BRILL

Timing & Time Perception 9 (2021) 161–197

Timing & Time Perception
brill.com/time

When Data Transformations are Appropriate or Even Necessary: A Response to Cohen-Shikora, Suh and Bugg (2019)

James R. Schmidt*

Université Bourgogne Franche-Comté, LEAD-CNRS UMR 5022, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon, France

Received 3 January 2020; accepted 18 August 2020

Abstract

In this paper, I argue that common data transformations used for statistical modelling are not inherently problematic. Depending on the research question, transformation can be appropriate or even necessary. The paper also discusses the often-overlooked impact of decision-related processes (e.g., rhythmic timing) on behaviour and how such biases can often unintentionally confound research designs. More narrowly, the current paper considers a recent debate about the list-level proportion congruent (LLPC) effect, which is the finding that congruency effects (e.g., in the Stroop task) are reduced when most trials are incongruent relative to when most trials are congruent. The LLPC effect is typically interpreted as evidence for conflict-driven attentional control (conflict monitoring). However, another view proposes that a rhythmic responding bias (temporal-learning) explains the effect. In a recent article, Cohen-Shikora, Suh, and Bugg (2019) challenged some of the evidence for the latter account. One key question they raise is whether it is appropriate to inverse transform (essentially: de-skew) response times when using linear mixed effect modelling. The authors argued that this transform is problematic and presented a series of analyses that they argued demonstrate both (a) that there are minimal concerns about temporal-learning confounds, and (b) that conflict monitoring clearly contributes to the LLPC effect. The present article presents new analyses and demonstrates that neither of these two key conclusions of Cohen-Shikora and colleagues are justified. More global implications for linear mixed effect modelling are discussed, including an analysis of when data transformations should or should not be used.

Keywords

Temporal learning, data transformations, conflict monitoring, cognitive control, attention, proportion congruent effect, mixed models

*To whom correspondence should be addressed. E-mail: james.schmidt@ubfc.fr

DOI: 10.1163/22134468-bja10019

1. Introduction

In experimental psychology, the influence of decision-based factors on performance is often overlooked (see Grosjean et al., 2001). Frequently, for instance, a researcher might be interested in how quickly participants can respond to different types of stimuli, where the influence of the *content* of the items on processing speed, attention, etc. is of interest. However, many popular manipulation types can also influence decision-based processes, such as the evidence accumulation criterion that participants set for themselves before accepting a potential response alternative and executing it. As will be discussed in the present report, one example of this is studies in which the proportion of different filler item types is varied and researchers assess how performance on some target items is influenced by the type of filler. The typical experimental logic in this type of study is that performance on target trials might be influenced by the content of fillers (e.g., congruent vs. incongruent, easy vs. hard, positive vs. negative, etc.). However, as I will argue in this manuscript, this type of design can also produce rhythmic timing biases. In particular, a faster pace in a condition with faster-to-identify fillers relative to a condition with slower-to-identify fillers can influence performance on target trials simply by virtue of the task rhythm. The current report will focus on one very specific example of this from the attentional control domain, but the same concerns equally apply to any other domain making use of similar manipulations.

The current report will also consider the appropriateness of data transformations when analysing data. In particular, this paper will consider a series of papers from Balota et al. (2013), Lo and Andrews (2015), and, more centrally, Cohen-Shikora, Suh, and Bugg (2019) that have asked whether or not the typical process of applying inverse transformations (i.e., 1/observation) to heavily skewed data typical of response times is appropriate when making use of linear mixed effect (LME) models. While the abovementioned reports have been rather critical of this standard analysis approach, the present paper will present a defence of this analysis approach. More precisely, it will be argued that the preference for analyses on raw vs. transformed data often depends on the research question: sometimes transformation is not only acceptable, but also appropriate. Again, the present report will focus primarily on one specific research question from the attentional control domain, but the present manuscript will highlight how transformations of data are either merely acceptable or absolutely necessary for a wide range of problems. I will return to each of the two abovementioned broader issues shortly, but I will first outline the more specific question of interest in the following section.

2. Conflict Monitoring

One popular cognitive control theory of attention is the *conflict-monitoring* (or *conflict adaptation*) account (Botvinick et al., 2001). According to this account, each experience of conflict between competing response tendencies leads to an upregulation of control, with a downregulation in the absence of conflict. For instance, in a Stroop task (Stroop, 1935) participants respond to the print colour of colour words, which produces conflict on *incongruent trials* — where the word and colour mismatch (e.g., 'red' in green) — but not on *congruent trials* — where the word and colour match (e.g., 'red' in red) (see Note 1). According to the conflict-monitoring account, then, control is increased after incongruent trials and decreased after congruent trials. One particular strain of evidence for conflict monitoring is the *proportion congruent* (*PC*) *effect* (Logan & Zbrodoff, 1979; Logan et al., 1984), which is the finding that congruency effects are reduced when trials are mostly incongruent (e.g., 80% incongruent) relative to mostly congruent (e.g., 80% congruent), as illustrated in Fig. 1. Although initially interpreted differently, the conflict-monitoring account proposes that this effect occurs because control of attention away from the distracting word (and/or toward the colour) is increased when conflict is more frequent, thereby reducing the congruency effect (Cohen et al., 1990; Lowe & Mitterer, 1982).

Much debate has centred around whether PC effects like this are due to conflict monitoring or to other biases (for reviews, see Abrahamse et al., 2016; Bugg & Crump, 2012; Schmidt, 2013a, 2019). For instance, simple stimulus–response contingency learning biases (Hazeltine & Mordkoff, 2014; Schmidt, 2013b; Schmidt & Besner, 2008) and binding biases (Risko et al., 2008) confound the PC effect. Relevant to the present article, however, one issue in the literature is whether there is a PC effect independent of any item-specific biases (Cheesman & Merikle, 1986; Glaser & Glaser, 1982; Kane & Engle, 2003; Lindsay & Jacoby, 1994; Shor, 1975; West & Baylis, 1998). This I will refer to as the *list-level proportion congruent* (LLPC) *effect*. Typically, LLPC is assessed by manipulating the PC of the list (e.g., averaged across all items for one group of participants or block of trials) with some *biased* (or

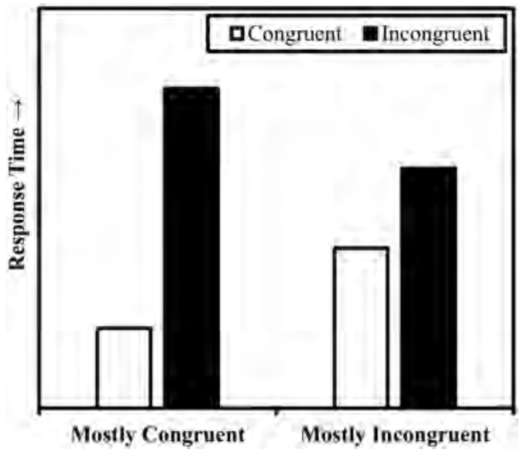

Figure 1. Example proportion congruent effect. The congruency effect is smaller when most trials are incongruent relative to when most trials are congruent.

inducer) *items*. For instance, 'blue' and 'red' might be mostly congruent (e.g., 'blue' most often in blue) in one condition and mostly incongruent (e.g., 'blue' most often in red) in another condition. Intermixed with these biased items are some other *transfer* (or *diagnostic*) *items* that are not directly manipulated. For instance, 'green' and 'brown' might be presented equally often in green and brown for all participants (i.e., the same congruent:incongruent ratio in both PC conditions). It is the PC effect for these transfer items that we term the LLPC effect.

Notably, a LLPC effect cannot be explained by contingency learning or binding, but could, in principle, be explained by transfer of control from the manipulated items to the transfer items. Some of the first, most straightforward manipulations of LLPC produced no effect (Blais & Bunge, 2010; Bugg et al., 2008). However, later reports have observed effects in a variety of tasks (e.g., Stroop, Simon, picture–word, prime–probe; Bugg, 2014; Bugg & Chanani, 2011; Bugg et al., 2011; Gonthier et al., 2016; Hutchison, 2011; Schmidt, 2017; Spinelli & Lupker, in press), including across tasks (Funes et al., 2010; Torres-Quesada et al., 2013; Wühr et al., 2015). There still remain alternative interpretations of these LLPC effects, however. For the present report, I will focus on one alternative mechanistic account of the LLPC effect: temporal learning.

3. Temporal Learning

Schmidt (2013c) first presented the notion that the LLPC effect might be due, wholly or in part, to temporal-learning biases (for a related idea in masked priming, see Kinoshita et al., 2011). The idea is not necessarily easy to grasp if one is used to thinking about the content of the items we manipulate (e.g., congruent vs. incongruent, high vs. low frequency, etc.). However, many times more systematic variance in response times is explained by how we time our responses than by the factors themselves (see Grosjean et al., 2001). For instance, we are highly biased to time our responses in a rhythmic way: my response time (RT) on the current trial will likely be similar to my RTs on very recent trials, and increasingly less similar to a given prior RT the further back in time it occurred. This systematic variability in response times is called *pink noise, 1/f noise,* or *flicker noise*. These autocorrelations in RTs are omnipresent in a broad range of cognitive paradigms, including mental rotation, lexical decision, visual search, and speeded classification (Gilden, 1997, 2001; Gilden et al., 1995). In Fig. 2a, the data of one randomly selected participant from Bugg (2014) are presented (Participant 312) (Note 2) which demonstrates the typical pink noise pattern. To better visualize the pink noise, Fig. 2b presents simulated data showing how current RT correlates with RTs of previous lags in autocorrelated data. In particular, current RT becomes less and less correlated with prior RTs the larger the lag between the two RTs (Note 3).

