is in fact required (i.e., because the item is a transfer item). As a result, on those (immediate-response) congruent trials for the transfer items, participants may delay their response somewhat because they feel a need to wait momentarily until it is determined that no delayed response is actually required. As a result of this slow-down for congruent transfer items, the congruency effect would diminish for those items. This pattern is indeed the one that we observed for congruent transfer items in the congruent-wait list in the long-wait group.

A similar strategy may be used in a list with a strong association between incongruent items and delayed responses such as was created in the incongruent-wait list in the long-wait group. Here, participants could anticipate on an incongruent trial that a delayed response to that trial would be required even when an immediate response is in fact required (i.e., because that item was a transfer item), potentially delaying the response to those items initially. As a result of this slow-down for incongruent transfer items, the congruency effect would increase for those items. Consistent with this idea, a (small) numerical slow-down for incongruent transfer items did emerge in the incongruent-wait list in the long-wait group (although a (larger) numerical speed-up for congruent items also emerged in this situation which this idea would have difficulty explaining). Thus, because of the blocked nature of the design used, the results of Experiment 1 (and by extension, the results reported by Schmidt (2017), who also used a blocked design) may be interpreted as the outcome of response strategies based on associations between congruency and response delay. This issue was addressed in Experiment 2, where a mixed rather than a blocked design was used.

Experiment 2

The purpose of Experiment 2 was to attempt to replicate the results of Experiment 1 in a situation where participants could not develop response strategies based on associations between congruency and response delay, strategies which could explain the data pattern obtained in Experiment 1 without temporal learning being necessarily implicated. To this end, congruency and response delay for context items were manipulated in a mixed fashion rather than in a blocked fashion. Thus, rather than having one list where all delayed-response context items were congruent and another list where all delayed-response context items were incongruent, a single list was used where both congruent and incongruent immediate- and delayed-response context items appeared with equal probability. The implication is that for the transfer items, each combination of the type of response requested for the previous (context) item (immediate vs. delayed), the congruency of the previous (context) item (congruent vs. incongruent), and the congruency of the subsequent transfer item (congruent vs. incongruent), was equally probable in the list as a whole. Therefore, in this situation, any potential modulation of congruency effects on transfer items based on the nature of the previous (context) item could not be attributed to associations between congruency and response delay because no such associations existed.

Method

<u>Participants</u>

Fifty-six students (18 males) at the University of Western Ontario (age 17–28 years) participated for course credit, a sample size similar to that used in Experiment 1. All participants were native English speakers and had normal or corrected-to-normal vision. None of them had participated in Experiment 1.

Materials

The materials were the same as in Experiment 1. What changed was how those materials were arranged in the two 200-trial blocks comprising the experiment. In this experiment, unlike in Experiment 1, there was no difference between those two blocks. In both blocks, half of the context items were congruent and half were incongruent and, orthogonal to this manipulation, half required an immediate response and half required a delayed response. The transfer items were the same as in Experiment 1, as was the rest of the materials, the presentation of context and transfer items on odd and even trials, respectively, and the counterbalancing. The average frequency of the stimuli in one of the counterbalancing versions of the experiment (across both blocks of trials) is presented in Table 3 (again, the reason that this frequency is an average is because stimuli were selected at random with replacement).

Table 3

Template for the Average Frequency of Prime-Probe Combinations in Experiment 2

			Context primes			
	Transfer primes		Immediate response		Delayed response	
Probes	Left	Right	Up	Down	Up	Down
Left	50	50				
Right	50	50				
Up			25	25	25	25
Down			25	25	25	25

Note. All transfer items required an immediate response.

Procedure

The procedure was the same as in Experiment 1.

Results

Since all participants produced correct responses on more than 70% of the trials, no participant was excluded. There were 28 participants in the short-wait group and 28 participants in the long-wait group. Prior to the analyses, we discarded trials following trials in which an error or a too slow response was made (9.4% of the observations). For both context and transfer items, the analyses were conducted in the same way as in Experiment 1.

As for Experiment 1, we report the analysis for the context items in the Supplementary materials for simplicity. Briefly, the results from these items confirmed that the delayed-response manipulation was effective in this experiment as well. While these items showed a regular congruency effect when an immediate response or a short delay was requested (ranging from 86 to 121 ms in the latencies and from 5.8% to 5.9% in the error rates), this effect was not just reduced but completely eliminated when a long delay was required (i.e., in the long-wait, delayed-response condition, the congruency effect was -1 ms and -0.1% in the latencies and error rates, respectively). For the transfer items, their mean RTs and error rates are presented in Table 4, the mean RTs are presented in Figure 5, and the results of the analyses are reported below.

Table 4

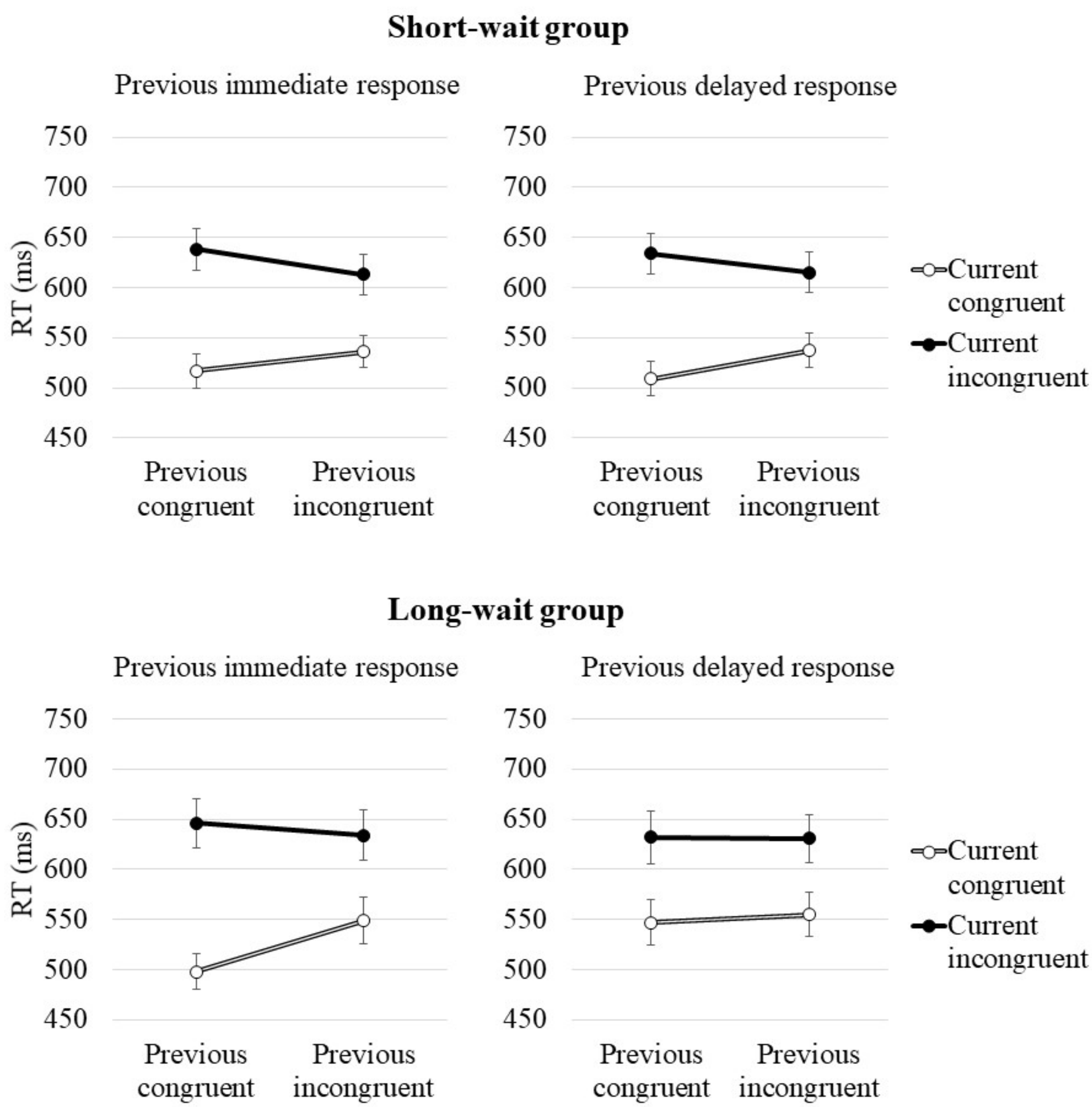
Mean RTs and Error Rates (and Corresponding Standard Errors) for the Transfer Items in Experiment 2

	RTs			Error rates		
Current congruency	Previous immediate response	Previous delayed response	Previous response type effect	Previous immediate response	Previous delayed response	Previous response type effect
<u>Short-wait group</u>						
Previous congruent						
Current congruent	517 (17)	509 (17)	-8	.034 (.007)	.044 (.010)	.010
Current incongruent	638 (21)	634 (20)	-4	.117 (.017)	.118 (.019)	.001
Current congruency effect	121	125	4	.083	.074	-.009
Previous incongruent						
Current congruent	536 (16)	537 (17)	1	.039 (.009)	.041 (.008)	.002
Current incongruent	613 (20)	615 (20)	2	.076 (.014)	.113 (.016)	.037
Current congruency effect	77	78	1	.037	.072	.035
<u>Long-wait group</u>						
Previous congruent						
Current congruent	498 (18)	547 (23)	49	.067 (.014)	.047 (.015)	-.020
Current incongruent	646 (25)	632 (26)	-14	.150 (.026)	.108 (.021)	-.042
Current congruency effect	148	85	-63	.083	.061	-.022
Previous incongruent						
Current congruent	549 (23)	555 (22)	6	.054 (.011)	.049 (.013)	-.005
Current incongruent	634 (25)	631 (24)	-3	.110 (.021)	.095 (.018)	-.015
Current congruency effect	85	76	-9	.056	.046	-.010

Note. The current congruency effect is the difference between responses to incongruent and congruent items on the current trial. The previous response type effect is the difference between (immediate) responses to the current trial preceded by a delayed-response trial and (immediate) responses to the current trial preceded by an immediate-response trial.

Figure 5

Mean RTs (with Error Bars Indicating 1 Standard Error Below and Above the Mean) for the Transfer Items in Experiment 2



<u>RTs</u>

There was a main effect of Current Congruency (congruent faster than incongruent), $F(1, 54) = 204.81$, $MSE = 5389$, $p < .001$, $\eta_p^2 = .791$. There was numerical tendency for overall shorter latencies when the previous trial was congruent than when it was incongruent, however, the main effect of Previous Congruency was not significant, $F(1, 54) = 3.50$, $MSE = 1184$, $p = .067$, $\eta_p^2 = .061$. The main effect of Previous Response Type was not significant, $F(1, 54) = .98$, $MSE = 1428$, $p = .327$, $\eta_p^2 = .018$. The main effect of Wait Group was not significant either, $F(1, 54) = .17$, $MSE = 86035$, $p = .678$, $\eta_p^2 = .003$, reflecting the fact that in this experiment (unlike in Experiment 1), there was no evidence for a general slow-down in the long-wait group compared to the short-wait group. Current Congruency and Previous Congruency interacted, $F(1, 54) = 28.56$, $MSE = 1628$, $p < .001$, $\eta_p^2 = .346$. This interaction indicated a regular congruency sequence effect, with an overall larger congruency effect on the subsequent trial when the previous trial was congruent (122 ms) than when it was incongruent (80 ms). Current Congruency also interacted with Previous Response Type, $F(1, 54) = 6.33$, $MSE = 1175$, $p = .015$, $\eta_p^2 = .105$, indicating an overall smaller Current Congruency effect following delayed-response trials than following immediate-response trials. More importantly, there was a three-way interaction between Current Congruency, Previous Response Type, and Wait Group, $F(1, 54) = 8.66$, $MSE = 1175$, $p = .005$, $\eta_p^2 = .138$, and a four-way interaction between Current Congruency, Previous Congruency, Previous Response Type, and Wait Group, $F(1, 54) = 4.76$, $MSE = 1301$, $p = .033$, $\eta_p^2 = .081$.

To explore these interactions, follow-up analyses were conducted for each wait group separately. In the short-wait group, the only significant interaction was that between Current Congruency and Previous Congruency, $F(1, 27) = 21.04$, $MSE = 1389$, $p < .001$, $\eta_p^2 = .438$, indicating a regular congruency sequence effect, with a larger congruency effect on the subsequent trial when the previous trial was congruent (123 ms) than when it was incongruent (79 ms). Notably, in this group, Previous Response Type had no effect and interacted with no other factors (all *p*s > .20). In the long-wait group, Current Congruency and Previous Congruency interacted as well, $F(1, 27) = 9.62$, $MSE = 1867$, $p = .004$, $\eta_p^2 = .263$, indicating an overall congruency sequence effect, with a larger congruency effect on the subsequent trial when the previous trial was congruent (121 ms) than when it was incongruent (81 ms). Current Congruency also interacted with Previous Response Type, $F(1, 27) = 11.71$, $MSE = 1496$, $p = .002$, $\eta_p^2 = .302$, indicating that, overall, the Current Congruency effect was smaller following delayed-response trials than following immediate-response trials. More importantly, the three-way interaction was significant, $F(1, 27) = 5.74$, $MSE = 1753$, $p = .024$, $\eta_p^2 = .175$.

To interpret this three-way interaction, we split the data of the long-wait group by Previous Response Type. What this analysis showed is that, when the previous trial required an immediate response, Current Congruency and Previous Congruency interacted, $F(1, 27) = 11.14$, $MSE = 2465$, $p = .002$, $\eta_p^2 = .292$, i.e., there was a regular congruency sequence effect, with a larger congruency effect on the subsequent trial when the previous trial was congruent (148 ms) than when it was incongruent (85 ms). In contrast, when the previous trial required a delayed response, although the congruency effect on the subsequent trial was still a little larger when the previous trial was congruent (85 ms) than when it was incongruent (76 ms), Current Congruency and Previous Congruency did not interact, $F(1, 27) = .49$, $MSE = 1156$, $p = .489$, $\eta_p^2 = .018$. The Bayes Factor for the comparison between the model with the interaction and the model without it was $BF_{01} = 3.51$, indicating “moderate” evidence for the absence of the interaction. Thus, a delayed response to the previous trial eliminated the congruency sequence effect on the subsequent trial.