At first glance, this sort of rhythmic timing bias may seem to be orthogonal to the experimental manipulation of content (e.g., the proportion of congruent

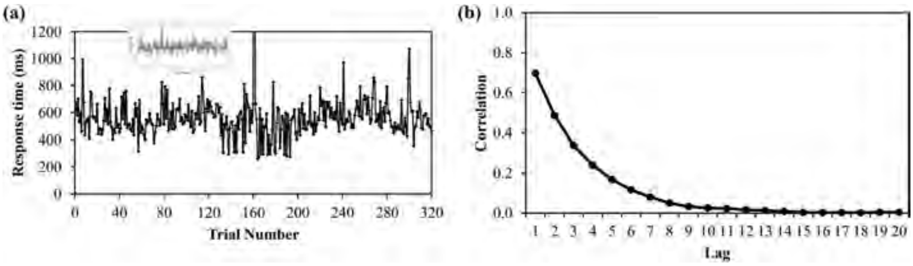

Figure 2. (a) Classic 1/*f* (or pink/flicker) noise in a participant from Bugg (2014). Note the local correlations [similar adjacent response times RT)] with both short and long-term fluctuations in response times. The insert displays the same data randomized to produce zero-correlation white noise. (b) Simulated data showing an autocorrelation between current RT and prior RTs of varying lags.

trials), but it is not. It has been repeatedly observed in a number of domains that timing biases produce interactive effects between study factors that have relatively little to do with the factor manipulations themselves (Kinoshita & Lupker, 2003; Kinoshita et al., 2008, 2011; Lupker et al., 1997; Mozer et al., 2004; Schmidt, 2014, 2016a). Indeed, Kiger and Glass (1981; see also, Kinoshita et al., 2011) stress that such decision-related (rather than content-related) effects 'will continue to be rediscovered in many circumstances … and will be mistakenly attributed to a multiplicity of causes' (p. 697).

Rhythmic timing biases can produce a LLPC effect because such biases can affect congruent and incongruent trial types differentially in conditions with a faster vs. slower task pace. Naturally, the task pace in a mostly congruent list will be much faster than in a mostly incongruent list (i.e., more fast congruent trials in the former). This is illustrated in the top panels of Fig. 3 with imagined data: because there are so many congruent trials in the mostly congruent list, incongruent trial RTs largely fall in the right tail of the overall distribution as outliers, whereas the reverse is the case in the mostly incongruent list. Schmidt (2013c) argued that timing biases will benefit response speed selectively for trials in which participants have sufficient evidence to select a response at the expected time. A simplified illustration is presented in the bottom panels of Fig. 3. In particular, the threshold for selecting a response is decreased (i.e., the trigger to respond is loosened) at the expected time, allowing for faster responses if the task pace can be maintained (i.e., if there is sufficient evidence to cross the temporarily-decreased threshold). When the task pace is fast (e.g., mostly congruent), congruent trials will tend to benefit from temporal expectancies. That is, participants will have enough evidence to select a response at the expected time and maintain their task pace. For the occasional incongruent trial, however, there will typically not be enough evidence for a response at the expected time (e.g., due to ongoing resolution of conflict), and responding will therefore be

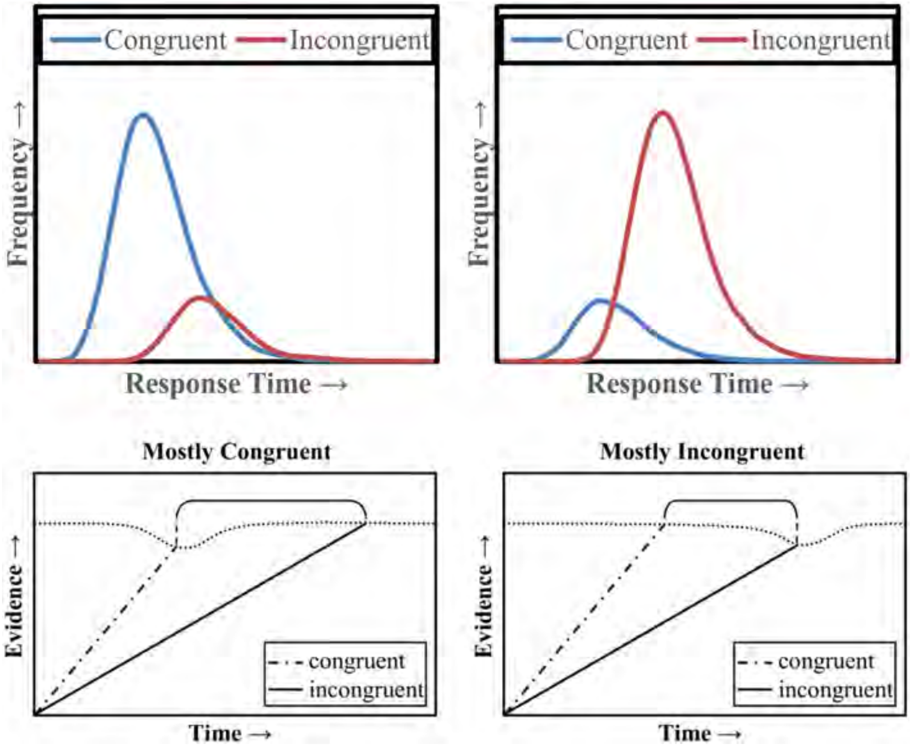

Figure 3. Top panels illustrate the different frequencies of fast to slow responses in the mostly congruent (left) and mostly incongruent (right) lists. Bottom panels present a simplified illustration of how temporal expectancies can produce a list-level proportion congruent (LLPC) effect. Dotted line = response threshold; slopes indicate rate of evidence accumulation for the correct response.

delayed. The net effect is an inflated congruency effect. In the mostly incongruent condition, the situation is largely reversed. The task pace is slower and an early response is therefore not expected. Expectancy for a later response might therefore benefit incongruent trials. The occasional congruent trials, however, do not benefit in the same way as in the mostly congruent condition. The net effect is a smaller congruency effect.

At a very rough level, the notion is that the faster the 'pace' of responding, the more likely it is that a given congruent trial will benefit from temporal expectancies (i.e., the temporarily reduced response threshold) and the less likely that a given incongruent trial will benefit. We could therefore consider previous trial RT as a rough proxy for pace, with the prediction that the congruency effect should be overall larger the faster the previous RT. As one of several lines of evidence for temporal-learning biases in the LLPC effect, Schmidt (2013c) tested this notion. In particular, congruency, PC, previous trial RT, the interaction between congruency and PC, and the interaction between congruency and previous RT were used as predictors of current-trial RT (along with subject and item random effects) in a LME regression on LLPC data from Hutchison (2011). The predictions of the temporal-learning account were met. First, the standard autocorrelations in response times were observed (i.e., a sizeable interaction between previous and current RT). Second and more importantly, previous RT and congruency interacted. That is, the faster the previous trial RT, the larger the congruency effect on the current trial. Third, this previous RT by congruency interaction explained variance in the LLPC effect, with the latter effect diminishing after accounting for previous RT biases. The LLPC effect (i.e., interaction between PC and congruency) was still significant, but (as explained later) this was expected. As will be expanded on in further detail in sections to follow, these analyses were a far from perfect test of the temporal-learning account, but did provide positive evidence of temporal learning (e.g., the conflict adaptation account should not have predicted the observed results).

4. Challenge to the Temporal-Learning Account

In a recent report, Cohen-Shikora and colleagues (2019) presented a strong challenge to the temporal-learning account of the LLPC effect. As a one key point of their critique, they question a particular detail of the LME analyses in Schmidt (2013c): previous- and current-trial RTs were inverse-transformed (−1000/RT) (Note 4). This transform was used for three reasons. First, this transform (along with all other data treatments) was based directly on past reports (esp., Kinoshita et al., 2011). Second, an inverse transform normalizes the response time distribution (Gamma and log transforms are also relatively effective and similar, but inverse is typically the most optimal). The typical response time distribution is not normal, but rather ex-Gaussian in shape, with a heavy positive skew. This violates the distributional assumptions of LME, so it would be inappropriate to interpret the LME results without a correction. As illustrated in Fig. 4, an inverse transform normalizes the distribution by reducing the right tail and increasing the left tail, thereby resolving the problem. The third reason for using an inverse transform relates to the reason why raw RT is ex-Gaussian-distributed in the first place and will be returned to later.

Cohen-Shikora and colleagues (2019) reproduced the LME analyses of Schmidt (2013c) and further replicated the analyses on two more datasets from Bugg (2014) and Gonthier and colleagues (2016). The Hutchison (2011) dataset was a colour–word Stroop task with 226 participants (Note 5) each of which performed 180 trials in either a mostly congruent or mostly incongruent condition. The Bugg dataset was also a colour–word Stroop task with 72 participants from their Experiments 1a and 2a, and each participant completed 320 trials in either the mostly congruent or mostly incongruent condition. The Gonthier dataset was a picture–word Stroop task with 93 (Note 6) participants from their Experiments 1a and

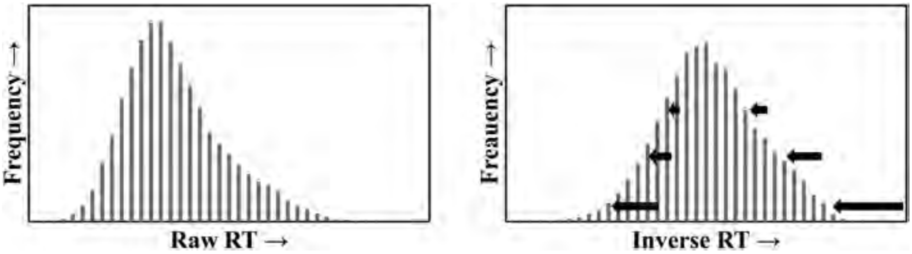

Figure 4. Ex-Gaussian simulated data (left) and the same data after an inverse transform (right). Arrows illustrate how the transform moves the tails of the distribution.