Another way to evaluate the three-way interaction between Current Congruency, Previous Congruency, and Previous Response Type found in the long-wait group is to split the data by Previous Congruency. When the previous trial was congruent, Current Congruency and Previous Response Type interacted,

$F(1, 27) = 16.73$, $MSE = 1618$, $p < .001$, $\eta_p^2 = .383$. This interaction revealed a Previous Response Type effect selectively for (current) congruent trials: Congruent trials following a delayed-response congruent trial were significantly slower (49 ms) than congruent trials following an immediate-response congruent trial, $t(27) = -3.38$, $p = .002$, whereas for incongruent trials, a delayed response to the previous congruent trial led to a nonsignificant speed-up (14 ms) compared to when the previous congruent trial required an immediate response, $t(27) = 1.31$, $p = .200$, $BF_{01} = 2.30$. In contrast, when the previous trial was incongruent, only the Current Congruency effect (congruent faster than incongruent) was significant, $F(1, 27) = 56.48$, $MSE = 3218$, $p < .001$, $\eta_p^2 = .677$. That is, there was no evidence for an effect of Previous Response Type on responses to the subsequent trial, $F(1, 27) = .03$, $MSE = 1621$, $p = .872$, $\eta_p^2 = .001$, $BF_{01} = 4.50$. Thus, as in Experiment 1, it would appear that in the long-wait group, the primary reason why the congruency sequence effect was eliminated following a delayed-response trial was a slow-down for congruent trials when the previous (delayed) trial was also congruent.

Error rates

There were main effects of Current Congruency (congruent more accurate than incongruent), $F(1, 54) = 45.92$, $MSE = .010$, $p < .001$, $\eta_p^2 = .460$, and Previous Congruency (overall less accurate responses following congruent trials than following incongruent trials), $F(1, 54) = 6.04$, $MSE = .003$, $p = .017$, $\eta_p^2 = .101$. Current Congruency and Previous Congruency interacted as well, $F(1, 54) = 4.47$, $MSE = .003$, $p = .039$, $\eta_p^2 = .076$. This interaction once again indicated a regular congruency sequence effect, with an overall larger congruency effect on the subsequent trial when the previous trial was congruent (7.4%) than when it was incongruent (5.2%). The only other effect that reached significance was the interaction between Previous Response Type and Wait Group, $F(1, 54) = 7.59$, $MSE = .004$, $p = .008$, $\eta_p^2 = .123$, indicating that while in the short-wait group, accuracy was, if anything, overall lower following a delayed-response trial than following an immediate-response trial (but not significantly so, $F(1, 27) =$

2.38, *MSE* = .004, *p* = .134, η_p^2 = .081), in the long-wait group responses to the subsequent trial were overall more accurate following a delayed-response trial than following an immediate-response trial, *F*(1, 27) = 5.41, *MSE* = .004, *p* = .028, η_p^2 = .167. No other effect reached significance (all *p*s > .09), including the four-way interaction between Current Congruency, Previous Congruency, Previous Response Type, and Wait Group, *F*(1, 54) = .52, *MSE* = .004, *p* = .372, η_p^2 = .010. The Bayes Factor for the comparison between the model with this four-way interaction and the model without it was BF_{01} = 2.37, indicating "anecdotal" evidence for the absence of the interaction. As in Experiment 1, wait group did not appear to have an influence on the congruency sequence effect in the error rates.

Discussion

The goal of Experiment 2 was to determine whether the results of Experiment 1 would replicate in a situation which did not allow the use of response strategies to anticipate on what type of trial (congruent or incongruent) a delayed response would be required. Because, as noted, those strategies could be used in Experiment 1 (as in Schmidt's (2017) experiments) and could contribute to the elimination of the congruency sequence effect obtained in the long-delay condition in that experiment, it was important to establish that that elimination of the congruency sequence effect could be obtained even in the absence of those strategies.

The results of Experiment 2 were consistent with this idea. First, the delayed-response manipulation was again effective, with a complete elimination of the congruency effect for delayed-response context items in the long-wait group. Second, transfer items showed a congruency sequence effect in most conditions again even though binding of stimulus features could not have played a role in this effect. Third and most importantly, we replicated the critical result of Experiment 1 that, in the latencies, the congruency sequence effect on transfer items was eliminated following delayed-response trials in the long-wait group. Also similar to Experiment 1, this elimination mainly resulted from a slow-down of congruent items following delayed-response congruent trials which caused the congruency effect following delayed-response congruent trials to diminish. In contrast, in this case, performance was essentially unaltered following a delayed-response incongruent trial both when the delay was short and when it was long. Thus, although response strategies may have played a role in Experiment 1 (and in Schmidt's (2017) experiments), they likely were not the factor that produced the pattern of results in those experiments.

Generalized Linear Mixed-Effects Model (GLMM) Re-Analyses

Overall, the results of Experiment 1 and 2 are in general agreement with the temporal-learning account of the congruency sequence effect (Schmidt & Weissman, 2016). As noted, this account assumes that a fast vs. slow temporal expectancy would be formed following congruent vs. incongruent trials, respectively. The fast temporal expectancy would be met by congruent trials but not incongruent trials, leading to a speed-up for the former but not for the latter producing an increased congruency effect. The slow temporal expectancy would be met by incongruent trials but not congruent trials, leading to a (smaller) speed-up for the former but not the latter producing, if anything, a decreased congruency effect. When a delayed response is required on the previous trial, however, a (very) slow temporal expectancy would be formed following both congruent and incongruent trials. As a result, both the speed-up normally observed for congruent trials following another congruent trial and the (smaller) speed-up normally observed for incongruent trials following another incongruent trial would be lost, resulting in the elimination of the congruency sequence effect. Because this pattern of results is precisely the one that we observed in Experiments 1 and 2, these results appear consistent with the temporal-learning account of the congruency sequence effect.

On the other hand, because a delayed-response procedure may dramatically alter the way in which participants perform the task in normal circumstances, these results do not appear to be inconsistent with Botvinick et al.'s (2011) conflict-monitoring account of the congruency sequence effect either, when some additional assumptions are made which take into account the altered version of the task. For example, as noted in the Introduction, it can be assumed that a delayed response may essentially prevent conflict from task-irrelevant information (for incongruent stimuli) from having any real impact during that trial. As a result, the congruency of the delayed-response item would have no impact on subsequent performance (i.e., no congruency sequence effect following a delayed-response trial would be observed). Thus, before concluding that the temporal-learning account provides a better explanation of effects that have typically been considered as support for conflict-monitoring accounts (e.g., the congruency sequence effect), it would seem appropriate to review the evidence from different approaches that have been used to support the temporal-learning account.

As noted in the Introduction, in addition to Schmidt's (2017) delayed-response procedure, two other approaches have been used to investigate temporal-learning accounts. One of these approaches consists of examining situations in which participants experience some sort of difficulty in performing a task, but this difficulty does not derive from processing of irrelevant information. The rationale for this approach is that, from the perspective of the temporal-learning account, any task in which the proportion of easy-to-process versus hard-to-process items is manipulated should produce differences in the magnitude of difficulty effects, differences that parallel those observed for congruency effects in the PC paradigm (i.e., a smaller difficulty effect in a list in which most of the items are hard in comparison to the difficulty effect in a list in which most of the items are easy). Indeed, Schmidt obtained evidence for this proportion-easy effect in a number of studies in which no interfering irrelevant information was presented (Schmidt, 2013b; 2014; 2016; see also Schmidt et al., 2014). For example, in a letter identification task, Schmidt (2013b) manipulated difficulty by using contrast, i.e., comparing high-contrast letters (easy items) with low-contrast letters (hard items). In line with

Schmidt's proposal, there was a larger difficulty effect in a list in which most of the items were easy than in a list in which most of the items were hard, a proportion-easy effect.

Because PC manipulations in interference tasks also involve manipulating the proportion of easy (congruent) and hard (incongruent) items, the possibility exists that the temporal-learning process that underlies the proportion-easy effect in the situations examined by Schmidt is also the process that underlies the PC effect in PC manipulations in interference tasks. Recently, however, we failed to find a proportion-easy effect in a picture naming task in which, similar to Schmidt's experiments, the difficulty of the hard pictures did not derive from the conflict produced by an irrelevant dimension (Spinelli et al., 2019, Experiment 2). Instead, the pattern that we observed was more in line with the literature on blocking effects, a literature in which both easy- and hard-to-process items are typically slower in situations that encourage slow vs. fast temporal expectancies (e.g., Chateau & Lupker, 2003; Lupker et al., 1997; Lupker et al., 2003; Kinoshita & Mozer, 2006; Rastle et al., 2003). Thus, although Schmidt's (2013b) proportion-easy approach does suggest that his conception of temporal learning may contribute to producing PC effects in some situations, it does not inevitably do so (for an explanation, see Schmidt, in press).

The other approach used to support the temporal-learning account, an approach used to address not only the PC effect but also the congruency sequence effect, consisted of re-analyzing data from experiments that produced those effects while controlling for temporal learning. To accomplish this goal, Schmidt and Weissman (2016; Schmidt, 2013b) used linear mixed-effects models (LMMs), a type of statistical analysis that, unlike traditional mean-based ANOVAs, allows the evaluation of trial-level predictors in addition to the typical factors evaluated in congruency sequence and PC paradigms. The relevant additional predictor Schmidt and Weissman used was the (inverse-transformed) latency on the most recent (i.e., immediately previous) trial, which was used as an index of the temporal expectancy that participants had at that point in the task (and, hence, was presumably a reflection of the temporal expectancy driving performance on the subsequent trial). The rationale for this analysis is that, because congruent items (i.e., easy-to-process stimuli) are more likely to benefit from fast temporal expectancies (i.e., following a fast RT) whereas incongruent items (i.e., hard-to-process stimuli) are more likely to benefit from, if anything, slower temporal expectancies (i.e., following a slow RT), the congruency effect on a given trial should be larger following faster responses than following slower responses. That is, if Schmidt's (2013b) temporal-learning process is engaged, the congruency effect should be modulated by temporal expectancies induced by the prior trial(s) and not by the list type (MC vs. MI) or the congruency of the previous trial. Therefore, in these re-analyses, the (inverse-transformed) latency of the previous trial (previous RT) was used as a predictor of the inverse-transformed latency on the subsequent trial in addition to the predictors normally used in analyzing PC and congruency sequence effects. In two such re-analyses Schmidt and Weissman (2016; Schmidt, 2013b) found that the (inverse-transformed) latency of the previous trial not only had a main effect on performance on the subsequent trial (the smaller the inverted RT on the previous trial, the smaller the inverted RT on the subsequent trial or, put another way, the higher the previous RT, the higher the subsequent RT) but it also interacted with the congruency effect on the subsequent trial. That is, although higher previous RT led to a higher subsequent RT for both congruent and incongruent items, the effect of previous RT was stronger for congruent than incongruent items and, as result, the congruency effect decreased the higher the previous RT.

Schmidt and Weissman (2016; Schmidt, 2013b) made the argument that that this two-way interaction is crucial for the temporal-learning account as it implies that temporal expectancies based on the latency of the previous trial have a differential impact on subsequent performance depending on the congruent vs. incongruent nature of the subsequent trial, that is, on whether the subsequent trial allows participants to meet a fast temporal expectancy (typically, when the subsequent trial is congruent) vs. a slow temporal expectancy (typically, when the subsequent trial is incongruent). Interestingly, Schmidt and Weissman (2016; albeit not Schmidt, 2013b) also reported a three-way interaction between previous RT, current congruency, and previous congruency indicating that the two-way interaction (i.e., reduced congruency effects with higher previous RT) held only when the previous item was congruent but not when it was incongruent, a pattern they attributed to the fact that easier items are more sensitive to temporal expectancies than harder items are. (note 7)

Schmidt and Weissman's (2016; Schmidt, 2013b) LMM analysis (see also Kinoshita et al., 2011) represents an interesting approach to examining the role of temporal learning in effects of adaptation to conflict because 1) it addresses those effects in interference tasks directly rather than by use of a parallel situation in a non-conflict task (as is the case for the proportion-easy approach), and 2) it does not involve altering the processes participants normally use in performing those interference tasks (as occurs in Schmidt's (2017) delayed-response procedure). However, as with the proportion-easy approach, this approach has also faced a challenge.

This challenge comes from the fact that Schmidt and Weissman's (2016) LMM analyses were conducted on RT data which were inverse-transformed (invRT = -1000/RT) in order to accommodate the assumption made by those models that the residuals produced by those data be normally distributed. What is crucial to note is that nonlinear transformations of the dependent variable can systematically alter the pattern and size of interaction terms, potentially creating interactive patterns in situations where the raw data show additivity and vice versa (Balota et al., 2013). An inverse transformation, in particular, systematically moves the effect sizes in an underadditive direction in the presence of two main effects (e.g., previous RT and congruency). That is, in conditions in which the overall RTs are higher (e.g., following a slow trial) the relative impact of an inversion on equal effect sizes will be larger than in conditions in which the RTs are lower (e.g., following a fast trial). For example, a 100-ms difference between two long latency conditions, e.g., 700 versus 800 ms, will be reduced to a difference of .00018 by inverting those latencies while a difference of 100-ms between two short latency conditions, e.g., 400 and 500 ms, will be reduced to a difference of .00050 by inverting those latencies, creating an underadditive pattern. Because an underadditive pattern for the congruency by previous RT interaction is a crucial prediction for the temporal-learning account, it seems unlikely that any useful conclusion can be drawn by using a data transformation procedure that inherently biases the results in the predicted direction.

Indeed, in a recent series of re-analyses targeting the temporal-learning account of the PC effect in the color-word Stroop task and the picture-word interference task, Cohen-Shikora et al. (2019) reported that while inverse-transformed RTs in LMMs typically return the regular two-way interaction between previous RT and current congruency (measured in terms of inverse latencies), the same is not true for similar analyses performed with generalized linear mixed-effects models (GLMM), a class of models that does not assume a normally distributed dependent variable and, therefore, requires no RT transformation (Lo & Andrews, 2015). Notably, these GLMM analyses not only failed to show a regular two-way interaction in many cases but also returned a reversed (i.e., overadditive) interaction in some cases, with previous RT affecting subsequent congruent trials *less* than subsequent incongruent trials, with the result being that the subsequent trial congruency effect *increased*, rather than decreased, the higher the previous RT.