1b, and each participant performed 384 trials in both the mostly congruent and mostly incongruent conditions. In all three datasets, all of the original predictions of the temporal-learning account were met in the LME as originally performed by Schmidt (2013c). Specifically, in all three datasets (a) previous trial RT was significantly correlated with current-trial RT, (b) previous trial RT was significantly negatively correlated with current-trial congruency, and (c) the LLPC effect beta was reduced after accounting for (a) and (b).

However, the authors argued that the use of an inverse transform on current RT (and, of less importance, previous RT) was problematic. Although inverse (or similar) transforms are increasingly being adopted in LME analyses as standard practice (e.g., Andrews & Lo, 2012; Kinoshita et al., 2011; Kliegl et al., 2010; Masson & Kliegl, 2013), there are scenarios in which this might be undesirable (Balota et al., 2013; Lo & Andrews, 2015), as I will expand on in a section to follow. Cohen-Shikora and colleagues therefore re-conducted the analyses on raw RTs (for both previous and current RT) with a *generalized linear mixed model* (GLMM), which can correct for the skewed distribution by modelling the skew (subsequently also applied by Spinelli et al., 2019, which will be discussed later in the manuscript). This was done with a Gamma distribution (similar to an inverse) and an identity link function (which specifies that factors should have a linear relationship with the dependent variable, as in an ANOVA or simple linear regression). Unlike the LME results with the inverse transform, the results with GLMM on raw RTs can best be described as inconsistent, with no clear evidence for a temporal-learning bias in the LLPC effect across datasets (e.g., the beta for the LLPC effect actually increased in two of the three datasets).

The authors also fairly pointed out that the influence of timing biases on the LLPC interaction was merely eyeballed in Schmidt (2013c). That is, the beta for the interaction was reduced with temporal-learning controls (as predicted), but this decrease was not tested statistically. Because of this, Cohen-Shikora and colleagues (2019) performed a series of additional analyses, which (for the most part) provided better support for the temporal-learning account in the LME than in the GLMM. For instance, the Akaike information criterion (AIC) and the

Bayesian information criterion (BIC) improved with previous trial RT in the LME for all three datasets but reversed in two of the three datasets in the GLMM.

Although both approaches have limitations, they also observed in the LME that variance explained by the LLPC interaction increased in one dataset (Hutchison, 2011) and the change in R^2 by adding the LLPC interaction was not reduced with temporal-learning metrics in this same dataset (for the other two datasets, these metrics were consistent with the temporal-learning account). The authors themselves pointed out a limit with the former approach (see their Footnote 12): variance explained for the LLPC interaction increased in this one dataset when including timing controls, even though the beta got smaller. This may seem contradictory, but only if the analysis of Cohen-Shikora and colleagues is misunderstood as a test of change in the size of the LLPC with the introduction of a temporal-learning control. Instead, their contrast tests whether there is a change in variance explained, which is influenced not only be the size of the effect, but also its precision. And, indeed, the LLPC estimates did become more precise (i.e., reduced standard errors) when modelling away the substantial noise variance introduced by pink noise. A similar concern also applies to the R^2*-change* analyses, where the authors tested whether the increase in variance explained by introducing the LLPC interaction was smaller in a model with previous RT controls than in a model without previous RT controls: again, this models variance explained, rather than the effect magnitude.

Indeed, none of the supplementary tests that these authors provided directly tested the significance of the change in beta weights for the LLPC interaction by introducing temporal-learning metrics, which is the actual question of interest. For this, we can use a test for beta changes with nested data (Clogg et al., 1995), which is designed to directly measure the significance of a change in betas with the introduction of one or more additional control factors to the regression. Consistent with the temporal-learning account, the beta did significantly reduce in all three LME datasets when previous RT and the interaction with congruency were added to the regression: Hutchison (2011), $t(223) = 4.116$, SE = 0.002, $p < 0.001$; Bugg (2014), $t(69) = 5.057$, SE = 0.003, $p < 0.001$; and Gonthier and colleagues (2016), $t(86) = 3.729$, SE = 0.002, $p < 0.001$.

The authors also performed analyses on the three-way interaction between previous RT, congruency, and PC, with the notion that the temporal-learning account should predict such an interaction (not observed in two of the three datasets). However, they appropriately acknowledge that no claims have been made about the presence or absence of this interaction in past argumentation for the temporal-learning account. Indeed, the present author is unsure why the temporal-learning account should make any strong predictions about this three-way interaction and said interaction does not speak directly to the LLPC effect, anyway.

Also interesting, the authors aimed to improve the proxy for pace by averaging three previous response times (first in an unweighted average and then in an exponentially-weighted average). Although the authors suggested that improving the temporal-learning proxy in this way should have eliminated the LLPC effect if the temporal-learning account is correct, this is not justified. The influence of timing on the LLPC effect would have to be modelled perfectly for this to be true, but, as will be explained in section 6 *Previous RT and Pace*, this is much more difficult to achieve than the simple addition of a few more prior RTs. What certainly could be predicted from the temporal-learning account, however, is that adding extra prior RTs to the regression should explain a bit more variance. Consistent with this, there was a further decrease in the beta for the LLPC effect in the LME in all three datasets. Again, results with GLMM were less favourable to the temporal-learning account, with an increase in the LLPC beta in two of the three datasets. Globally, then, the temporal-learning account fared much worse in the GLMM analyses than in the LME. Indeed, with LME, the temporal-learning predictions were met in all three datasets with the original analyses in addition to the newly-introduced AIC/BIC measures and the improved timing measures introduced by Cohen-Shikora and colleagues (2019), but this was definitely not the case with the GLMM analyses. As I also demonstrated above, the betas for the LLPC effect also significantly decrease with the introduction of a temporal-learning control, which is the most direct test of the temporal-learning account. The present report will focus primarily on the differences between LME and GLMM in the simple analyses as performed initially by Schmidt (2013c), but will return to some of these additional analyses later. I will also introduce some new ways of statistically assessing the impact of temporal-learning biases on the LLPC effect, including on raw response times.

Based on the above analyses, the conclusion of Cohen-Shikora and colleagues (2019) was that there is no clearly established evidence for temporal-learning confounds in LLPC effects and that such confounds can be safely ignored. This seems surprising to the present author, as the LME data clearly seem to provide consistent support for the temporal-learning account. The strong claims of Cohen-Shikora and colleagues, therefore, seem to be based on a favouring of the GLMM data (which were much less favourable for the temporal-learning account) and a dismissal of the LME data (but with no explanation for the consistent patterns across datasets). The authors further pointed out that the LLPC effect (i.e., congruency by PC interaction) remains robust regardless of how the data are analysed, with the implication that this is inconsistent with a pure temporal-learning account of the LLPC effect and that a conflict-monitoring contribution is difficult to contest.

5. Response to the Challenge

Cohen-Shikora and colleagues (2019) thus present a challenge to the temporal-learning account and in the rest of this paper I will address this challenge. In particular, I will attempt to convince the reader of five things. First, the fact that a LLPC effect remains after modelling pace with previous RT (or even a weighting of several past RTs) does not argue against a pure temporal-learning account. This is because the temporal-learning account predicts a priori that previous trial RT is only a weak proxy of 'pace' and will therefore only capture a small part of the true rhythmic timing bias. In fact, if previous RT completely explained away the LLPC effect, this would actually be inconsistent with the temporal-learning account explained above. Second, inverse transforms are not inherently problematic and may even be regarded as a more sensible metric than raw RT analyses, particularly for the present case. Third, the reason that GLMM with raw RTs produced notably 'worse' results than LME with transformed RTs was due to distortion of the autocorrelation in response times in a raw RT scale relative to the inverse scale. The difference is not, in contrast, due to a distortion of a true effect of LLPC. Fourth, even in raw RTs one can still observe clear evidence for temporal-learning confounds in the LLPC effect if previous RT is allowed to predict variance in the correct way. Fifth, there are compelling lines of empirical evidence that provide convergent support for temporal learning in the LLPC effect. Each of these points will be addressed in a separate section below. Most importantly, this report will provide an explanation for why one approach (LME) provides relatively consistent support for temporal learning across multiple datasets, whereas another approach (GLMM) finds only noise. That is, if the GLMM results — which provide no clear evidence of temporal learning — are to be trusted as the true story (i.e., that there is no temporal-learning bias), then there should be some account of why the LME results provide clearer support for temporal learning. That is, how are the inverse-scaled response times repeatedly providing evidence for temporal learning if there is no temporal-learning bias to start with?