A recent study from our lab (Spinelli et al., 2019) confirmed and extended those results. In two picture-word interference tasks which eliminated contingency learning and binding confounds, not only did regular PC and congruency sequence effects arise, there was also no evidence in the GLMM analyses (with raw RTs) for the two-way interaction between previous RT and subsequent trial congruency that the temporal-learning account requires in order to explain those effects (with, again, some evidence for an overadditive interaction in one of the experiments). Thus, when what would appear to be a more appropriate technique than Schmidt and Weissman’s (2016; Schmidt, 2013b) original technique is used for analyzing the impact of temporal learning in effects of adaptation to conflict, there appears to be no real evidence in favor of the temporal-learning processes that they described, at least in Stroop and picture-word interference tasks.

The reason that these observations are relevant is that, in addition to the present Experiments 1 and 2, to our knowledge, Schmidt and Weissman's (2016) LMM analysis is currently the only piece of evidence in support of the temporal-learning account of the congruency sequence effect (the proportion-easy approach has only been used to support the temporal-learning account of the PC effect). What is worth noting is that Schmidt and Weissman's data come from the prime-probe task, the same task used by Schmidt (2017) and in the present experiments. This task differs from the Stroop and picture-word interference tasks examined by Cohen-Shikora et al. (2019) and Spinelli et al. (2019) in a few respects, for example, response modality (vocal in the interference tasks examined by Cohen-Shikora et al. and Spinelli et al. vs. manual in Schmidt and Weissman's prime-probe task) as well as the nature of the context/transfer stimuli (i.e., whereas there typically is no clear difference in Stroop-like tasks between the stimuli assigned to context vs. transfer sets, those stimuli in the prime-probe task do differ in their horizontal/vertical nature, that is, the designated response hand can be easily anticipated due to the odd/even alternating procedure, creating a scenario similar to task-switching; see also Algom & Chajut, 2019).

Although the temporal-learning account appears to describe a task-general mechanism, it does not deny the possibility that temporal-learning processes are actually implemented only in some circumstances (see, e.g., Schmidt, in press). Thus, although Spinelli et al. (2019) found no evidence for a role of temporal learning in the congruency sequence effect (and the PC effect) in the context of a picture-word interference task, the same may not be true for the prime-probe task. That is, it is possible that even when a GLMM analysis on raw RTs, rather than Schmidt and Weissman's (2016; Schmidt, 2013b) LMM analysis on inverse-transformed RTs, is used to examine the impact of temporal learning in the congruency sequence effect in the prime-probe task, the results would still produce evidence in favor of a temporal-learning explanation of this effect. Specifically, the GLMM analysis, like Schmidt and Weissman's LMM analysis, might show that congruency effects on a subsequent trial diminish with higher previous RT, the pattern of results the temporal-learning account of the congruency sequence effect predicts (potentially with an additional role of previous congruency in the present data – although see footnote 7).

On the other hand, in a response to Cohen-Shikora et al. (2019), Schmidt (in press) has recently presented arguments suggesting that inverse-transformed latencies are the most appropriate measure to use in these types of analyses rather than raw RTs for a number of reasons. First, Schmidt argued that, although inverse RTs represent a different dependent variable (i.e., response rate) than raw RTs (i.e., response time), inverse RTs are arguably a superior metric than raw RTs in most research situations (including the present situation) because by correcting for the tendency for effects (and associated variances) to "scale up" at higher latencies, inverse RTs would reveal underlying differences in processing rate, the type of differences that most theoretical accounts concern themselves with.

Second, the raw scale typically produces a smaller (but still significant) Pearson's *r* for the correlation between previous RT and subsequent RT (as well as for the correlation reflecting the interaction between previous RT and congruency) than does the inverse scale, suggesting that the raw scale is unlikely to be the scale where the "true" relationship between previous RT and subsequent RT exists. Schmidt's (in press) reason is that "correlations will necessarily be weaker in a scale that more poorly reflects the true relationship between two variables" (Schmidt, in press, p. 24).

Finally, evaluating the impact of previous RT on subsequent RT in the raw scale is problematic because it involves fitting a line through the combination of two distributions that, being both ex-Gaussian, form a fan-shaped pattern where most of the observations are concentrated in one corner (the corner where the faster previous and subsequent RTs are) and the remaining observations from the slower (previous and/or subsequent) RTs are spread out from that corner in a fan. As a result, “very little weight is given to the bulk of the observations and a very large weight is given to severely-outlying slow RT observations” (p. 22). In contrast, “the temporal learning account does not predict effects to be localized primarily in the extreme right tail of the distribution... Indeed… most of the movement to be around the peak of the response time distribution, not in the right tail” (p. 23).

For these reasons, Schmidt (in press) rejected the idea that GLMMs with raw RTs would represent a fair assessment of his temporal-learning account. A fairer assessment of temporal learning would be offered by another analytical procedure, called “corrected raw RT analysis”. Briefly, this procedure involves extracting the variance associated with temporal learning (without considering the PC effect) on the inverse scale (according to Schmidt, the theory-relevant scale), calculating the resulting residuals on the raw scale, and comparing the PC effect for the residual data vs. the observed data (i.e., the data, also on the raw scale, for which no variance had been extracted in the previous step). What the temporal-learning account predicts, and what Schmidt did report for three datasets, is that a smaller PC effect should be observed in the residual data compared to the observed data because in the former case, but not in the latter, an opportunity was given for temporal learning to explain some of the variance in the theory-relevant scale that might otherwise be “stolen” by the PC effect in the raw scale. The fact that that opportunity resulted in a reduction of the PC effect in the residual data would be proof that temporal learning does contribute to that effect, as the temporal-learning account assumes. Importantly, according to Schmidt, because the assessment of temporal learning is conducted on the inverse scale, this procedure would avoid the problems, reviewed above, associated with conducting this assessment on the raw scale. Further, because the crucial result (i.e., the reduction of the PC effect in the residual data) is evaluated in the raw scale, this procedure would also guard against the problem that the inverse transformation produces a bias towards underadditivity, as evidence for temporal learning would not be confined to the inverse scale but would also “transform out” to the raw scale.

Unfortunately, the arguments presented by Schmidt (in press) are not without problems. First, from the theoretical point of view, the idea that response rate (i.e., inverse RT) is a better metric than response time (i.e., raw RT) is at odds with the fact that, originally (Schmidt, 2013b), the temporal-learning account was presented as a model based on response time, not a model based on response rate. Schmidt (2013b) made this idea quite clear in defining the foundational hypotheses of the account, i.e., the *temporal-coding hypothesis* that “information about the latency between stimulus onset and when a participant responds (i.e., *response time*) might be encoded” [emphasis added] and the related *temporal-learning hypothesis* that that information “is used by participants in an anticipatory way on following trials” (p. 2). That is, what was being encoded and retrieved in that model was the response time associated with past events (i.e., raw RT), not the response rate (i.e., inverse RT). Indeed, the rate of processing was not at all assumed to be influenced by the temporal-learning process (see Figures 1 and 2 above). (note 8)

Independently from how one should model the influence of previous latencies on performance *statistically*, the raw scale would, therefore, seem to be the theoretically most relevant scale for the temporal-learning account, at least in its original version. Schmidt (2013b), however, left open the possibility of temporal learning operating in a different fashion than assumed in his model. Indeed, more recently, Schmidt (in press) has argued that, statistically *and* theoretically, the inverse scale is actually the most relevant for the temporal-learning account because it better captures changes in evidence accumulation across different trials, the true piece of information used in the temporal-learning process. Schmidt (in press), however, did not explain how the contrast between this new position and the original one is resolved (or even acknowledge that indeed there is such a contrast).

Also questionable is the idea that a stronger correlation between previous RT and subsequent RT in the inverse than the raw scale would suggest that it is in the former scale that the “true” relationship between those variables exists. The reason that this idea is questionable is that it is possible that the true relationship does exist in the raw scale but that relationship is not linear. For example, it is reasonable that the positive relationship between previous RT and subsequent RT would be strong at shorter latencies and weaker at longer latencies. In the raw scale, a line would not fit this relationship very well. A curve would be more appropriate, although curvilinear relationships in the raw scale have never been examined in this area of research, in which the focus has exclusively been on linear relationships. However, in terms of linear relationships, a linear relationship would naturally be fit worse in the raw scale than in the inverse scale, a scale in which the differences between fast RTs (where the relationship is presumably stronger) increase and those between slow RTs (where the relationship is presumably weaker) decrease. Thus, in this scenario, a higher Pearson's *r* for the correlation between previous RT and subsequent RT would be produced in the inverse than the raw scale even though the true relationship between those two variables would exist in the latter scale. (note 9)

Further, neither the argument that the inverse scale is theoretically most appropriate nor the argument that that scale offers a better fit addresses the main problem with that scale, i.e., the fact that, in the presence of two main effects, that scale produces a bias for the pattern predicted by the temporal-learning account for the congruency by previous RT interaction, an underadditive pattern. Schmidt's (in press) "corrected raw RT analysis" does not overcome this problem either. The reason is that the crucial control for temporal learning in that procedure occurs in the inverse scale, not the raw scale, and in a situation in which the effect of list type (MC vs. MI) is not modelled. In that scale, for the reasons noted above, it is virtually inevitable that evidence for temporal learning would emerge in the form of an underadditive congruency by previous RT interaction. Further, because the effect of list type is not modelled, that underadditive interaction would explain part of the variance that would be normally associated with list type. The reason is that the two variables are often confounded, i.e., congruency effects tend to be larger in the MC list, a list where previous RTs tend to be fast, whereas congruency effects tend be smaller in the MI list, a list where previous RTs tend to be slow. It follows, then, that later in Schmidt's procedure, the effect of list type would be larger in the observed data, data in which no variance relevant to that effect had been modelled in a previous step, than in the residual data, data in which part of that variance *had* been modelled by the congruency by previous RT interaction in a previous step.

This pattern, however, would not necessarily be observed had the crucial control for temporal learning occurred in the raw scale in the initial step of the procedure. The reason is that there is no mathematical artefact in that scale that would systematically make the congruency by previous RT interaction explain part of the variance associated with the congruency by list type interaction, i.e., the PC effect. Indeed, when reanalyzing the three datasets on which Schmidt conducted his corrected raw RT analysis (i.e., Hutchison, 2011; Bugg, 2014; Gonthier et al., 2016) by conducting the control for temporal learning in the raw rather than the inverse scale (i.e., without using an inverse transformation at any point in Schmidt's "corrected raw RT analysis"), we found no reduction of the PC effect in the residual data in either Hutchison's (2011) dataset, $F(1,224) = .15$, $MSE = 101$, $p = .700$, $\eta_p^2 = .001$, Bugg's (2014), $F(1,70) = 1.09$, $MSE = 45$, $p = .300$, $\eta_p^2 = .015$, or Gonthier et al.'s (2016), $F(1,88) = 1.91$, $MSE = 13$, $p = .171$, $\eta_p^2 = .021$. These results suggest that in Schmidt's (in press) analytical procedure, evidence for temporal learning only seems to emerge when an inverse transformation is applied to the dependent variable initially in the procedure. When no such transformation is applied, that evidence completely vanishes.

These results are important because they make it clear that what is crucial in Schmidt's corrected raw RT analysis is not the scale in which the final test for temporal learning is conducted (i.e., the raw scale), but the scale in which the "correction" is conducted (i.e., in Schmidt's original analysis, the inverse scale). On the other hand, these results are certainly not inconsistent with Schmidt's (in press) temporal-learning account according to which, for the reasons noted above, little or no evidence for temporal learning should emerge in the raw scale because that scale is essentially inappropriate for examining temporal-learning processes. However, they are also fully consistent with the idea that evidence for temporal-learning only emerges in the inverse scale because that evidence is a by-product of a mathematical artefact created by the inverse transformation.

In any case, while these observations point to the weaknesses of using the inverse scale in the present situation, we do not intend to suggest that the raw scale would have no weaknesses at all. In particular, we agree with Schmidt (in press) that, given the ex-Gaussian distribution of raw RTs, it is likely that, in a GLMM in which previous RT is used as a predictor for subsequent RT, observations for which either the previous or the subsequent RT was in the right tail of the respective distribution would have a disproportionate influence on the results. Although an inverse transformation would handle this problem by bringing those slow observations closer to the center of the distribution (and, thus, decreasing their influence on the results), that solution is arguably not ideal, not only for the reasons noted above but also because it would imply maintaining in the analyses observations that Schmidt (in press) considered "severe outliers" when examining the raw scale. After an inverse transformation, those observations would typically not be severe outliers anymore in distributional terms, however, they would still be those observations for which Schmidt claimed that temporal learning should have little or no impact.

A better and more standard solution to reduce the impact of slow observations on the results, however, is to remove those observations from the analyses. A popular procedure to do so is to remove, for each participant, observations that are above or below a certain number of Standard Deviations (SDs, typically 2.5 or 3) from the mean of that participant (Howell, 1998). Evaluating the impact of this procedure in the present situation would appear to be useful because it allows an investigation of the impact of regulating the rejection criterion from quite lenient (e.g., when only observations above or below 4 or 3.5 SDs from the participant's mean are removed) to quite strict (e.g., when all observations above or below 1 or 1.5 SDs from the participant's mean are removed – although very strict criteria are unusual).

What the temporal-learning account seems to predict regarding this procedure is that, in the raw scale, weak or no evidence for temporal learning should emerge when a lenient rejection criterion is used. The reason is that, in that situation, many observations would remain in the right tail of the latency distribution that would have a disproportionate influence on the results, making the test of the temporal-learning account quite noisy. In contrast, it is reasonable to conclude that evidence for temporal learning should emerge more clearly when a strict rejection criterion is used, i.e., when most of the observations in the tails of the distribution, particularly in the right tail, are removed from the analyses and, therefore, prevented from unduly influencing the results. In the latter situation, most of the observations left will be those that were in the peak of the original distribution, where most of the movement relevant to temporal learning should be. The implication would seem to be that, even in the raw scale, the test of temporal learning would certainly be less noisy in this situation than when a more lenient rejection criterion is used. Therefore, evidence for temporal learning in the form of an underadditive congruency by previous RT interaction should start to emerge in the raw scale in this situation, although this scale is normally deemed inappropriate for examining temporal learning in the current version of the temporal-learning account (Schmidt, in press).