6. Previous RT and Pace

Let me first start by agreeing with one aspect of the conclusions of Cohen-Shikora and colleagues (2019): attempting to 'model away' temporal-learning biases with an LME (or GLMM) using previous RT as a proxy for temporal learning is unlikely to work very well. In fact, this was my conclusion from the outset (Schmidt, 2013c). Indeed, it was predicted on an a priori basis that including previous RT in the LME would reduce but not eliminate the LLPC effect. The temporal-learning account predicts this because previous RT is only a weak proxy for pace, meaning that (a) much of the temporal-learning bias will not be captured by previous RT, and (b) the LLPC interaction will continue to 'steal' this unmodelled timing bias. These two conclusions follow from theory, and were also demonstrated with a computationally modelled implementation of the theory. In particular, a large simulated dataset was created with the Parallel Episodic Processing (PEP) model. The PEP model implements the temporal-learning mechanism discussed above and produces the LLPC effect as a direct result of this temporal-learning mechanism exclusively (e.g., with appropriate lesion studies to localize the effect to this specific mechanism). When these simulated LLPC data were analysed in the same way as the participant data, the LME revealed the same previous RT effects as in the participant sample, and a large remaining LLPC effect.

This finding in the simulated data itself was expected: the temporal-learning account simply does not predict that inclusion of previous RT in the LME should eliminate the LLPC effect. This is because the LME effectively tests the hypothesis that the magnitude of the congruency effect linearly increases with a linear decrease in previous RT (or inverse RT, as the case may be). For instance, following an unusually fast RT, the congruency effect should be unrealistically gigantic (i.e., as the congruency effect should continue to grow the faster and faster the prior RT), whereas following an unusually slow RT the congruency effect should be trivially small or even negative (i.e., the congruency effect should continue to shrink the slower the prior RT, eventually crossing zero and reversing sign). This is simply not what the temporal-learning account predicts (especially a negative congruency effect). In fact, the specific notion implemented in the PEP model (and represented visually in Fig. 3) predicts that most of the 'movement' should be around the peak pace of the RT distribution (which can be observed in changes in skewness and kurtosis; see Schmidt, 2016a, for detailed analyses), with much less movement in the tails. Thus, a priori, there is not a one-to-one relationship between previous RT and current-trial congruency effects. The test for a linear slope therefore only partially captures the temporal-learning bias and the LLPC interaction term should continue to 'steal' some of this missed variance. In fact, if the LLPC effect were eliminated by controlling for previous RT then it would indicate that the temporal expectancy account explained above is wrong (e.g., being inconsistent with the PEP model data and the logical implications of the verbal model, as explained above). As such, the persistence of a LLPC effect in such analyses should not be taken as strong evidence against a pure temporal-learning view nor as strong evidence for a conflict-monitoring contribution to the effect. Instead, it is ambiguous, favouring neither the temporal-learning nor the conflict-monitoring account. This ambiguity, of course, is problematic, but the results from the statistical modelling analyses do provide some evidence in favour of a temporal-learning contribution to the LLPC effect (see also section 10 *Other Lines of Evidence for Temporal Learning*).

7. The Scale of Time

The previous section explored why previous RT only roughly measures what it serves as a proxy for. In the section to follow, I will demonstrate that analyses on raw (rather than inverse) RT distort this proxy even further. First, the present section will explore what an inverse transform actually does and when such transforms are and are not problematic. Cohen-Shikora and colleagues (2019) rightly point out that inverse transforming RTs changes the nature of the question being assessed:

> "Furthermore, these transformations change the nature of the variable being explored; what was an analysis of raw RT (response time, what researchers are typically formulating predictions about, as in the case of Schmidt's (2013c) predictions) becomes an analysis of response rate (a different DV) once transformed to inverse RT. As Lo and Andrews (2015) and Robidoux (2017) pointed out, researchers should take this into account when justifying a transformation that is appropriate for their predictions."

Indeed, there are clear cases in which transforming data makes the test of a specific hypothesis inappropriate. That is, the transform can distort the original scale of a variable in a potentially undesirable way (Stevens, 1946). In these cases, we should indeed prefer analyses on non-transformed data. In other cases, the reverse is true: raw RTs do not test the theory appropriately and transformed data do.

First, let us consider an instance of the former case: tests for additive main effects. Lo and Andrews (2015; see also, Balota et al., 2013) were specifically concerned with word-reading models, some of which explicitly propose additive relations between factors. In other words, said models adhere to additive factors logic (Sternberg, 1969), proposing, for instance, that word frequency and stimulus quality should each have an independent influence on naming times, with no interactions between the two factors (e.g., the effect of stimulus quality should be equivalent for high- and low-frequency words). The idea is that the processes that produce one effect (e.g., stimulus quality) are separate from the processes that produce another effect (e.g., word frequency), so the two should not interact. Prima facie, it may seem like a null interaction is necessary with independent mechanisms, though this is not necessarily the case (e.g., in a cascading system; see Ridderinkhof et al., 1995; Smid et al., 1991). Some word-reading models do assume strong additivity whereas others (e.g., Borowsky & Besner, 2006) do not. When testing these sorts of additive-factor models, we do want to ensure that response times are not distorted by a transform. For example, with an inverse transform, longer response times from the extended right tail of an ex-Gaussian distribution are 'squished' toward the centre of the distribution and faster response times from the left tail are 'stretched out' (see Fig. 4). When you have two factors with a large main effect, the result of such a transform will be a relative decrease in the differences between the slower cells of the design relative to the faster cells. This can change the effective form of the interaction, as illustrated in a simplified example with one observation per cell in Fig. 5 (see also Balota et al., 2013).

First, it is noteworthy that the above-discussed issue with transformed data is applicable to assessing noncrossover interactions in which there is a main effect for both factors. The same concern is not applicable to crossover-type interactions (Loftus, 1978), which the LLPC effect is (Note 7). That is, a test for the additivity of two main effects can change direction (positive, negative, or additive) with a transform. However, this is not true of a crossover interaction, which will remain

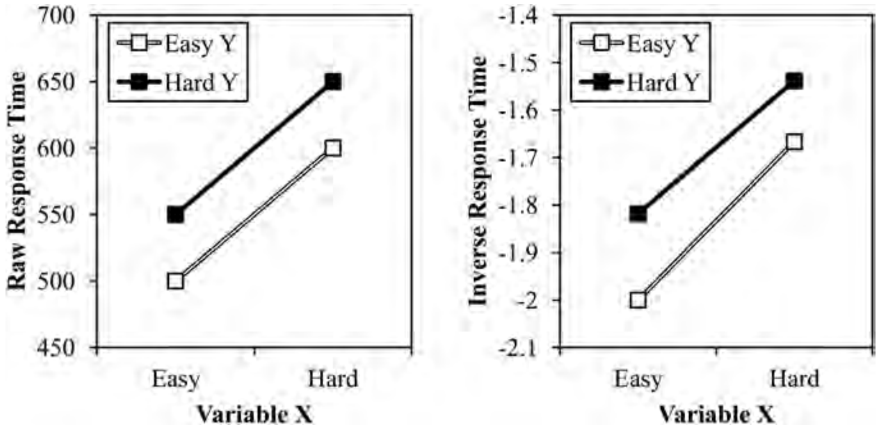

Figure 5. Simplified example of how an inverse transform can influence an interaction. Note that while the raw response times (left) are additive, an underadditive interaction is present in the transformed data (right).

a crossover interaction (in the same direction) after a transform. It is similarly not true of a simple main effect (Kliegl et al., 2010), which will also maintain its direction after a transform. In this sense, the issue with inverse transforms discussed in Lo and Andrews (2015) and Balota and colleagues (2013) is not applicable to the LLPC effect and does not argue for a switch from inverse to raw RT (though I will return to a different scenario in LLPC data later that might raise a related concern).

More generally, analyses on transformed data are not inherently problematic. In fact, there are scenarios in which raw RT is clearly the wrong scale. Lo and Andrews (2015) give a clear example of this. Certain theories of cognitive ageing propose a general slowing with age, whereby a given effect (e.g., a Stroop effect) will be larger in an ageing population simply because response time effects scale up with slower responses (e.g., Salthouse, 1985). In this case, we are interested in knowing whether the magnitude of an effect-proportional-to-mean RT is any different in young and elderly populations. A log or similar transform (e.g., effect-proportional-to-mean or z transformations; Note 8) is thus the only way to assess the viability of this general cognitive slowing theory (i.e., that the effect-proportional-to-mean RT is no larger or smaller in an elderly population), and analyses on raw RT are inappropriate.

Also outside of the speeded response time domain, there are a number of domains in which an inverse scale is theory-appropriate. For instance, Weber's law states that the just-noticeable difference between two things — in many sensory modalities, such as luminance, length, mass, or sound perception — is proportional to the reference level (though not always perfectly; see Holway & Pratt, 1936). For instance, if a temporal duration (e.g., of a tone) x is just noticeably different from a temporal duration $i \bullet x$, then durations y and $i \bullet y$ will be similarly distinguishable. That is, any multiplicative difference of i will be just noticeable. Similarly in memory for temporal order, a larger difference in time is required to distinguish two events at a comparable rate the farther ago the events occurred — termed temporal distinctiveness — ranging all the way from the scale of milliseconds for very recent events to the scale of years for very distant events (Brown et al., 2007). There are many other examples of inversed scaling, such as the scalar property of time estimation (e.g., French et al., 2011), where accuracy in time estimation scales with the reference duration.

By convention, most cognitive theories are developed and tested on raw RT. This does not imply, however, that raw RT is the most sensible metric for a given theory. Though views differ (cf. Balota et al., 2013; Lo & Andrews, 2015), there is even an argument as to why inverse RT is actually a better metric for most questions we might ask. This is related to the very reason why RT data are likely ex-Gaussian-distributed in the first place. Consider a simplified (Note 9) example of the drift-diffusion model (Ratcliff, 1978), presented in Fig. 6. Note how arithmetic increases in the slope (i.e., speed of processing information) do not translate into arithmetic increases in RT (or variance). Put a different way, a Gaussian distribution of slopes (or 'drift rates' in diffusion model terms) will produce an ex-Gaussian distribution of response times. This is why the diffusion model fits raw RT distributions (Wagenmakers & Brown, 2007). Though Balota and colleagues rightly point out that an inverse transform would be inappropriate for standard diffusion model analyses, an inverse transform is akin to transforming RTs back into their underlying parameter value (viz., slope).

More globally, response time effects and variances tend to 'scale up' the slower one responds (Schmidt, 2016b; Schmidt & De Houwer, 2016; Stevens et al., 2002; Urry et al., 2015), also linearly in relation to one another (Wagenmakers & Brown, 2007). Thus, if we want to know whether there are differences in the underlying learning or (as the case may be) attentional control in a given response time effect, then inverse-transformed RT (which corrects for the abovementioned mean and variance scaling) is probably a better measure of the underlying processes of interest than raw RT. Indeed, the conflict-monitoring account is exactly about processing rate: evidence accumulation for the target and distracting dimensions proceeds at different rates (slopes) depending on attentional control settings. The typical analyses may be on raw RT, but this does not necessarily mean that raw RT analyses best reflect the underlying theory (*argumentum ad antiquitatem*). Of course, this drift-diffusion example illustrates why, from many theoretical perspectives, transforming data may make more sense than one initially imagines, but this will not be the case in all instances. For instance, accounts that assume additive effects of two or more discrete processes (discussed earlier) are fundamentally incompatible with this sort of thinking in terms of a single drift process (except, perhaps, if there are two separated drifts).

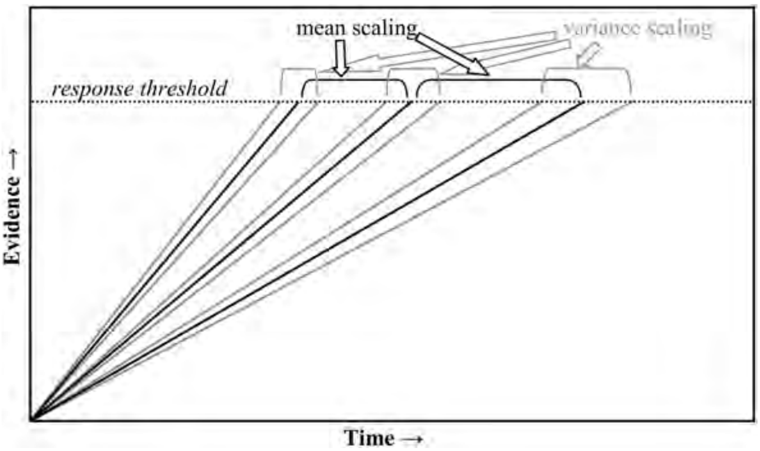

Figure 6. Simplified example of why diffusion models produce ex-Gaussian response times. Linear increases in drift produce greater-than-linear increases in response time (RT) and variance. Black lines indicate mean trajectory and the surrounding grey lines indicate variances.

How frequently we might prefer analyses on inverse or raw RT (or yet another scale) is open for debate. Indeed, there has been a long discussion of the up- and downsides of restricting analyses to simple means (Sternberg, 1969) vs. considering the distribution of effects (Heathcote et al., 1991) and/or performing transformations of non-normal data (Kliegl et al., 2010). At minimum, however, this section aimed to show that (inverse) transformed response times are not merely a distortion of what we should (always) be interested in. It is true that in some cases inverse transforms clearly are undesirable (e.g., when aiming to assess additivity of two main effects). However, in many cases transformation is appropriate or even necessary. Most critically, concerns about distortion of an interaction are not applicable to the crossover-type interaction observed in the LLPC. The reader may therefore wonder why LME on inverse RT and GLMM on raw RT produced seemingly contradictory results in Cohen-Shikora and colleagues (2019). The next section will explain this discrepancy and why analyses on inverse RTs are more appropriate for assessing the temporal-learning account.

8. Inverse RT Better than Raw RT for Assessing Autocorrelations

While transforming response times can influence the nature of an interaction, particularly between two factors with a large main effect for each, this is not the reason why LME and GLMM seemingly gave different answers to the same question in Cohen-Shikora and colleagues (2019). More precisely, the decision whether to transform does not meaningfully influence the PC by congruency interaction. Instead, analyses on raw (rather than inverse) current and previous RT distort the autocorrelation between previous and current RT. These two assertions will be clearly demonstrated in this and the following section.

The distortion of the response time autocorrelation can most easily be visualized with scatterplots of the relation between raw previous and current RT and between inverse previous and current RT, presented in Fig. 7 for each of the three datasets used by Cohen-Shikora and colleagues. Note that in the transformed data the scatterplots are relatively normal, with most observations in the centre oval with a positive slope indicating the standard autocorrelation. The exact same correlations in the raw RTs, however, are very atypical. Most of the observations are 'squished' into the bottom left corner of the scatter plots and the relatively slower (previous and current) response times are spread out distantly from this in a fan. This is a predictable consequence of correlating two ex-Gaussian-distributed variables with exactly the sort of autocorrelation predicted by the temporal-learning account. What this pattern means is that most of the trials simply anchor the regression line (bottom left) and the slope is almost exclusively determined by massively outlying response times in the right tails of the previous and current RT distributions. That is, by asking the regression to plot a straight line through this 'fan' pattern, very little weight is given to the bulk of the observations and a very large weight is given to severely-outlying slow RT observations. We are essentially asking the regression to fit the outliers and not the rest of the data. This is related to the familiar textbook example of the heavy oversensitivity of Pearson's r to outliers.

The raw RT scatter plots are not only atypical but are also diagnostic for why the LME results on the normalized RTs produce different results than the GLMM on raw RT. Put simply, the raw scale is not the right metric for the hypothesis. The temporal-learning account does not predict effects to be localized primarily in the extreme right tail of the distribution, but this is exactly what is tested with the identity link function on raw RT in the GLMM (Note 10). Indeed, as mentioned earlier, the temporal-learning account actually predicts most of the movement to be around the peak of the response time distribution, not in the right tail. This can also be observed in the amount of autocorrelation between previous and current RT. This was done by first removing subject and item noise using LME to get residual RT and previous RT estimates for each participant in both raw and inverse data. The resulting correlation between previous RT and current RT was significantly larger in the inverse scale by 27% in Hutchison (2011), $z = 5.978$, $p < 0.001$, 14% (Note 11) in Bugg (2014), $z = 2.262$, $p = 0.024$, and 19% in Gonthier and colleagues (2016), $z = 5.016$, $p < 0.001$ (the z modification from Silver et al., 2004, is reported, but all six tests from the cocor package converge on the same inference for all reported tests). There are two important things to note about these changes. First, the autocorrelation is significantly reduced, but certainly not eliminated in the raw scale. Second, note that the autocorrelation between previous and current RT, while a strong prediction of the

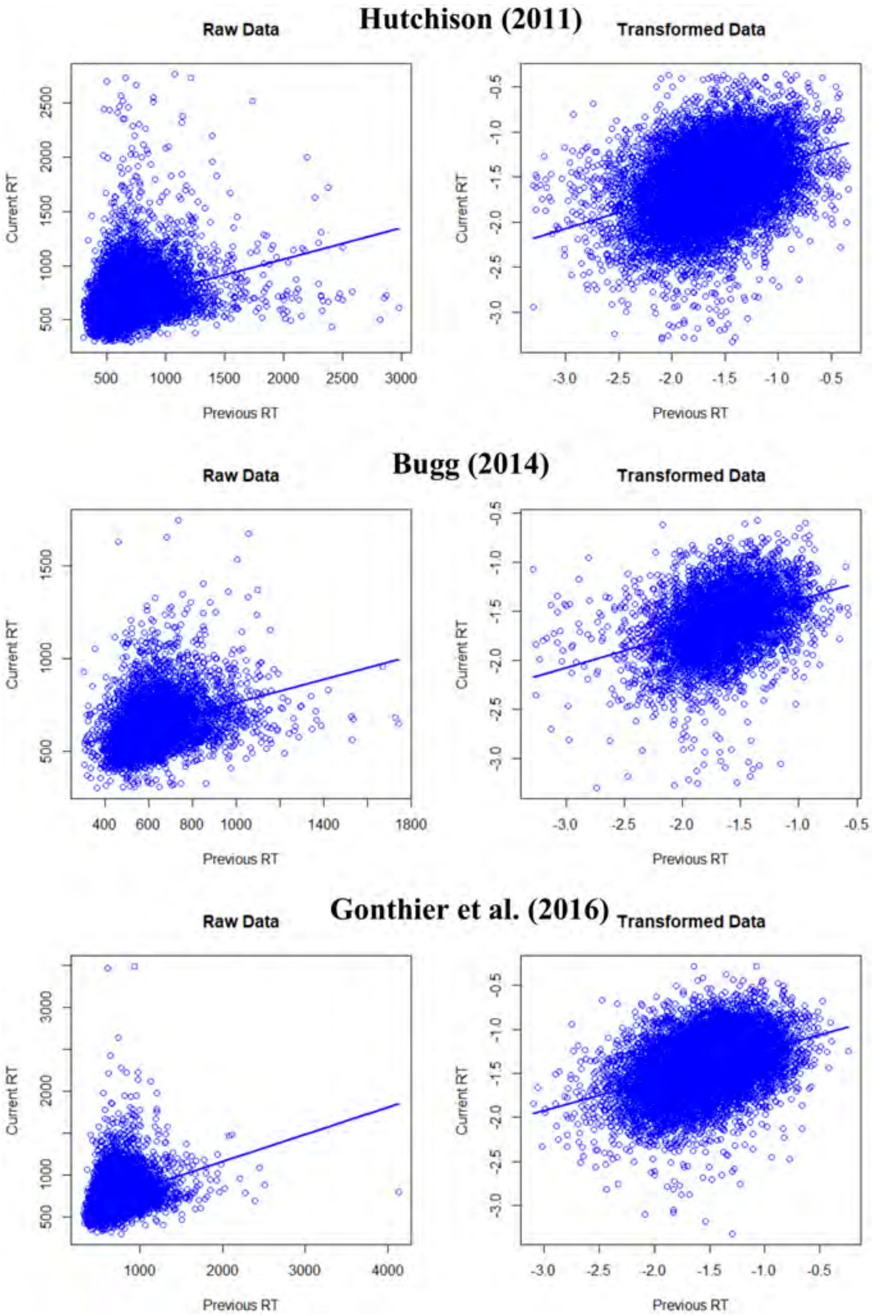

Figure 7. Scatterplots of the relation between current and previous response time (RT) on raw (left) and inverse-transformed (right) scales.