Based on the fact that the support for the temporal-learning account's explanation of the congruency sequence effect reported by Schmidt and Weissman (2016) (using the prime-probe task, the task also used in the present experiments) could have been a product of the inverse transformation of the RT data they performed, we conducted two re-analyses using a GLMM analysis (and raw RTs) in an attempt to evaluate the strength of that support. The first re-analysis was a re-analysis of the RT data from the short-wait groups in the present Experiments 1 and 2. The reason that we decided to re-analyze the data for the transfer items from short-wait groups but not long-wait groups is that the former represent a situation close to normal because the delay required for “delayed”-response trials was actually negligible. Thus, it is reasonable to assume that in this situation, participants performed the task as they would have done in a situation in which all responses were immediate. The second re-analysis was a GLMM re-analysis of Schmidt and Weissman’s (2016) dataset, a situation where all responses were, in fact, immediate. Importantly, in both re-analyses, separate GLMMs were conducted for both the full dataset and subsets of the full dataset obtained by applying an increasingly strict rejection criterion (from 4 to 1 SDs above or below each participant’s mean) in order to examine whether evidence for temporal learning would emerge more clearly in the datasets with a stricter rejection criterion.

The reason these re-analyses are important is that, as noted, the results of Experiments 1 and 2, although consistent with the temporal-learning account of the congruency sequence effect, do not provide evidence that can only be explained by that account. However, converging evidence from another approach would certainly strengthen the support for that account. Conversely, a failure to reproduce Schmidt and Weissman’s (2016) pattern of results in the re-analyses would weaken any support that the present data provide for the temporal-learning account.

Method

The full dataset for the re-analysis of the present Experiments 1 and 2 was obtained by collapsing the RT data of the short-wait groups in those experiments. Further, because in the analyses of Experiments 1 and 2 the congruency sequence effect was evaluated for transfer items (the items appearing on even trials), for the sake of consistency, this re-analysis was also based on those items. Thus, data for context items (the items appearing on odd trials) were discarded prior to the re-analysis. (note 10)

The full dataset for the Schmidt and Weissman (2016) re-analysis is the same dataset they used in their own re-analysis of Schmidt and Weissman’s (2014) data. Briefly, this dataset includes the data from Experiments 1 and 2 in Schmidt and Weissman (2014). There were 16 participants for each experiment and each experiment included 768 experimental trials. Both experiments used a prime-probe task, the difference being that the stimuli were arrows in Experiment 1 (e.g., ">" rather than the word "right") and words in Experiment 2 (as in Schmidt, 2017, and in the present experiments). The congruent/incongruent ratio was 50:50 and left/right trials alternated with up/down trials so that there were no feature repetitions across consecutive trials. Congruent-congruent, congruent-incongruent, incongruent-congruent, and incongruent-incongruent sequences were equally probable. All trials required a normal (i.e., immediate) response to the probe. Note that, in these experiments, unlike the present Experiments 1 and 2, there was no distinction between context and transfer items, i.e., the items appearing on odd trials did not serve a different function (e.g., the function of context items) than the items appearing on even trials (e.g., the function of transfer items). Therefore, all trials were analyzed. Readers are invited to consult Schmidt and Weissman (2014, 2016) for further details.

For both datasets, subsets were obtained by applying a rejection criterion ranging from 4 SDs above or below each participant's mean (the most lenient criterion) to 1 SD above or below each participant's mean (the strictest criterion; we decided to stop at 1 SD as using an even stricter criterion would have meant removing more than half of the original observations). For the present Experiments 1 and 2, compared to the full dataset (9328 observations), the 4-SD dataset had .85% fewer observations (9249), the 3.5-SD dataset 1.61% (9178), the 3-SD dataset 2.87% (9060), the 2.5-SD dataset 4.91% (8870), the 2-SD dataset 8.95% (8493), the 1.5-SD dataset 18.72% (7582), and the 1-SD dataset 43.91% (5232). Confirming that this removal procedure deflated the right tail of the distribution, skewness values decreased from 1.37 for the full dataset, to 1.31 for the 4-SD dataset, 1.26 for the 3.5-SD dataset, 1.21 for the 3-SD dataset, 1.07 for the 2.5-SD dataset, .95 for the 2-SD dataset, .88 for the 1.5-SD dataset, and .84 for the 1-SD dataset (note that positive values suggest a right skew; this skew was reduced in narrower datasets but was still present since skewness did not reach zero). For Schmidt and Weissman's (2014) Experiments 1 and 2, compared to the full dataset (21652 observations), the 4-SD dataset had

1.10% fewer observations (21414), the 3.5-SD dataset 1.68% (21288), the 3-SD dataset 2.73% (21061), the 2.5-SD dataset 4.44% (20961), the 2-SD dataset 7.68% (19990), the 1.5-SD dataset 17.19% (17930), and the 1-SD dataset 42.95% (12353). In this case as well, skewness values decreased, from 1.75 for the full dataset, to 1.57 for the 4-SD dataset, 1.53 for the 3.5-SD dataset, 1.42 for the 3-SD dataset, 1.21 for the 2.5-SD dataset, 1.07 for the 2-SD dataset, .97 for the 1.5-SD dataset, and .88 for the 1-SD dataset.

Results

For the re-analysis of Experiments 1 and 2, the data treatment was the same as reported above. For the re-analysis of Schmidt and Weissman's (2016) data, the data treatment was the same treatment that they used. Specifically, only correct RTs slower than or equal to 300 ms preceded by correct RTs slower than or equal to 300 ms were used. (note 11) Importantly, however, raw RTs, rather than inverse-transformed RTs, on the current trial were used as the dependent variable in the present re-analyses. (note 12) For both datasets, subjects and items (i.e., the target probe) were regarded as random effects and Current Congruency (current congruent vs. current incongruent), Previous Congruency (previous congruent vs. previous incongruent), and Previous RT were regarded as within-subject fixed effects (Baayen, 2008; Baayen et al., 2008). Separate analyses were conducted for the full datasets and their subsets.

Prior to running the model in R (R Core Team, 2020), R-default treatment contrasts were changed to sum-to-zero contrasts (i.e., contr.sum) to help interpret lower-order effects in the presence of higher-order interactions (Levy, 2014; Singmann & Kellen, 2019). The model was fit by maximum likelihood with the Laplace approximation technique. The lme4 package, version 1.1-23 (Bates et al., 2015), was used to run the GLMM and obtain probability values. A Gamma distribution was used to fit the raw RTs, with an identity link between fixed effects and the dependent variable (Lo & Andrews, 2015). Model estimation was performed using the BOBYQA optimizer and a maximum of 1 million iterations, settings which typically return equivalent estimates as the default settings but with fewer false-positive convergence warnings (Bolker, 2020). For each dataset, Previous RT was standardized (i.e., centered and scaled) in order to avoid spurious correlations between the intercept and the slope and to help evaluating and interpreting the model (Bolker, 2020; Kinoshita et al., 2011; Schielzeth, 2010). In the following, therefore, Previous RT stands for standardized previous RT.

For these analyses as well, in addition to traditional null-hypothesis significance testing analyses, we also performed Bayes Factor analyses for theoretically relevant non-significant effects in order to quantify the evidence for the absence vs. presence of those effects. These analyses were performed in R (rather than in JASP as was done for the ANOVAs conducted for Experiments 1 and 2 above) using the Bayesian information criterion (BIC) approximation of the Bayes factor (Wagenmakers, 2007). Specifically, the Bayes factor BF_{01} was computed using the BIC value obtained for the model containing the effect of interest, but no interactions with that effect (interpreted as the alternative hypothesis H_1), and the BIC value obtained for the equivalent model stripped of that effect (interpreted as the null hypothesis H_0) using the formula $BF_{01} = \exp((\mathrm{BIC}(H_1) - \mathrm{BIC}(H_0))/2)$ (Wagenmakers, 2007, p. 796; see also Colombo et al., 2020). $BF_{01} < 1$ would suggest evidence in support of H_1 (i.e., the presence of the effect), whereas $BF_{01} > 1$ would suggest evidence in support of H_0 (i.e., the absence of the effect; $BF_{01} = 1$ would suggest equal evidence for the two hypotheses.

<u>Experiments 1 & 2</u>

The results for the full dataset and its subsets are reported in Table 5. Here, we focus on the most relevant results. First, all datasets showed main effects of Current Congruency (current congruent faster than current incongruent), Previous Congruency (indicating that a previous congruent trial tended to slow down latencies overall), and Previous RT (slower responses with higher Previous RT). Note, however, that in all datasets except for the 1.5-SD and 1-SD datasets the means actually showed faster latencies when the previous trial was congruent than when it was incongruent (e.g., 572 and 577 ms, respectively, in the full dataset). Therefore, the fact that a Previous Congruency effect was obtained in the opposite direction for all datasets is at least partially an artifact of the inclusion of Previous RT, a highly correlated predictor, in the model. Consistent with this idea, for all datasets, the Previous Congruency effect disappeared when Previous RT was removed from the model.

More importantly, all datasets also showed a two-way interaction between Current Congruency and Previous Congruency, indicating a regular congruency sequence effect: Current Congruency had a larger effect when the previous trial was congruent than when the previous trial was incongruent.

The two-way interaction between Current Congruency and Previous RT was also significant in all datasets. However, contrary to what would be expected from the temporal-learning account, this two-way interaction indicated that the Current Congruency effect tended to increase, rather than decrease, the higher the Previous RT. This overadditive pattern is represented in Figure 6. In this Figure, for each dataset, the regression lines representing the impact of the standardized previous RT (i.e., the predictor used in the GLMMs) on the current RT for congruent and incongruent items are presented on top of a scatterplot of the raw observations. These graphs indicate that although latencies tended to increase with higher previous RT (as reflected by the fact that regression lines for congruent and incongruent items go up in the graph), this tendency was stronger for incongruent than for congruent items, resulting in larger, rather than smaller, Current Congruency effects the higher the Previous RT.

Crucially, this overadditive pattern was observed not only for datasets with a more lenient rejection criterion, datasets for which, as Schmidt (in press) noted, the scatterplot showed the typical fan shape where observations in the right tail of the latency distribution, rather than the observations around the peak of the distribution, could have a major role in determining the slope of the regression lines. Instead, this pattern of results was also observed for datasets with a stricter rejection criterion, particularly the 1.5-SD and 1-SD datasets which retained slightly more than four fifths and slightly more than half, respectively, of the original observations. Although these datasets were arguably less fan-shaped and more oval-shaped than the others and contained mostly observations that were around the peak of the original distributions, there still was no evidence in those datasets for the underadditive pattern that one would expect based on the temporal-learning account. To the contrary, although the coefficients for the Current Congruency by Previous RT interaction were reduced (i.e., closer to zero) in those datasets (as were the coefficients for most of the other fixed effects; see Table 5), that interaction was still significantly overadditive.

Table 5

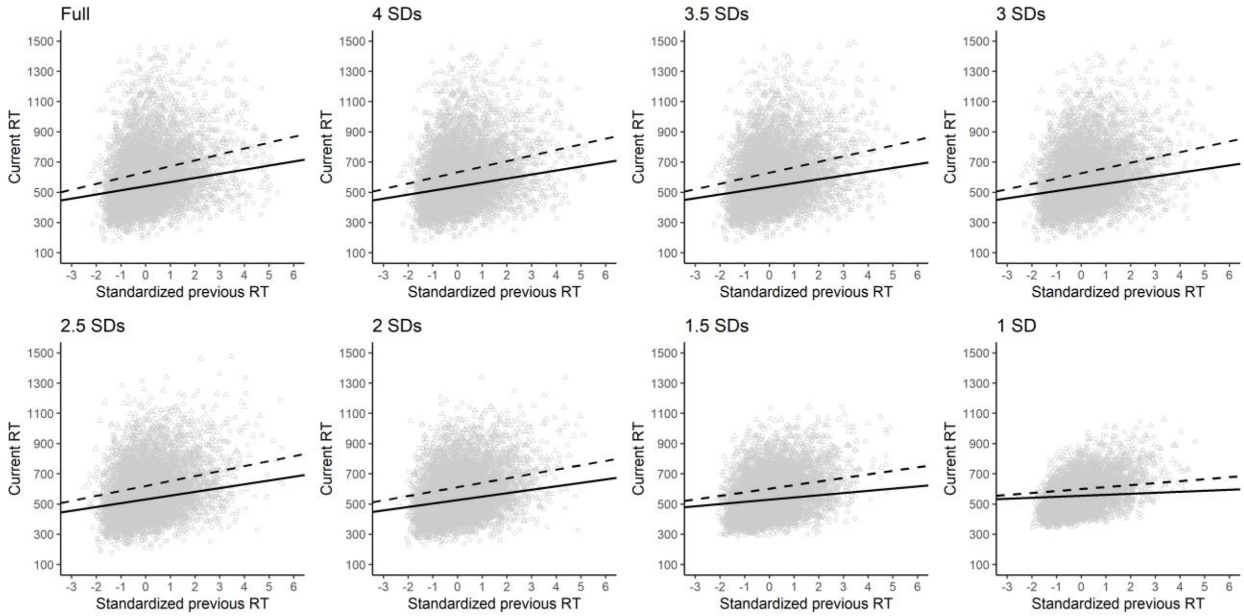
Coefficients, standard errors, z *and* p *values for the full dataset of the Present Experiments 1 and 2 and its subsets*