temporal-learning account, is not how the temporal-learning account explains the LLPC effect. Instead, the temporal-learning account explains the LLPC effect via the previous RT by congruency interaction. Importantly, a much larger increase in the correlation for the interaction between previous RT and congruency is observed in the inverse scale of 149% in Hutchison (2011), $z = 8.722$, $p < 0.001$, 109% in Bugg (2014), $z = 5.229$, $p < 0.001$, and 351% in Gonthier and colleagues (2016), $z = 7.038$, $p < 0.001$. That is, the correlations in the inverse RT scale (0.084, 0.079, and 0.050, respectively) are substantially larger than those in the raw RT scale (0.034, 0.038, and 0.011, respectively). Thus, while the simple autocorrelation is decreased moderately, what is lost in the autocorrelation is exactly the variance that the temporal-learning account predicts is important for explaining the LLPC effect.

Indeed, what changes between the LME and GLMM analyses is not the presence of a LLPC effect (this remains stable with or without a transform). The direction of this crossover interaction (along with the main effects of congruency and PC) is simply not impacted by an inverse transform. What changes, instead, is how well previous RT predicts current RT and, more critically, congruency (and indirectly: LLPC). That is, a continuous predictor (like previous RT) will be influenced strongly by the scaling of the variable. When response times are adjusted to their theory-anticipated inverse scale (see section 7 *The Scale of Time*), the continuous previous RT variable does a good job of explaining variance in both current RT and congruency. We should naturally expect that this effective predictive power should be undermined when distorting this continuous predictor to a heavily skewed scale (along with the continuous dependent variable).

It is further important to note that a correlation that does not exist will not emerge out of a transform. That is, if previous RTs are not actually related to the LLPC effect, then a transform will not make it appear as if they are. Any changes in slopes will simply be random (i.e., not systematic). In contrast, a correlation that does truly exists can be distorted, even eliminated entirely, by a transform. Analogically, this is similar to trying to fit a straight regression line to an inverted-U shaped curvilinear relationship, or vice versa. As a logical consequence, correlations will necessarily be weaker in a scale that more poorly reflects the true relationship between two variables (Note 12). As with standard model-fitting techniques, then, the results reported here demonstrate that the inverse scale better fits the true relationship of previous RT to current RT and congruency. Thus, the inverse scale should be preferred. Of course, while a better-fitting model provides evidence against a worse-fitting model, the better-fitting model can still be wrong. The next section will consider (but eventually reject) a potential caveat with the conclusions thus far.

9. Corrected Raw RT Analysis

It is never explicitly stated by Cohen-Shikora and colleagues (2019) why the GLMM results provided notably different results than LME. The authors do reference Lo and Andrews (2015) as a reason to be sceptical of analyses on inverse-transformed data. The concerns raised in that paper, however, are that the direction of an interaction might be reversed (or otherwise qualitatively changed) by 'stretching' of the response time distribution with a transform. Although this should not apply to the LLPC effect (as explained in section 7 *The Scale of Time*) one might nevertheless propose that the variance that inverse previous RT explains in the LLPC effect in inverse RTs does not 'stretch out' to the raw RT scale. For example, a reviewer (Giacomo Spinelli) notes that the interaction between previous RT and congruency (critical for the temporal-learning account in explaining the LLPC effect) is between two factors with a main effect. Hypothetically then, if we assume that an autocorrelation does exist in response times (as proposed by the temporal-learning account), but that this autocorrelation is equivalent for congruent and incongruent trials (unlike the temporal-learning account prediction), then an inverse transform could reduce the slope for the (typically slower) incongruent trials relative to the (typically faster) congruent trials. This would create an underadditive interaction between previous RT by congruency in the inverse scale that does not exist in the raw scale. That is, previous RT may be explaining an effect that only 'exists' in the inverse scale (Note 13). According to this view, the apparent temporal-learning effect in the inverse scale is an artefact that is not applicable to the LLPC effect in the raw scale. In the present section, I will perform an analysis that tests this notion directly. This section will additionally demonstrate that previous RT does explain the LLPC effect in the raw RT scale.

Admittedly, it is difficult to test for a potential impact of previous RT on the LLPC while both (a) allowing previous RT to predict variance on the inverse time scale and (b) measuring the LLPC on the raw RT scale. There is, however, a two-step analysis approach that can achieve these goals, which was applied to each of the three datasets separately. To avoid confusion, the procedure is illustrated in Fig. 8. The first step involves computing raw RT residuals from the temporal-learning model assessed on the inverse scale. To do this, previous and current RT were first inverse-transformed (Fig. 8a). As in Cohen-Shikora and colleagues (2019), current and previous RTs faster than 300 ms were trimmed (i.e., to normalize the *Q*-*Q* plots) and previous RT was centred on the mean. Next, an LME was performed as before, but without PC as a factor. That is, previous RT, congruency, and the previous RT by congruency interaction were entered as fixed factors, with subject, item, and (for the applicable datasets) experiment random intercepts. This initial LME is used for only one purpose: to compute individual-trial-predicted inverse RT (Fig. 8b). This predicted inverse RT can then be simply transformed back to the

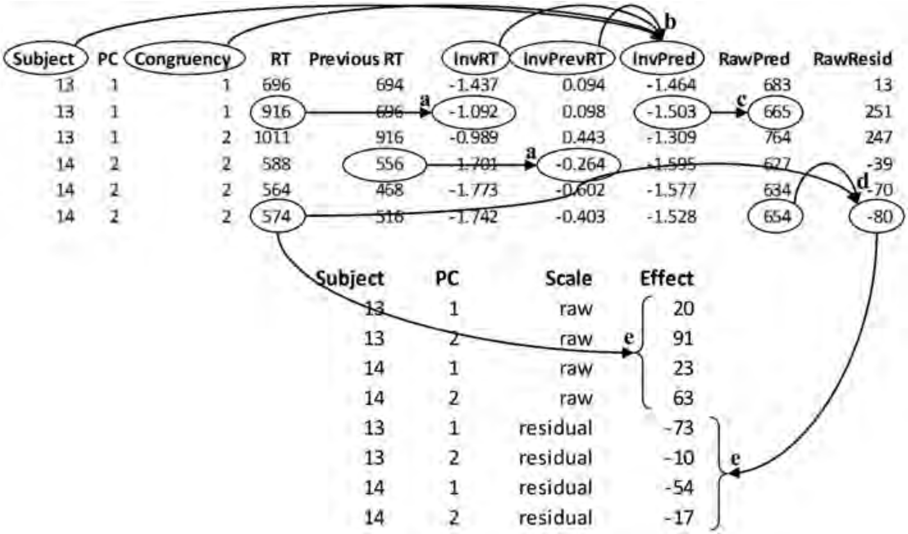

Figure 8. Illustration of the two-step modelling procedure: (a) current response time (RT) is inverse-transformed (InvRT) and Previous RT is inverse-transformed and centered (InvPrevRT); (b) Subject, Congruency, and InvPrevRT are entered into a linear mixed effect (LME) model to generate predicted InvRT values (InvPred); (c) InvPred is inverse-transformed back to the raw scale (RawPred) to have the temporal-learning prediction in the raw scale; (d) RT (raw) is subtracted from RawPred to get a raw residual (RawResid); (e) the congruency effect as a function of proportion congruent (PC) is computed separately in RawResid and simple (raw) RT, and a simple PC × Scale ANOVA is conducted to see whether the list-level proportion congruent (LLPC) effect is reduced in raw residuals.

raw RT scale (i.e., by applying the −1000/RT transform again) to get a temporal-learning-predicted RT on the raw scale (Fig. 8c). Next, this predicted RT can be subtracted by the observed raw RT for each observation to get a raw residual RT (Fig. 8d). The net result of this first step is simply to subtract out the temporal-learning prediction from the raw response times.

The second step of the analysis merely involves comparing the magnitude of the LLPC effect in the pure raw scale to that in the raw residuals. That is, the impact of previous RT on the LLPC effect is being assessed on the raw scale (as in the GLMM analyses), except that previous RT has been allowed to predict variance in current RT and congruency in an inverse scale. At this point, previous RT has already been allowed to explain what variance it can, so neither an LME nor a GLMM are needed. Instead, we can simply compute the congruency effect (i.e., incongruent − congruent) for each participant on raw RTs and the raw residual RTs and run an ANOVA on the PC (mostly congruent vs. mostly incongruent) by scale (raw vs. residual) congruency effects (Fig. 8e). This produces a significant decrease in the LLPC effect in the residual relative to raw scale of 30% (6.49 ms) in Hutchison (2011), $F(1,224) = 4.437$, MSE = 57627, $p = 0.036$, 20% (6.34 ms) in Bugg (2014), $F(1,70) = 4.761$, MSE $= 5324$, $p = 0.032$, and 11% (2.85 ms) in Gonthier and colleagues (2016), $F(1,88) = 20.051$, MSE $= 794$, $p < 0.001$.

What these analyses show is that the variance that (inverse) previous RT is explaining in the LLPC effect is not only within the inverse scale of response times. Instead, it also 'transforms out' to the raw RT scale. Stated more simply, controlling for temporal-learning biases *does* significantly decrease the LLPC effect in raw RTs as long as the temporal-learning assessment is fair. Thus, we can clearly see that the reason why a GLMM on raw RTs produced different results was simply because of the negative impact of the skew in the key predictor variable (previous RT) and dependent variable (current RT): the effect is present in raw RTs, but it simply is not captured well when previous RT is asked to predict variance in a purely linear way in abnormal scatterplots. Together, these results provide a clear explanation for the discrepancies between the LME and GLMM analyses in Cohen-Shikora and colleagues (2019) (Note 14).

10. Other Lines of Evidence for Temporal Learning

The previous section demonstrated that previous RT does explain variance in the LLPC effect in the raw response timescale if predictions are generated from the theory-anticipated and data-supported inverse form. Notably, however, the LLPC effect remains robust in all datasets with this approach. As already mentioned, this is anticipated a priori from the temporal-learning account, as previous RT is only a rough proxy of pace. The same applies to the other analyses performed by Cohen-Shikora and colleagues (2019): attempts to eliminate the LLPC effect with previous RT, or even multiple previous RTs (see their Analysis 3), should not succeed unless timing biases are completely modelled (which they should not be according to the temporal-learning account). The obvious limitation, however, is that it is difficult to determine whether the remaining LLPC effect is due to temporal learning or conflict monitoring, as both accounts predict the same effect. In that sense, the current data are not sufficient to argue against a contribution of conflict monitoring to the LLPC effect. Instead, they demonstrate that concern is warranted. In this final section, I will discuss both converging evidence for a temporal-learning bias in the LLPC effect and some potentially problematic data for the simple learning view. In my view, some of the empirical lines of evidence provide compelling support for the temporal-learning view, though other results raise questions and further research will certainly be needed.

First, I have shown in a series of reports that one does not have to manipulate conflict proportions to produce a PC-like interaction. Simply manipulating the pace of the task with more 'easy' relative to 'hard' items, and vice versa, produces the same interaction pattern. As one example, Schmidt (2013c) used a simple letter identification task. On each trial, participants saw only a letter (D, F, J, or K) and were simply required to press the corresponding key on the keyboard. Unlike a conflict task (e.g., Stroop), there were no distracting stimuli and thus no conflict. The only manipulations were the contrast of the target digit on a given trial (high vs. low) and the proportion of high- vs. low-contrast trials (mostly easy vs. mostly hard). Of course, participants respond faster to high-contrast (easy to see) targets than to low-contrast (slightly harder to see) targets, but this contrast effect was also moderated by *proportion easy*. Just like a PC effect, the contrast effect was larger in the mostly easy context relative to the mostly hard context. This is exactly what the temporal-learning account would predict. Conflict is not relevant, only the pace, and the pace is faster in the mostly easy condition. The conflict-monitoring account, of course, cannot explain this finding, as there is no conflict to monitor or adjust to. Schmidt (2014) further confirmed that this proportion-easy effect is not specific to items by using the same sort of biased/transfer item design as described earlier for the LLPC procedure. What these results illustrate is a relatively pure example of why we should expect a PC-like interaction in a LLPC procedure even without conflict monitoring.

Of course, observing a temporal-learning effect in one procedure does not necessarily imply that the same learning effect will be observed in another procedure. On the other hand, at least some post hoc explanation seems necessary to explain why a general impact of trial pace on performance would be eliminated in a conflict task environment. Still, even if we assume that temporal-learning biases should equally well apply in proportion congruent experiments as they do in proportion-easy experiments, it does not necessarily follow that said learning biases explain *all* of the LLPC effect. Of course, this caveat should not lead us to either favour or disfavour the idea that conflict monitoring additionally contributes to the LLPC effect, but does leave open the possibility.

In that vein, a recent set of experiments by Schmidt (2017) aimed to more clearly adjudicate between a pure temporal-learning view and conflict monitoring. Prime-probe conflict tasks with direction word distracters and targets (essentially word–word direction Stroop) were conducted with the typical LLPC design. That is, some biased words (e.g., ‘up’ and ‘down’) were manipulated for PC and some intermixed transfer items (e.g., ‘left’ and ‘right’) were not manipulated. In a control condition, this produced a robust LLPC effect. In the critical ‘long wait’ condition, however, task pace was manipulated by presenting ‘wait cues’ on some of the biased item trials. Participants had to wait for a brief amount of time (until the cue disappeared) before making a response. This, at least roughly, served to match response speed and accuracy in the mostly congruent and mostly incongruent conditions. This eliminated the LLPC effect. Note that in the control (short wait) condition, the same wait cues were presented but more briefly. These experiments provided a clear dissociation between the pure temporal-learning and control views. According to the temporal-learning account, only the pace of responding matters. Thus, the LLPC effect should be eliminated. According to the conflict-monitoring view, however, conflict matters. The long wait manipulation preserved the conflict proportions, so a LLPC effect still should have been observed.

It might be argued that the wait cues somehow interfered with conflict processing or control adjustments (albeit only in the long wait condition). It is not clear why the conflict-monitoring account should predict this a priori, however. It is also worth noting that Cohen-Shikora and colleagues (2019) incorrectly state that 'there was no congruency effect in the longer-wait condition that eliminated the list-wide PC effect', such that there was no conflict to adjust to. This is incorrect. The speed of responding to congruent and incongruent trials was only equated for the filler 'wait trials', not by eliminating conflict but by requiring temporary withholding of the selected response. Congruency effects were large and robust for the critical test trials, and this congruency effect merely did not change in magnitude across the mostly congruent and mostly incongruent lists. One might alternatively assume that matching response times on the filler trials equates conflict in the two lists, but it is difficult to see how this should be the case. Response times were effectively matched for congruent and incongruent items with the wait manipulation, but not by eliminating the mismatch in stimulus or response information on incongruent trials (and also not by introducing a mismatch for congruent trials). One might postulate that the wait duration prevents conflict from occurring, but the timings of the stimulus events in the wait cue procedure seem to present difficulties for this view. For instance, the distracter was presented in advance of the target, and then removed from the screen (after 133 ms). After an additional blank screen (67 ms), the target was then presented briefly (133 ms) along with the wait cue, only the latter of which remained on the screen. Thus, participants could not know whether or not they could 'wait' while the distracter was on the screen and only had a brief amount of time to process the target. As such, it is not clear why conflict should not occur while processing said target. At minimum, supplemental ad hoc assumptions seem necessary to explain why conflict monitoring would not occur under these conditions. The wait cue data may not be the last word on the subject, but do currently favour a pure temporal-learning view. Another critique might also be that the wait cue data were from a prime-probe task, rather than, say, Stroop. The present author sees no compelling reason to favour Stroop, however, particularly when a robust LLPC effect was observed in prime-probe, then eliminated with the same materials. Follow-up research with different task procedures (e.g., Stroop, Simon, flanker, etc.) is certainly welcome, of course.

These added lines of evidence for temporal-learning biases supplement evidence from the modelling approach. With one exception to be discussed shortly, these studies represent the only investigations (that I am aware of) directly aiming to put the temporal-learning account of the LLPC effect to an empirical test, and all data point in one direction. Until conclusive evidence can be presented that are contrary to the pure temporal-learning view (along with some alternative explanation for the modelling results above), these findings should therefore be worrying for the conflict-monitoring perspective.

Another series of studies by Spinelli et al. (2019), however, took a different approach to controlling for temporal-learning effects, which produced results less favourable for a (pure) temporal-learning view. Their first two experiments made use of picture–word Stroop tasks, where participants either identified the category of (Experiment 1a) or named (Experiment 1b) pictures and ignored superimposed words that were either congruent or incongruent in meaning with the picture. Their primary evidence against a contribution of temporal learning to their LLPC effects was the absence of a congruency by previous RT interaction. However, they conducted their analyses with GLMM (in the identical fashion as Cohen-Shikora et al., 2019), which the present paper has shown to be problematic. An additional potential complication is that they did not use an inducer-diagnostic type design, but instead used a large set of non-repeated stimuli. Their reasoning for this is that eliminating stimulus repetitions eliminates the contingency biases. However, this is only true at the level of exact stimulus matches. At a categorical level, a contingency still exists, as illustrated in Table 1. In the categorization experiment, the distracting word category is predictive of the categorization response, similar to non-conflict categorical contingency learning experiments (Schmidt et al., 2018). Granted, in the naming experiment, the distracting word category is not predictive of a particular response, but rather only of the category of potential target responses. Thus, in addition to limitations with the control for temporal-learning biases, it is not completely clear whether the measured LLPC effect is free of indirect contingency biases. Interestingly, in a second experiment the authors failed to find 'proportion-easy' effects (discussed above) with a resolution manipulation (i.e., high- vs. low-resolution stimuli), similar to the abovementioned contrast manipulations, on their picture stimuli (i.e., with the distracting words removed) using a naming response. They did find evidence of timing biases in the proportion-easy experiment, but not of the same form as observed by Schmidt (2013c, 2014, 2016a). In particular, participants were overall slower in the mostly hard list, but the resolution effect was not modulated by proportion easy. This suggests that, at least in the context of their experiments with naming of a large set of non-repeated stimuli, the response criteria are set by participants in a different manner. In particular, the results seem consistent with the time criterion account (Lupker et al., 1997), according to which a fixed (i.e., rather than dynamic) criterion is set for each of the mostly easy and mostly hard lists. If so, then this could indeed prove to be a less problematic approach to controlling away temporal-learning biases: a timing effect is certainly present, but not one that should produce a spurious LLPC effect.