	Dataset															
	Full				4 SDs				3.5 SDs				3 SDs			
Effect	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*
Intercept	586.75	4.03	145.57	< .001	584.29	5.23	111.67	< .001	582.22	3.95	147.31	< .001	579.29	5.71	101.47	< .001
Current Congruency	47.01	1.27	36.94	< .001	47.16	1.26	37.48	< .001	46.67	1.23	38.06	< .001	46.61	1.20	38.92	< .001
Previous Congruency	-6.12	1.30	-4.69	< .001	-5.99	1.27	-4.71	< .001	-5.68	1.24	-4.58	< .001	-5.77	1.24	-4.63	< .001
Previous RT	33.09	1.59	20.78	< .001	31.77	1.55	20.47	< .001	30.66	1.57	19.58	< .001	29.62	1.56	18.99	< .001
Current Congruency × Previous Congruency	-11.17	1.28	-8.70	< .001	-10.05	1.28	-7.88	< .001	-9.73	1.26	-7.74	< .001	-9.72	1.22	-7.95	< .001
Current Congruency × Previous RT	**5.72**	**1.39**	**4.13**	**< .001**	**5.24**	**1.44**	**3.64**	**< .001**	**5.55**	**1.38**	**4.02**	**< .001**	**5.45**	**1.31**	**4.15**	**< .001**
Previous Congruency × Previous RT	-2.48	1.47	-1.68	.092	-1.97	1.42	-1.39	.163	-1.36	1.36	-1.00	.316	-1.68	1.28	-1.31	.190
Current Congruency × Previous Congruency × Previous RT	4.40	1.37	3.22	.001	4.54	1.40	3.23	.001	4.99	1.39	3.60	< .001	4.79	1.32	3.63	< .001
	2.5 SDs				2 SDs				1.5 SDs				1 SD			
Effect	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*
Intercept	574.52	4.16	137.97	< .001	568.60	4.96	114.63	< .001	564.76	4.28	131.85	< .001	577.02	6.60	87.48	< .001
Current Congruency	44.65	1.21	37.00	< .001	42.44	1.16	36.64	< .001	36.52	1.05	34.87	< .001	21.95	1.00	21.94	< .001
Previous Congruency	-5.91	1.23	-4.79	< .001	-5.74	1.12	-5.12	< .001	-6.08	1.05	-5.79	< .001	-3.17	1.01	-3.15	.002
Previous RT	28.91	1.49	19.37	< .001	25.98	1.48	17.59	< .001	19.08	1.38	13.82	< .001	9.79	1.49	6.59	< .001
Current Congruency × Previous Congruency	-9.36	1.19	-7.89	< .001	-8.41	1.10	-7.65	< .001	-7.16	1.06	-6.76	< .001	-4.55	1.00	-4.56	< .001
Current Congruency × Previous RT	**3.94**	**1.30**	**3.03**	**.002**	**3.14**	**1.18**	**2.66**	**.008**	**4.44**	**1.12**	**3.95**	**< .001**	**3.15**	**1.05**	**3.01**	**.003**
Previous Congruency × Previous RT	-1.27	1.28	-.99	.322	-.95	1.21	-.79	.432	-1.83	1.15	-1.59	.111	.09	1.05	.09	.928
Current Congruency × Previous Congruency × Previous RT	4.27	1.28	23.35	< .001	3.24	1.18	2.75	.006	3.74	1.10	3.42	< .001	-.05	1.05	-.05	.960

Note. For the full dataset, no observations were removed. For the other datasets, the name of the dataset indicates the number of Standard Deviations (SDs) above or below each participant's mean that were used as a rejection criterion. Previous RT stands for standardized previous RT. The Current Congruency by Previous RT interaction is in bold because it is the most relevant effect for the temporal-learning account.

Figure 6

The Impact of Standardized Previous RT on Current Congruency in the Present Experiments 1 and 2



Note. The scatterplots represent the relation between Previous RT and Current Congruency for the full dataset and its subsets. For the full dataset, no observations were removed. For the other datasets, the name of the dataset indicates the number of Standard Deviations (SDs) above or below each participant's mean that were used as a rejection criterion. Individual observations for congruent and incongruent trials are marked with triangles and circles, respectively. Regression slopes for current congruent and current incongruent conditions are marked with solid and dashed lines, respectively.

Interestingly, in all datasets except for the 1-SD dataset, the three-way interaction between Current Congruency, Previous Congruency, and Previous RT was also significant. What this interaction indicated is that, for those datasets, the overadditive pattern for the two-way interaction between Current Congruency and Previous RT was mainly driven by trials preceded by incongruent trials. In that situation, Current Congruency and Previous RT always interacted, all *p*s $< .001$, with the effect of Current Congruency being larger when Previous RT was higher. In contrast, for trials preceded by congruent trials, Current Congruency and Previous RT did not interact, all *p*s $\geq .599$, with the Bayes Factor for the comparison between the model with the interaction and the model without it, $BF_{01} \geq 60.23$, always suggesting "very strong" evidence for the absence of the interaction. As noted, however, this pattern of results was not observed for the 1-SD dataset (see Table 5). For that dataset, there was no evidence for the three-way interaction between Current Congruency, Previous Congruency, and Previous RT, with the Bayes Factor for the comparison between the model with the interaction and the model without it, $BF_{01} = 72.24$, suggesting "very strong" evidence for the absence of the interaction.

Schmidt and Weissman (2016)

The results for the full dataset and its subsets are reported in Table 6. Again, here, we focus on the most relevant results. Similar to the results reported above for the present Experiments 1 and 2, there were main effects of Current Congruency (current congruent faster than current incongruent), Previous Congruency (indicating that a previous congruent trial tended to slow down latencies overall), and Previous RT (slower responses with higher Previous RT) in all datasets. Note, however, that in this case as well, in all datasets except for the 1.5-SD and 1-SD datasets the means actually showed minimally faster latencies when the previous trial was congruent than when it was incongruent (e.g., 506 and 507 ms, respectively, in the full dataset). Therefore, the fact that a Previous Congruency effect was obtained in the opposite direction for all datasets is at least partially an artifact of the inclusion of Previous RT in the model. Consistent with this idea, for all datasets except for the 1.5-SD and 1-SD datasets, the Previous Congruency effect disappeared when Previous RT was removed from the model.

All datasets also showed a two-way interaction between Current Congruency and Previous Congruency, indicating a regular congruency sequence effect: Current Congruency had a larger effect when the previous trial was congruent than when the previous trial was incongruent. The two-way interaction between Current Congruency and Previous RT was also significant in all datasets, however, the pattern of this interaction was the opposite compared to what originally reported by Schmidt and Weissman (2016) for this dataset: The effect of Current Congruency tended to increase, rather than decrease, the higher the Previous RT. This overadditive pattern is represented in Figure 7. In this Figure, similar to the results reported above for the present Experiments 1 and 2, latencies tended to increase with higher previous RT for both congruent and incongruent items in all datasets, but more so for the latter than for the former. Further, this tendency, present in the wider (fan-shaped) datasets, was present also in the narrower (less fan-shaped) datasets, although in the latter the coefficients for the Current Congruency by Previous Congruency interaction (as for the other fixed effects) were reduced.

Finally, also similar to the results for the present Experiments 1 and 2, all datasets except for the 1-SD dataset showed a three-way interaction between Current Congruency, Previous Congruency, and Previous RT. In this case as well, this interaction indicated that, for those datasets, there was an overadditive interaction between Current Congruency and Previous RT when the previous trial was incongruent, all *p*s < .001. In contrast, when the previous trial was congruent, there was no interaction between Current Congruency and Previous RT, all *p*s ≥ .363, with the Bayes Factor for the comparison between the model with the interaction and the model without it, $BF_{01} \geq 69.40$, always suggesting "very strong" or "extreme" evidence for the absence of the interaction. As noted, however, this pattern of results was not observed for the 1-SD dataset (see Table 6), for which there was no evidence for the three-way interaction between Current Congruency, Previous Congruency, and Previous RT, and for which the Bayes Factor for the comparison between the model with the interaction and the model without it, $BF_{01} = 97.56$, suggested "very strong" evidence for the absence of the interaction. (note 13)

Table 6

Coefficients, standard errors, z *and* p *values for the full dataset of Schmidt and Weissman's (2016) Experiments 1 and 2 and its subsets*

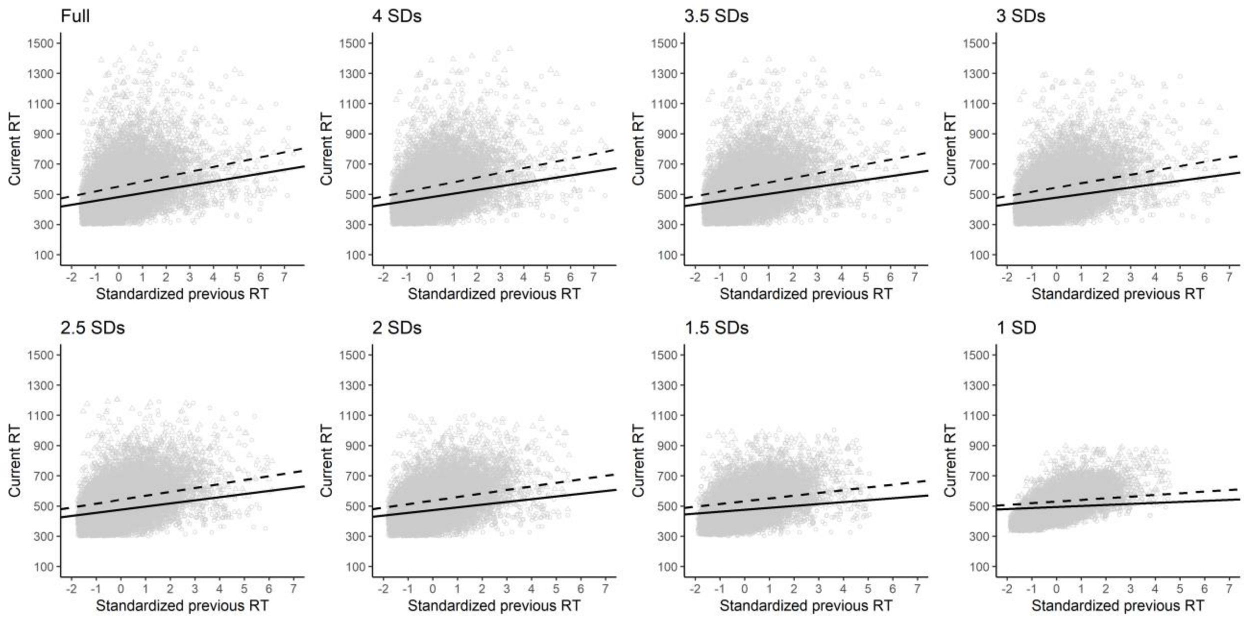
	Dataset															
	Full				4 SDs				3.5 SDs				3 SDs			
Effect	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*P*	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*
Intercept	516.93	2.24	230.28	< .001	514.50	1.83	280.64	< .001	513.42	3.19	160.76	< .001	511.40	3.07	166.51	< .001
Current Congruency	34.65	.60	57.79	< .001	34.48	.56	61.99	< .001	34.28	.55	62.03	< .001	33.61	.54	62.02	< .001
Previous Congruency	-7.35	.62	-11.94	< .001	-7.52	.55	-13.70	< .001	-7.29	.58	-12.61	< .001	-6.95	.55	-12.65	< .001
Previous RT	29.05	.83	35.13	< .001	27.60	.74	37.30	< .001	26.80	.76	35.11	< .001	25.22	.75	33.52	< .001
Current Congruency × Previous Congruency	-5.22	.59	-8.86	< .001	-5.65	.55	-10.20	< .001	-5.60	.56	-10.10	< .001	-5.40	.54	10.07	< .001
Current Congruency × Previous RT	**3.32**	**.66**	**5.01**	**< .001**	**3.41**	**.60**	**5.67**	**< .001**	**3.45**	**.62**	**5.60**	**< .001**	**2.98**	**.59**	**5.01**	**< .001**
Previous Congruency × Previous RT	-.74	.66	-1.12	.263	-1.07	.61	-1.76	.079	-1.17	.63	-1.85	.065	-.87	.61	-1.43	.154
Current Congruency × Previous Congruency × Previous RT	3.31	.67	4.30	< .001	2.54	.62	4.08	< .001	2.62	.63	4.13	< .001	2.80	.59	4.71	< .001
	2.5 SDs				2 SDs				1.5 SDs				1 SD			
Effect	*β*	*SE*	*z*	*P*	*β*	*SE*	*z*	*P*	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*
Intercept	508.57	2.82	180.34	< .001	504.46	3.37	149.53	< .001	503.08	2.27	221.78	< .001	511.68	4.25	120.29	< .001
Current Congruency	32.81	.54	59.83	< .001	32.14	.51	63.44	< .001	28.46	.47	60.27	< .001	17.91	.45	10.19	< .001
Previous Congruency	-6.69	.55	-12.21	< .001	-6.16	.51	-11.97	< .001	-5.60	.50	-11.25	< .001	-3.44	.46	-7.51	< .001
Previous RT	23.34	.75	31.15	< .001	20.74	.71	29.25	< .001	15.41	.70	22.14	< .001	8.88	.77	11.58	< .001
Current Congruency × Previous Congruency	-5.09	.54	-9.40	< .001	-5.09	.51	-10.04	< .001	-3.90	.48	-8.20	< .001	-2.07	.44	-4.68	< .001
Current Congruency × Previous RT	**2.66**	**.59**	**4.51**	**< .001**	**2.64**	**.55**	**4.83**	**< .001**	**2.76**	**.51**	**5.40**	**< .001**	**2.07**	**.46**	**4.47**	**< .001**
Previous Congruency × Previous RT	-.89	.59	-1.53	.126	-.22	.55	-.39	.695	-.78	.52	-1.52	.129	-.61	.47	-1.32	.189
Current Congruency × Previous Congruency × Previous RT	3.29	.58	5.68	< .001	2.89	.55	5.23	< .001	2.35	.51	4.64	< .001	.24	.46	.51	.608

Note. For the full dataset, no observations were removed. For the other datasets, the name of the dataset indicates the number of Standard Deviations (SDs) above or below each participant's mean that were used as a rejection criterion. Previous RT stands for standardized previous RT. The Current Congruency by Previous RT interaction is in bold because it is the most relevant effect for the temporal-learning account. For the

3.5-SD and 1.5-SD datasets, the results are from the GLMM restarted from the apparent optimum reached in the initial estimation, which failed to converge (as per the recommended troubleshooting procedures; see “convergence” help page in R).

Figure 7

The Impact of Standardized Previous RT on Current Congruency in Schmidt and Weissman's (2016) Experiments 1 and 2



Note. The scatterplots represent the relation between Previous RT and Current Congruency for the full dataset and its subsets. For the full dataset, no observations were removed. For the other datasets, the name of the dataset indicates the number of Standard Deviations (SDs) above or below each participant's mean that were used as a rejection criterion. Individual observations for congruent and incongruent trials are marked with triangles and circles, respectively. Regression slopes for current congruent and current incongruent conditions are marked with solid and dashed lines, respectively.

Discussion

The results of the re-analyses of the present Experiments 1 and 2 and Schmidt and Weissman's (2016) dataset, re-analyses conducted with GLMMs using raw RTs and an increasingly strict rejection criterion for removing observations in the tails of the latency distribution, are straightforward. First, in all datasets, there was a regular congruency sequence effect, with a larger congruency effect when the previous trial was congruent than when it was incongruent. Second, in all datasets, as is often reported (e.g., Cohen-Shikora et al., 2019; Kinoshita et al., 2011), previous RT had an overall impact on subsequent performance, with slower RTs on the previous trial leading to slower RTs on the subsequent trial. Third, and most importantly, there was no evidence in any of the datasets that the effect of previous RT was stronger for congruent items than for incongruent items, the crucial underadditive two-way interaction for the temporal-learning account which would lead to smaller congruency effects on the subsequent trial the higher the previous RT. Instead, in all datasets, Current Congruency and Previous RT were overadditive, with the effect of previous RT being stronger for incongruent items than for congruent items, resulting in larger, rather than smaller, congruency effects on the subsequent trial the higher the previous RT.

Examining the coefficients, this overadditive interaction became weaker in narrower datasets. This weakening was to be expected under the temporal-learning account because, for the overadditive interaction to become underadditive, it is reasonable that the overadditive pattern would gradually diminish in strength first. Note, however, that this weakening could also be due to the fact that the increase in congruency effects on the subsequent trial with higher previous RT was simply more pronounced at longer latencies, latencies that are missing from narrower datasets. More importantly, however, the interaction between Current Congruency and Previous RT was still significantly overadditive even in the 1-SD datasets, the narrowest datasets, in which a reversal of the overadditive pattern (i.e., an underadditive pattern) was to be expected under the temporal-learning account.

Further, except for the 1-SD datasets, this overadditive pattern was mainly driven by trials preceded by incongruent trials. For trials preceded by congruent trials, the trend, if anything, was for Congruency and Previous RT to be additive.

Note that in Schmidt and Weissman's (2016) LMM analysis using inverse-transformed RTs, the scenario was the complete opposite: They found an underadditive relationship between congruency on the subsequent trial and previous RT (congruency effects on the subsequent trial decreased with higher previous RT) when the previous trial was congruent and an additive relationship when the previous trial was incongruent. As such, the contrast between Schmidt and Weissman's analysis and the present re-analyses is particularly instructive of how large the mismatch between raw data and nonlinear transformations of those data can be for interaction terms.

More specifically, the present re-analyses are compatible with two ideas. The first idea is that, as in other situations (Cohen-Shikora et al., 2019), Schmidt and Weissman's (2016) pattern of results (particularly the interactions that were supposed to support a role of temporal learning in the congruency sequence effect) was likely produced by the RT transformation they used. The second, alternative, idea is that Schmidt and Weissman's pattern of results reflect a temporal-learning effect that can only ever be captured in inverse RTs. In any case, because no evidence of that effect emerged in the present re-analyses even when a very strict rejection criterion was applied to prevent observations in the tails of the distribution, particularly the right tail, from unduly affecting the results, the temporal-learning account of the congruency sequence effect appears to be somewhat less convincing than Schmidt and Weissman have suggested. (note 14)

General Discussion

The Congruency Sequence Effect Disappears After Delayed Responses: Evidence for Temporal Learning

In recent years, there has been increasing research interest in the idea that effects traditionally attributed to conflict-induced control, effects such as the PC effect and the congruency sequence effect, might actually reflect more general learning processes that are not directly related to conflict (Schmidt, 2013a, 2019). One such process is temporal learning, a process that involves learning and using temporal expectancies for response emission based on the latencies at which responses were emitted on the previous trials, particularly the most recent trial (Schmidt, 2013b; Schmidt & Weissman, 2016). Because temporal expectancies are typically slower in situations thought to tighten control (e.g., following an incongruent trial) than in situations thought to relax control (e.g., following a congruent trial), a temporal-learning process would potentially be able to explain the fact that PC and congruency sequence effects are observed when those situations are contrasted without invoking adaptation to conflict as an explanation. The implication is that those effects would not be observed if a manipulation were used that eliminated the typical differences in temporal expectancies existing between those situations.

Using a delayed-response procedure to implement such a manipulation in the context of a PC paradigm in the prime-probe task, Schmidt (2017) obtained evidence consistent with this idea: A regular PC effect was observed for transfer items (items which were identical in lists differing in conflict frequency) in a situation in which an immediate or a virtually immediate response was required for the context items (the items that differed across the two lists, being mainly congruent in the MC list and mainly incongruent in the MI list). This effect was, however, eliminated in a situation in which a delayed response was imposed on the list-defining context items (i.e., congruent context items in the MC list and incongruent context items in the MI list) in order to substantially slow down and equate temporal expectancies across the two lists. Based on these results, Schmidt suggested that, when other non-conflict learning processes are controlled for, the PC effects one obtains likely result from differences in temporal expectancies between MC and MI lists rather than differences in conflict frequency between the two lists.

In the present research, we first addressed the question of whether Schmidt's delayed-response procedure in the prime-probe task would have a similar impact on the congruency sequence effect as it did on the PC effect. Such a result is one that the temporal-learning account would appear to predict based on the assumption that, in unconfounded experiments, the PC effect and the congruency sequence effect should be produced by the same temporal-learning process. Consistent with this idea, in two experiments, the congruency sequence effect was eliminated in similar situations as those that produced the elimination of the PC effect in Schmidt's experiments. That is, while a regular congruency sequence effect emerged for transfer items when they were preceded by context items requiring an immediate or a virtually immediate response, this effect was eliminated when the transfer items were preceded by context items requiring a delayed response.

Notably, this elimination mainly resulted from a slow-down of congruent items following delayed-response congruent items, a pattern of results which is consistent with the idea that, in normal circumstances, congruent-congruent sequences are the ones that benefit the most from temporal learning (Schmidt, 2013b; Schmidt & Weissman, 2016). Further, the same pattern of results was observed both in an experiment where, like in Schmidt's (2017) experiments, the context allowed participants to use the trial congruency to anticipate whether a delayed response would be required (Experiment 1) and in an experiment which did not create such contextual expectations (Experiment 2). Thus, the observed elimination of the congruency sequence effect following delayed-response trials was likely the result of the delayed-response procedure itself and not just the result of response strategies about delayed responding that some situations could allow to develop.

Overall, these results appear to be in general agreement with the idea that temporal expectancies play a key role in the congruency sequence effect, in that when differences in temporal expectancies following congruent vs. incongruent trials are eliminated by the use of a delayed-response procedure, so is the congruency sequence effect itself. Extending the results obtained by Schmidt (2017) for the PC effect to the congruency sequence effect, these results would, therefore, appear to offer strong support for the idea that temporal learning, rather than conflict adaptation, may be the source of effects traditionally attributed to conflict-induced control.

Congruency Effects Do Not Diminish After Regular Slow Responses: No Evidence for Temporal Learning

A different story emerges from the GLMM re-analyses of Experiments 1 and 2 and Schmidt and Weissman's (2016) dataset. The purpose of these re-analyses was to examine the role of temporal learning in the congruency sequence effect in a normal situation, i.e., a situation in which temporal expectancies were not directly manipulated by means of a delayed-response procedure. According to the temporal-learning account (Schmidt, 2013b; Schmidt & Weissman, 2016), temporal expectancies for a trial should be faster when the latency of the previous trial was also fast, a situation that would allow a speed-up selectively for congruent items, and hence, an increased congruency effect when the previous trial is a congruent trial. Congruency effects on the subsequent trial should, therefore, decrease the higher the previous RT.

Although Schmidt and Weissman (2016) did obtain evidence for such a pattern in a congruency sequence paradigm in the prime-probe task, the inverse transformation that they applied to the RT data in their LMM analysis might have been responsible for producing those results (Balota et al., 2013; Cohen-Shikora et al., 2019; Lo & Andrews, 2015). Using GLMM analyses that do not require an RT transformation, we re-assessed the role of temporal learning in the congruency sequence effect in a less biased situation, i.e., one where raw RTs were used even though, more recently, Schmidt (in press) made several arguments against the raw scale being appropriate for examining temporal learning. Taking into account one of those arguments, i.e., that GLMMs with raw RTs are vulnerable to observations in the tails of the distribution (especially the right tail) that could distort a genuine underadditive relationship between congruency and previous RT, we combined GLMM analyses with an increasingly strict criterion to remove those observations. Interestingly, not only did we find no evidence for the crucial temporal-learning pattern in either our experiments or Schmidt and Weissman's (2016) original dataset, not even when the strictest rejection criterion was used, but we also found evidence for the reversed pattern, although mainly when the previous trial was incongruent: Congruency effects on the subsequent trial increased, rather than decreased, the higher the previous RT.

Therefore, overall, our GLMM analyses present what could be considered an ambiguous pattern, one that is compatible with, on one hand, the idea that there is no true temporal-learning effect and that evidence for that effect only emerges when introducing a bias into the analyses by inverse-transforming latencies. On the other hand, one could argue that the pattern of results that we obtained is also compatible with Schmidt's (in press) idea that a true temporal-learning effect does exist but that effect could only be reliably captured in the inverse scale, because according to that version of the temporal-learning account, it is the response rate, not the response time, on which the temporal-learning mechanism is based (but see Schmidt, 2013b). In sum, the evidence in favor of the temporal-learning account of the congruency sequence effect offered by the present research appears, at best, mixed.

Can the Conflict-Monitoring Account Explain Our Results from the Delayed-Response Procedure?

Even though the present research presents mixed evidence regarding the processes underlying the congruency sequence effect, it would seem that the temporal-learning account is presently (Schmidt, in press) better equipped to explain the pattern of results from the delayed-response procedure than control-based accounts (in particular Botvinick et al.'s (2001) conflict-monitoring account) are. Because, however, those control-based accounts were not built to address the manipulations and analyses conducted in the present research, a relevant question is how, and to what extent, those accounts could be modified to do so.

Indeed, reiterating what was noted by Schmidt (2017) in discussing his results, although in the present Experiments 1 and 2 the delayed-response procedure produced a pattern of results which is consistent with the temporal-learning account of the congruency sequence effect, that account is likely not the only one able to explain that pattern of results. Although that explanation would involve making assumptions not included in the original conflict-monitoring theory (Botvinick et al., 2001), this theory would appear to be able to accommodate those results in a relatively straightforward (though somewhat ad hoc) manner if, for example, it is assumed that conflict in a delayed-response incongruent trial does not arise at all or is simple to resolve by the time the response is requested. Thus, the fact that this conflict is only nominally present in a delayed-response trial would have no consequences for subsequent performance (i.e., it would produce no congruency sequence effect on the following trial).

Another relevant consideration in thinking about how control-based accounts could explain our results is that, in line with our previous discussion about potential strategies, a delayed-response procedure may inevitably change the normal processing involved in a task, possibly recruiting processes that would not be used otherwise. For example, it is reasonable to assume that Schmidt's delayed-response procedure requires some form of response inhibition, and that this process may, in some way, carry over to the next trial even if that trial requires an immediate response. For example, in the Go/NoGo paradigm, responding to a Go trial (i.e., a trial that requires a response) is overall slower if that trial is preceded by a NoGo trial (i.e., a trial for which no response should be made) than if it is preceded by another Go trial (Smith et al., 2010). A possible reason for this slow-down following a NoGo trial is that participants form expectancies for the subsequent trial based on the nature of the previous trial. Specifically, following a NoGo trial, participants may be biased toward anticipating having to inhibit their response on the following trial. This type of anticipation would lead to a slower response if the trial following the NoGo trial is a Go trial, i.e., a trial in which a response *should* be made (see also Niewenhuis et al., 2003). Although a delayed-response trial of the sort used by Schmidt (2017) (i.e., a trial in which a response needs to be withheld only for a short period of time) is not equivalent to a NoGo trial (a trial in which no response at all should be made), in both cases a response needs to be inhibited, at least initially.

Indeed, from a control-based perspective, it is conceivable that inhibitory processes of this sort, rather than temporal-learning processes, led to the patterns of results reported by Schmidt (2017) and in the present experiments. For example, it is possible that in a delayed-response paradigm, because a response can be determined faster for congruent trials than for incongruent trials, the time during which the response is inhibited will typically be longer in the former situation than in the latter. That is, when a delayed response is required, response inhibition should last longer for congruent trials because this process typically begins earlier (as the response is determined earlier for those trials). This prolonged inhibition, similar to what is found in the Go/NoGo task (Smith et al., 2010), may induce a bias for response inhibition that carries over to the next trial even if that trial requires an immediate response, a bias that would slow down the emission of responses that are normally fast, such as responses to congruent items. It is also reasonable to assume that this bias fades over time, that is, that following a delayed-response trial, the bias for response inhibition would likely be stronger in the early moments after the delayed-response trial than later. If so, this bias would have a smaller impact on responses that are normally slower, such as responses to incongruent items, because participants would have partially overcome that bias by the time conflict on that trial has been resolved and a response has been determined. In sum, an immediate-response congruent item following a delayed-response trial, particularly a congruent one, may be especially prone to slow down. Such an effect would cause the congruency sequence effect to be reduced, if not eliminated (which is, essentially, the pattern we observed).

The point of these considerations is that a delayed-response trial may do more than simply modifying temporal expectancies for future performance: It could prevent conflict from arising, it could allow participants to resolve that conflict completely, and/or it could promote use of additional processes (e.g., response inhibition) that would not be used otherwise. Thus, although, paralleling Schmidt (2017), our delayed-response procedure did produce the pattern of results predicted by the temporal-learning account, these considerations, in our view, mitigate the challenge that those results would seem to pose to control-based accounts. Indeed, it seems unlikely that the results obtained using this procedure can ever provide especially strong, let alone “the strongest” (Schmidt, 2017, p. 58), evidence in favor of the temporal-learning account of effects traditionally attributed to conflict-induced control. On the other hand, it must be acknowledged that the delayed-response procedure created by Schmidt, when applied to the prime-probe task, produces a pattern of results that the conflict-monitoring account does not readily explain without some additions to the model.

Challenges for the Temporal-Learning Account

As noted, while the results from the delayed-response procedure may well be consistent with the temporal-learning account, they are not necessarily inconsistent with control-based accounts. In order for the temporal-learning account to present a serious challenge to control-based accounts, it would seem appropriate that other paradigms, in particular, paradigms examining more typical situations, provide evidence in favor of the temporal-learning account. As noted, the present GLMM re-analyses not only failed to do so but, in most situations, also presented a pattern that does not appear easily reconcilable with the temporal-learning account: Why would congruency effects tend, if anything, to increase, rather than decrease, with higher previous RTs if the temporal-learning account were correct?

In fact, such a pattern is quite consistent with that typically found in the literature (Cohen-Shikora et al., 2019; Spinelli et al., 2019; see also De Jong et al., 1999) (i.e., with higher previous RT, latencies on the subsequent trial also increase, and as is found in many situations, effects are often larger when latencies are longer). One possibility is that this increase merely reflects mean scaling, i.e., the fact, noted by Schmidt (in press), that effects tend to “scale up” at longer latencies. On the other hand, the fact that this pattern was only found in the present experiments when the previous trial was incongruent (except when a very strict rejection criterion was used; see also footnote 13) may suggest a more complicated story.

For example, it is possible that strategic processes play a role in this pattern. Specifically, although responding to an incongruent trial is often faster following another incongruent trial, participants may feel the need to be especially careful (i.e., increase response caution) on an incongruent trial whenever responding to the previous incongruent trial was quite difficult. In that situation, participants would have just had a hard (but successful) experience with conflict and may even have come close to responding to the distractor before they caught themselves. Although an error was not committed, similar to what happens in post-error slowing (Rabbitt & Rodgers, 1977), participants in that situation may feel the need to slow down if the subsequent trial is also incongruent because they just learned that considerable effort may be required to make sure that they respond correctly to that type of trial. Such may not be the case if they have dealt with an incongruent item reasonably well on the previous trial. These ideas may explain why those incongruent trials after difficult incongruent trials generate such long latencies (i.e., why the line for incongruent trials goes up faster than that for congruent trials).

In any case, a potential role of the congruency of the previous trial would once again suggest that, unlike what the temporal-learning account appears to propose (see footnote 7), the latency of the previous trial, per se, is not the only determinant of performance on the subsequent trial. The fact that conflict did or did not arise on the previous trial may also be important in determining the relation between the latency on that trial and performance on the subsequent trial.

It is also worth noting that the present GLMM analyses (along with similar analyses reported by Cohen-Shikora et al., 2019, and Spinelli et al., 2019) are not the only pieces of evidence that appear to challenge the temporal-learning account. The temporal-learning account assumes that what is crucial to producing modulations of congruency effects in interference paradigms (e.g., the PC effect and the congruency sequence effect) is the difficulty associated with the items (i.e., congruent (easy) vs. incongruent (hard)), not the level of conflict associated with them (low for congruent items vs. high for incongruent items). Thus, if congruent and incongruent items in interference paradigms are replaced with easy-to-process and hard-to-process items which involve little or no conflict in conflict-free paradigms, one should observe modulations of difficulty effects that would parallel those observed in interference paradigms (e.g., the PC or congruency sequence effect). As noted, this parallel does arise in the proportion-easy paradigm, at least in some situations (Schmidt, 2013b, 2014, 2016; but see Spinelli et al., 2019): Similar to the typical PC pattern in PC paradigms, difficulty effects are larger in mostly-easy than mostly-hard lists in the proportion-easy paradigm even though there is, presumably, no conflict involved in either list. However, to our knowledge, there is currently no evidence that a similar pattern emerges when considering basic sequential analyses in other tasks, that is, that difficulty effects are larger following an easy trial than following a hard trial in conflict-free situations.

Further, there is considerable evidence to the contrary. For example, in a lexical decision task, Lima and Huntsman (1997) obtained the typical result of longer latencies for nonwords than for words (a lexicality effect). In one of their experiments, they also found that for both types of stimuli, latencies were faster when the stimulus was preceded by a word than when it was preceded by a nonword. More importantly, the difficulty effect (in this case, the lexicality effect) remained the same size following easy stimuli (words) vs. hard stimuli (nonwords), thus failing to produce a pattern similar to the congruency sequence effect (in their other experiment, hard stimuli were affected by the nature of the previous stimulus but easy stimuli were not, producing the opposite pattern one would expect from a temporal-learning account – larger, rather than smaller, difficulty effects following hard stimuli than following easy stimuli). Taylor and Lupker (2001) reported similar results in the context of a naming task: Both stimuli easier to name and stimuli harder to name were named faster when preceded by an easier than a harder stimulus, a pattern that they observed not only when the easy vs. hard contrast involved words vs. nonwords but also when it involved high-frequency vs. low-frequency words and easy nonwords vs. hard nonwords.

What these results suggest is that the difficulty of the previous trial has an overall impact on performance on the subsequent trial. This finding, along with results from the literature on blocking effects (Chateau & Lupker, 2003; Lupker et al., 1997; Lupker et al., 2003; Kinoshita & Mozer, 2006; Rastle et al., 2003), is certainly suggestive of a role for temporal processes in speeded performance. For example, Taylor and Lupker (2001) interpreted the sequential effects they found in terms of a time-criterion account (Lupker et al., 1997). According to this account, participants in speeded tasks establish a time criterion representing the time at which they expect, and will attempt, to initiate a response. This time criterion, similar to Schmidt's (2013b) notion of temporal expectancies, would be updated on a trial-by-trial basis based on the difficulty of the stimuli encountered in the task, with easy stimuli causing the criterion to be set at an early position (thus leading to a shorter latency on the next trial) and harder stimuli causing it to be set at a later position (thus leading to a longer latency on the next trial).

What the results from data examining sequential effects do not provide support for, however, is the idea that the difficulty of the previous trial would have a differential impact on easy vs. hard items on the subsequent trial, the pattern of results the temporal-learning account predicts. That is, in general, the difficulty effect on the subsequent trial typically remained the same size regardless of the nature of the previous stimulus, meaning that there appears to be no parallel to the congruency sequence effect in conflict-free tasks. In fact, the finding that, in this type of task, latencies for hard stimuli are typically *slower* when preceded by another hard stimulus is not easily reconcilable with the fact that in interference tasks, latencies for incongruent items are typically *faster* when preceded by another incongruent item. Similarly, in Schmidt's (2013b) original proportion-easy experiment, the hard items were slower in the mostly-hard list than in the mostly-easy list, rather than faster as Schmidt's account predicts. It is unclear how a temporal-learning process of any sort could reconcile these patterns (although, to be fair, similar inconsistencies that have not been fully explained have also emerged in interference tasks, e.g., why proportion-congruent manipulations would sometimes mainly impact congruent items, sometimes incongruent items, and sometimes both item types to a similar degree).

In the context of proportion manipulations, Schmidt (in press) recently argued that part of the reason for the inconsistencies in non-conflict paradigms may depend on the size of the stimulus set used, with inconsistent evidence for the temporal-learning account typically emerging in paradigms with large stimulus sets, such as those used in the literature on blocking effects (for a complementary argument, see also Schmidt, Cheesman, & Besner, 2013, for evidence that in conflict paradigms with large stimulus sets, interference effects do not seem to work in the same way that they do in more traditional conflict paradigms with small stimulus sets). Indeed, there is some suggestion in the literature that timing processes may be adjusted to the nature of the task and of the stimuli being used (e.g., Kinoshita & Mozer, 2006; Lupker et al., 2003). However, even assuming that this reasoning could be extended to the case of the congruency sequence effect, there is no mechanism in the temporal-learning account that would explain why timing processes should function differently with small vs. large stimulus sets. Further, it is not the case that paradigms with small stimulus sets inevitably produce evidence consistent with the temporal-learning account. For example, in a few cases (Blais & Bunge, 2010; Bugg, 2014; Bugg et al., 2008), PC effects were not observed for transfer items in Stroop paradigms even though, because only four to eight colors (and their corresponding color names) were used, the temporal-learning account (as well as the conflict-monitoring account) would have predicted a PC effect. Admittedly, supporters of neither non-conflict accounts nor control-based accounts seem to have been able to provide a full explanation for the inconsistencies emerging in conflict and non-conflict paradigms so far (Spinelli & Lupker, in review).

Conclusions

In sum, while the present research would appear to offer some evidence in support for a role of temporal learning in the congruency sequence effect, a careful consideration of the situation in which this evidence was obtained, of the performance literature in general, and of the lack of converging evidence from another approach we used, suggests considerable caution. A modified version of Botvinick et al.'s (2001) conflict-monitoring theory would still be able to account for the present results. Indeed, given the challenges to the temporal-learning account that we discussed, we do not think that it is likely that a purely time-based account could succeed at explaining effects such as the congruency sequence and PC effects unless that account is also modified. An idea that, in our view, is more promising is that temporal-learning processes work along with conflict-adaptation processes in shaping performance in congruency sequence paradigms, with the impact of the former process being more apparent for congruent items (explaining why those items tend to be slower following an incongruent item, i.e., following a slow temporal expectancy) and the impact of the latter process being more apparent for incongruent items (explaining why those items tend to be slower following a congruent item, i.e., an item that signals no conflict to the monitoring system; for a similar idea in the context of the PC effect, see Spinelli & Lupker, 2020). Future research should allow for additional, more stringent examinations of temporal-learning processes, conflict-adaptation processes, and their potential interplay.

Footnotes

1. Although in this manuscript we mainly refer to the original conflict-monitoring theory (Botvinick et al., 2001) when discussing control-based accounts of the congruency sequence effect and the PC effect, our doing so is not intended to suggest that that particular control-based account should necessarily be preferred over other control-based accounts that have been proposed in the literature – e.g., the perceptual-expectation hypothesis, the temporal-attention hypothesis, the negative-affect hypothesis, or the response-modulation account (see, e.g., Egner, 2014; Weissman et al., 2015). While the distinctions among these accounts are important, they are not particularly relevant in the present context where the main distinction we focused on was that between processes based on conflict and processes, particularly temporal learning, that are not based on conflict.
2. An interesting idea that is currently being explored is that the binding process and the contingency learning process may be one and the same (Schmidt et al., 2020). According to this idea, the reason high-contingency items elicit faster responses than low-contingency items would be that, for the former items, complete repetitions and complete changes represent the most frequent type of sequence, whereas for the latter items partial changes represent the most frequent type of sequence. Contingency learning would, therefore, be an artefact of binding. That is, the binding confound that may produce the congruency sequence effect and the contingency-learning confound that may produce PC effects may actually index the same non-conflict process, a binding process.

3. Initially, Schmidt (2013b) assumed that temporal expectancies in a PC manipulation would have to be based on the response latencies of multiple previous trials (i.e., 40). More recently, however, Schmidt (Schmidt et al., 2016; Schmidt & Weissman, 2016) proposed that in both congruency sequence and PC manipulations, temporal expectancies may be based on only a few of the previous trials (i.e., 5), with the most recent trial having a disproportionate influence. Temporal expectancies based on only the most recent trial would, therefore, be presumed to be the main factor in producing both the congruency sequence effect and the PC effect.
4. It is important to note, though, that when slow temporal expectancies do have an influence on hard-to-process items, this influence is usually to produce slower latencies, rather than faster latencies (e.g., Lupker et al., 1997; Taylor & Lupker, 2001), a pattern opposite to that predicted by the temporal-learning account. This pattern of results is more easily reconcilable with the time-criterion account (Lupker et al., 1997), according to which response emission for both easy- and hard-to-process items will be adjusted toward the point in time at which a response is expected to be emitted (i.e., earlier with a fast temporal expectancy vs. later with a slow temporal expectancy).
5. Although the temporal-learning account has typically been used to explain the emergence of PC and congruency sequence effects in the latencies, this account can explain emergence of such effects in error rates as well, when they do occur (Schmidt, 2013b). Specifically, the relatively large congruency effect that is sometimes observed in situations in which there is a fast temporal expectancy (i.e., an early drop in the response threshold, such as in an MC list or following a congruent trial), would result from the cases in which evidence for incongruent trials happens to accumulate at a fast enough rate to meet that fast expectancy (a situation that, as depicted in Figure 1A, ought to be uncommon as incongruent trials would typically miss fast temporal expectancies, but could happen for a portion of the incongruent trials). In those cases,

a response would be produced to those incongruent trials even though sufficient evidence may not have been accumulated yet to produce the correct response, resulting in an error. Such would not be the case for an incongruent trial in a situation in which there is a slow temporal expectancy (i.e., a late drop in the response threshold, such as in an MI list or following an incongruent trial). The reason is that, even when evidence accumulates at a fast enough rate to meet (or beat) that slow temporal expectancy, sufficient evidence would typically have been accumulated at that point in time to produce the correct response, resulting in fewer errors. Considering that errors for congruent trials are often at floor, the implication would be that a larger congruency effect would emerge in the error rates in situations leading to fast (vs. slow) temporal expectancies. In support of these ideas, Schmidt and Weissman (2016) reported two simulations in which a congruency sequence effect was produced in the error rates via a temporal-learning mechanism. On the other hand, to our knowledge, there have been no cases in which error rates from human data have been used to support the temporal-learning account of the congruency sequence effect, possibly because, in human data unlike in Schmidt and Weissman's simulations, this effect does not often show up in the errors.

6. On immediate-response trials, the long-wait group also showed a general slow-down compared to the short-wait group regardless of whether the previous trial required an immediate or delayed response, suggesting that imposing a long delay on some of the trials in the experiment may have induced caution for participants in the long-wait group, increasing their latencies.
7. Although Schmidt and Weissman (2016) suggested that this three-way interaction is one that would be expected based on the temporal-learning account, it does not seem to follow that such would be the case. The point of Schmidt and Weissman's (2016; Schmidt, 2013b) model appears to be that it is the latency of the previous trial, not its congruent vs. incongruent nature, that determines the temporal expectancy for the subsequent trial. Therefore, it is not clear why

the fact that the previous trial is incongruent would prevent the latency on that trial (a latency that, although typically slow, likely varies and may be equal or even faster than that for congruent trials in some occasions) from creating a temporal expectancy capable of affecting performance on the subsequent trial (see also Schmidt, in press).

8. This point is also made clear by how temporal learning is implemented in Schmidt's (2013b) computational model (the Parallel Episodic Processing model). In that model, the response threshold is lowered as a function of the proximity between the current *cycle* in the model and the *rt* stored for each previous episode, *i*, being considered (see Schmidt's (2013b) supplementary materials and Schmidt and Weissman (2016)):

$$proximity_i = 1 - \left(\frac{(cycle - rt_i)^2}{10000}\right)$$

What is crucial to note here is that *rt* stands for raw RT, i.e., the actual model cycle at which, for the previous episode, the response was emitted – not the response rate or any metric reflecting how quickly evidence accumulated during that episode.

9. This point can be demonstrated using a similar simulation as the simulation that Schmidt (in press) used to demonstrate that linear correlations created in one scale become weaker when the data are transformed to another scale. In this simulation, reported at https://osf.io/ynatu/, a curvilinear relationship is created between two Gamma distributions such that their values are positively associated, but more strongly so at the lower than the upper end. As a result, in this original (i.e., raw) scale, although a line would explain part of that relationship, it would not be the most appropriate function. In the inverse scale, however, a line would explain the relationship between the two distributions much better, as reflected by the fact that a larger Pearson's *r* is typically obtained for the correlation between the two distributions in the inverse scale compared to the raw scale. Note, however, that concluding that the true relationship "exists" in the inverse scale (i.e., that that relationship is the result of some form of processing

based on inverted latencies) because the inverse scale is the scale in which the test for a linear trend produced the best-fitting model, would be incorrect in this situation. Although the test for a linear trend in the raw scale produced the worse-fitting model, the fact remains that the true relationship exists in that scale in this situation (i.e., that that relationship is the result of processing based on raw latencies), even though the relationship is not a linear one. Therefore, in general, it would seem to be inappropriate to use model-fitting performance to adjudicate the scale in which the true relationship between two variables (e.g., previous RT and current RT) exists.

10. We also conducted an analysis including both context and transfer items. The pattern of results was the same.
11. For the present Experiments 1 and 2, we also conducted parallel analyses using a similar data treatment as Schmidt and Weissman's (2016), i.e., discarding trials for which either the current or previous RT was faster than 300 ms. The pattern of results remained the same. Note, further, that for both Experiments 1 and 2 and Schmidt and Weissman's Experiments 1 and 2, examination of Q-Q plots assuming a Gamma distribution as the theoretical distribution (the same distribution used for the GLMMs, see below) did not reveal a problem with fast RTs. Instead, the sample distributions deviated from the theoretical distributions more for slower RTs. However, this problem, present in the full datasets, was substantially reduced in the datasets with a stricter rejection criterion.
12. Before proceeding with the GLMM re-analysis, we replicated Schmidt and Weissman's (2016) original LMM analysis with inverse-transformed RTs on the full dataset. The pattern of results was the same as they reported although there were minor differences in the values, likely due to the analysis program we used (Schmidt and Weissman used SPSS rather than R) and to our choice to include the random effect of items (a factor not included in Schmidt and Weissman's

analysis) but not a random effect of Experiment (a factor included in Schmidt and Weissman's analysis).

13. For simplicity, Previous Response Type was not included as a fixed effect for the re-analysis of the present Experiments 1 and 2, however, including this fixed effect in the analysis did not change the pattern of results. Also for simplicity, we did not include Experiment as a fixed effect in either the re-analysis of Schmidt and Weissman's (2016) experiments or the re-analysis of our experiments. With one exception, including this fixed factor in the analysis for the full datasets did not change the general pattern of results. The exception is that, whereas in our Experiment 2 congruency effects increased with a higher RT on the previous trial only when that previous trial was incongruent, in our Experiment 1 congruency effects increased with a higher RT on the previous trial regardless of the congruency of that previous trial.
14. A minor result that also emerged in the analyses of both the present experiments and Schmidt and Weissman's (2016) is that the inclusion of Previous RT in the model seemed to artificially reverse the main effect of Previous Congruency. That is, the model predicted faster latencies following incongruent than congruent trials when, actually, latencies were slightly faster following *congruent* than incongruent trials. The most likely explanation for this result is that, although Previous RT and Previous Congruency are highly correlated, Previous RT "won" the competition for explaining the variance shared among them, leaving essentially nothing for the main effect of Previous Congruency to explain. In that situation, the estimates for the main effect of Previous Congruency would not be particularly reliable.

References

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*, 509-517. http://dx.doi.org/10.1037/xhp0000324

Algom, D., & Chajut, E. (2019). Reclaiming the Stroop effect back from control to input-driven attention and perception. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01683

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Balota, D., Aschenbrenner, A., & Yap, M. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*, 1563-1571. https://doi.org/10.1037/a0032186

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48. https://doi.org/10.18637/jss.v067.i01

Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *Journal of Cognitive Neuroscience*, *22*, 2758-2767. https://doi.org/10.1162/jocn.2009.21365

Bolker, B. (2020). *GLMM FAQ*. Retrieved from https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652. https://doi.org/10.1037/0033-295X.108.3.624

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J. C., Weissman, D. H., Notebaert, W., Egner, T. (in press). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences, 23*, 769-783.

Bugg, J. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*, 567-587. https://doi.org/10.1037/a0035032

Bugg, J. M., & Chanani, S. (2011). List-wide control is not entirely elusive: Evidence from picture–word Stroop. *Psychonomic Bulletin & Review, 18*, 930-936. https://doi.org/10.3758/s13423-011-0112-y

Bugg, J. M., & Crump, M. J. (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. https://doi.org/10.3389/fpsyg.2012.00367

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*, 1484-1494. https://doi.org/10.3758/MC.36.8.1484

Chateau, D., & Lupker, S. J. (2003). Strategic effects in word naming: Examining the route-emphasis versus time-criterion accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 139-151. https://doi.org/10.1037/0096-1523.29.1.139

Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 1703–1723. https://doi.org/10.1037/xlm0000670

Colombo, L., Spinelli, G., & Lupker, S. J. (2020). The impact of consonant–vowel transpositions on masked priming effects in Italian and English. *Quarterly Journal of Experimental Psychology*, *73*, 183-198. https://doi.org/10.1177/1747021819867638

Colvett, J. S., Nobles, L. M., & Bugg, J. M. (2020). The unique effects of relatively recent conflict on cognitive control. *Journal of Experimental Psychology: Human Perception and Performance*, *46*, 1344-1367. https://doi.org/10.1037/xhp0000860

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica*, *101*, 379-394. https://doi.org/10.1016/S0001-6918(99)00012-8

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The heterogeneous world of congruency sequence effects: An update. *Frontiers in Psychology*, *5*, 1001. https://doi.org/10.3389/fpsyg.2014.01001

Egner, T. (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, *5*, 1247. https://doi.org/10.3389/fpsyg.2014.01247

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*, 778-788. https://doi.org/10.3758/s13421-016-0591-1

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*, 480-506. https://doi.org/10.1037/0096-3445.121.4.480

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011a). Conflict, error likelihood, and RT: Response to Brown & Yeung et al. *NeuroImage*, *57*, 320-322. https://doi.org/10.1016/j.neuroimage.2011.04.027

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011b). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage*, *57*, 303-311. https://doi.org/10.1016/j.neuroimage.2010.12.027

Hazeltine, E., & Mordkoff, J. T. (2014). Resolved but not forgotten: Stroop conflict dredges up the past. *Frontiers in Psychology*, *5*, 1327. https://doi.org/10.3389/fpsyg.2014.01327

Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*, 1-17. https://doi.org/10.1007/s00426-003-0132-y

Howell, D. C. (1998). *Statistical methods in human sciences*. New York: Wadsworth.

Hutchison, K. A. (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*, 851-860. https://doi.org/10.1037/a0023437

JASP Team (2020). *JASP* (Version 0.14.1) [Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

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*, 271-284. http://dx.doi.org/10.1037/a0027734

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

Kim, S., & Cho, Y. S. (2014). Congruency sequence effect without feature integration and contingency learning. *Acta Psychologica*, *149*, 60-68. https://doi.org/10.1016/j.actpsy.2014.03.004

Kinoshita, S., & Mozer, M. C. (2006). How lexical decision is affected by recent experience: Symmetric versus asymmetric frequency-blocking effects. *Memory & Cognition*, *34*, 726-742. https://doi.org/10.3758/BF03193591

Kinoshita, S., Mozer, M. C., & Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology: General*, *140*, 622-636. https://doi.org/10.1037/a0024230

Kunde, W., & Wühr, P. (2006). Sequential modulations of correspondence effects across spatial dimensions and tasks. *Memory & Cognition*, *34*, 356-367. https://doi.org/10.3758/BF03193413

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Levy, R. (2014). Using R formulae to test for main effects in the presence of higher order interactions. *arXiv:1405.2094*.

Lima, S. D., & Huntsman, L. A. (1997). Sequential dependencies in the lexical decision task. *Psychological Research*, *60*, 264-269. https://doi.org/10.1007/BF00419412

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*, 166-174. https://doi.org/10.3758/BF03197535

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 570-590. https://doi.org/10.1037/0278-7393.23.3.570

Lupker, S. J., Kinoshita, S., Coltheart, M., & Taylor, T. E. (2003). Mixing costs and mixing benefits in naming words, pictures, and sums. *Journal of Memory and Language*, *49*, 556-575. https://doi.org/10.1016/S0749-596X(03)00094-9

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

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*, 450-452. https://doi.org/10.1038/nn1051

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 17-26. https://doi.org/10.3758/CABN.3.1.17

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology*, *29*, 727–743. https://doi.org/10.1080/14640747708400645

Rastle, K., Kinoshita, S., Lupker, S. J., & Coltheart, M. (2003). Cross-task strategic effects. *Memory & Cognition*, *31*, 867-876. https://doi.org/10.3758/BF03196441

Ridderinkhof, K.R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research*, *66*, 312-323. http://dx.doi.org/10.1007/s00426-002-0104-7.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*, 103-113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Schmidt, J. R. (2013a). Questioning conflict adaptation: proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*, *20*, 615-630. https://doi.org/10.3758/s13423-012-0373-0

Schmidt, J. R. (2013b). Temporal learning and list-level proportion congruency: conflict adaptation or learning when to respond?. *PLoS One*, *8*, e82320. https://doi.org/10.1371/journal.pone.0082320

Schmidt, J. R. (2014). List-level transfer effects in temporal learning: Further complications for the list-level proportion congruent effect. *Journal of Cognitive Psychology*, *26*, 373-385. https://doi.org/10.1080/20445911.2014.896367

Schmidt, J. R. (2016). Temporal learning and rhythmic responding: no reduction in the proportion easy effect with variable response-stimulus intervals. *Frontiers in Psychology, 7*, 634. https://doi.org/10.3389/fpsyg.2016.00634

Schmidt, J. R. (2017). Time-out for conflict monitoring theory: Preventing rhythmic biases eliminates the list-level proportion congruent effect. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*, 52-62. https://doi.org/10.1037/cep0000106

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review, 26,* 753-771. https://doi.org/10.3758/s13423-018-1520-z

Schmidt, J. R. (in press). When data transformations are appropriate or even necessary: A response to Cohen-Shikora, Suh, and Bugg (2019). *Timing & Time Perception*.

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 514-523. https://doi.org/10.1037/0278-7393.34.3.514

Schmidt, J. R., Cheesman, J., & Besner, D. (2013). You can’t Stroop a lexical decision: Is semantic processing fundamentally facilitative? *Canadian Journal of Experimental Psychology*, *67*, 130-139. https://doi.org/10.1037/a0030355

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

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

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

Schmidt, J. R., Lemercier, C., & De Houwer, J. (2014). Context-specific temporal learning with non-conflict stimuli: Proof-of-principle for a learning account of context-specific proportion congruent effects. *Frontiers in Psychology*, *5*, 1241. https://doi.org/10.3389/fpsyg.2014.01241

Schmidt, J. R., Notebaert, W., & van den Bussche, E. (2015). Is conflict adaptation an illusion?. *Frontiers in Psychology*, *6*, 172. https://doi.org/10.3389/fpsyg.2015.00172

Schmidt, J. R., & Weissman, D. H. (2014). Congruency sequence effects without feature integration or contingency learning confounds. *PLoS One*, *9*, e102337. https://doi.org/10.1371/journal.pone.0102337

Schmidt, J. R., & Weissman, D. H. (2016). Congruency sequence effects and previous response times: conflict adaptation or temporal learning?. *Psychological Research*, *80*, 590-607. https://doi.org/10.1007/s00426-015-0681-x

Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In D. Spieler, & E. Schumacher (Eds.), *New Methods in Cognitive Psychology.* New York: Routledge.

Smith, J. L., Smith, E. A., Provost, A. L., & Heathcote, A. (2010). Sequence effects support the conflict theory of N2 and P3 in the Go/NoGo task. *International Journal of Psychophysiology*, *75*, 217-226. https://doi.org/10.1016/j.ijpsycho.2009.11.002

Spinelli, G., & Lupker, S. J. (2020). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and color-word correlation confounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0000820

Spinelli, G., & Lupker, S. J. (in review). Robust evidence for proactive conflict adaptation in the proportion-congruent paradigm.

Spinelli, G., Perry, J. R., Lupker, S. J., (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*, 995-1014. http://dx.doi.org/10.1037/xhp0000656

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

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 117-138. https://doi.org/10.1037/0278-7393.27.1.117

Weissman, D. H., Egner, T., Hawks, Z., & Link, J. (2015). The congruency sequence effect emerges when the distracter precedes the target. *Acta Psychologica*, 156, 8-21. https://doi.org/10.1016/j.actpsy.2015.01.003

Weissman, D. H., Jiang, J., & Egner, T. (2014). Determinants of congruency sequence effects without learning and memory confounds. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 2022-2037. https://doi.org/10.1037/a0037454

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931- 959. https://doi.org/10.1037/0033-295X.111.4.931

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2011). Errors of interpretation and modeling: A reply to Grinband et al. *Neuroimage*, *57*, 316-319. https://doi.org/10.1016/j.neuroimage.2011.04.029

Author Note

This research was partially supported by Natural Sciences and Engineering Research Council of Canada Grant A6333 to Stephen J. Lupker and the Ontario Trillium Scholarship to Giacomo Spinelli. We wish to thank James Schmidt for sharing his E-prime files and his data with us. We also wish to thank Julie Bugg, Emily Cohen-Shikora, Corentin Gonthier, and Keith Hutchison, for sharing their data with us. The raw data and the scripts used for the analyses are publicly available at https://osf.io/ynatu/.

Correspondence concerning this article may be addressed to: Giacomo Spinelli, Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, Milan (MI), 20126, Italy (e-mail: giacomospinelli@hotmail.it) or Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario, N6A 5C2, Canada (e-mail: lupker@uwo.ca).