Globally, the interactive effects in proportion congruent and proportion-easy experiments are consistent with other findings in the timing literature, such as mixing costs. A *mixing cost* is the observation that performance on easy and hard

Table 1.
Contingency manipulation of Spinelli and colleagues (2019).

Picture category	Word category			
	Animal	Human being	Food	Man-made object
Mostly congruent				
Animal	27	3	3	3
Human being	3	27	3	3
Food	3	3	27	3
Man-made object	3	3	3	27
Mostly incongruent				
Animal	9	9	9	9
Human being	9	9	9	9
Food	9	9	9	9
Man-made object	9	9	9	9

items is reduced in mixed lists (i.e., with both easy and hard items intermixed) relative to pure lists (i.e., with pure easy and pure hard lists; Forrin, 1975; Grice, 1968; Grice & Hunter, 1964; Los, 1994, 1996, 1999a, b; Lupker et al., 1997; Niemi, 1981; Sanders, 1977; Van Duren & Sanders, 1988). The PEP model simulates mixing costs with the same mechanism that produces LLPC effects (Schmidt, De Houwer & Rothermund, 2016). Roughly, (fast) congruent trials in the mostly congruent list are akin to the (fast) easy trials in pure easy lists and (fast) incongruent trials in the mostly incongruent list are akin to the (fast) hard trials in pure hard lists. The (slower) congruent and incongruent trials in mixed lists are comparable to (slower) easy and hard items in mixed lists. Thus, the literature with such list mixing manipulations is also generally consistent with the basic premise of the temporal-learning account presented here. In some cases, though, an overall mixing cost is not observed, and *homogenization* is observed instead: easy items are notably faster in the pure easy lists relative to mixed lists, whereas hard items are somewhat *slower* in the pure lists (Chateau & Lupker, 2003; Kinoshita & Mozer, 2006; Lupker et al., 1997; Lupker et al., 2003; Rastle et al., 2003; Taylor & Lupker, 2001). This is effectively the same interaction minus the main effect of mixing, but this pattern is inconsistent with the LLPC effect and with the predictions of the temporal-learning account: incongruent trials should be responded to more slowly (not more quickly) in the mostly incongruent list if this homogenization pattern (without a mixing cost) is present.

The resolution data of Spinelli and colleagues' (2019) data are consistent with this homogenization-only pattern of results (i.e., slower responses to hard items in the mostly hard list). Notably, the homogenization-only pattern seems to only occur in studies with a large stimulus set, often but not exclusively with naming responses (Lupker et al., 2003), similar to that of Spinelli and colleagues, which may explain the discrepancy. The typical conflict paradigm used to study conflict monitoring is more akin to the procedures that have produced mixing costs and proportion-easy interactions than those that have found homogenization-only effects and that of Spinelli and colleagues. Indeed, as one added caveat, naming with large stimulus sets does not even produce a typical 'conflict' effect: congruent trials are responded to faster than incongruent trials, but *both* are responded to faster than neutral (Schmidt et al., 2013). This suggests positive priming, even for incongruent stimuli, and not conflict-driven interference. Whether it makes sense to talk about conflict monitoring when there is seemingly no conflict is thus uncertain, which may therefore be another reason to be cautious in interpreting findings with large stimulus lists like in Spinelli and colleagues.

On the other hand, Cohen-Shikora and colleagues (2019) discuss other empirical work that, though not directly related to temporal learning, may be interpreted in favour of the conflict-monitoring view. For instance, in completely unbiased lists (i.e., no PC manipulations) an 'LLPC effect' can be created via instructions that misleadingly tell participants about congruency proportions or attentional needs (Bugg et al., 2015; Entel et al., 2014). It is important to note, however, that the 'simple learning' view of LLPC effects does not propose that attention is, globally speaking, uncontrollable. Indeed, this would be an unsupportable view. The Stroop literature provides clear evidence of attentional control: participants can follow instructions to attend to the colour while ignoring the word or, conversely, to attend to the word and ignore the colour (i.e., in reverse Stroop; e.g., Blais & Besner, 2006). That instructions which (explicitly or implicitly) tell participants to increase or decrease attention to distracters lead to adjustments of attention (and thus the Stroop effect) is not surprising. It is also a different question than whether participants given a fixed goal of attending as best as they can to the target (while ignoring the distracter) dynamically adjust attention in response to monitored conflict.

Cohen-Shikora and colleagues (2019) also point to modulations of LLPC related to other factors which may be easily regarded as control-related. For instance, the LLPC effect in Hutchison (2011) was modulated by the working-memory capacity of participants. High-span participants produced a smaller LLPC effect, likely indicating more stable control of attention. Although only indirect, this might suggest that what the LLPC effect is measuring is related to attentional control. On the other hand, working-memory span could equally well influence temporal learning. Indeed, it has been observed that high working-memory capacity participants maintain focus on the target task better, whereas low working-memory capacity participants are less narrowly focused on the target task and can, perhaps unintuitively, learn more about task-irrelevant information such as timing regularities (Woehrle & Magliano, 2012). Alternatively, LLPC effects may have been reduced simply because congruency effects were much smaller in high working-memory capacity participants (e.g., due to better filtering of the distracter), leaving a smaller effect to be modulated by timing (or other mechanisms such as conflict monitoring). The above accounts, of course, assign a role of attention on working-memory capacity, but it does not clearly follow that modulations of LLPC effects by working-memory capacity imply that the base LLPC effect is due to conflict monitoring.

Relatedly, influences of the LLPC of one task can be observed on other tasks (Funes et al., 2010; Torres-Quesada et al., 2013; Wühr et al., 2015). For instance, Wendt et al. (2012) observed that the time to complete visual search of stimuli in normal flanking distracting positions was increased in mostly incongruent lists. Findings such as these might be less easy to explain in terms of simple temporal-learning biases, especially if the overall task pace is different for the two tasks. On the other hand, decisional processes (like temporal learning) have been proposed for a wide range of cross-task conditions like this outside of the conflict task domain (Kiger & Glass, 1981). For instance, response times on easy and hard math problem assessments are influenced by whether intermixed sentence verification trials are either uniformly easy or mixed easy and difficult (mixing cost). As stressed in the previously mentioned quote by Kiger and Glass, such observations are repeatedly observed in differing publications and paradigm-specific explanations (such as conflict monitoring) are repeatedly proposed that only explain the narrow effect of interest (e.g., LLPC effect) and not the trend across multiple unrelated paradigms.

Taken together, some of the most direct tests of timing biases on LLPC effects suggest strong support for the temporal-learning account. Other findings, though perhaps less direct tests of the dissociation between temporal learning and conflict monitoring, may or may not be as easily integrated into a simple learning view. For these reasons, further research is certainly needed. Indeed, the current picture is muddied by the inherent difficulty in controlling for a complicated influence like rhythmic timing. The hope of the present work is merely to highlight reasons why the temporal-learning account should not be tossed aside, as suggested by Cohen-Shikora and colleagues (2019; and as echoed by Spinelli et al., 2019).

11. Final Thoughts

In summary, the present article has aimed to show several things. First, the failure of previous RT to eliminate the LLPC effect in modelling approaches should not be taken as evidence in favour of conflict monitoring. The temporal-learning account (unlike the conflict-monitoring account) does predict that there should be some variance to capture in this way, but simply does not predict that this approach should work to fully eliminate the LLPC effect. Second, inverse (or similar) transformations of data are not inherently bad and can often even be the preferred or necessary approach. The specific application to LLPC effects is arguably the correct approach. Indeed, although certainly not convention, one may even argue that 'response rate' (with transformed data) is the more sensible way to assess response time data by default, or at least when investigating intertrial autocorrelations. Third, the inverse scaling does not actually distort the LLPC effect itself, but analyses on raw RT do distort the relation between previous and current RTs and this is supported by the significant decreases in the autocorrelations observed in the present report. Of particular import, the relationship between previous RT and the current-trial congruency effect is substantially decreased in the raw relative to inverse RT scale. It is this fact that explains why GLMM seemingly produces little more than noise in the temporal-learning tests, whereas the LME on transformed data provides clear and consistent evidence for temporal-learning biases across multiple datasets. Indeed, it is difficult to support the notion that we should prefer to quantify any variable on one scale (e.g., raw RT) when the same variable on another scale (e.g., inverse RT) explains notably more variance (using the same number of degrees of freedom). Fourth, even on raw RTs, an effect of temporal-learning biases is still observed if previous RT is allowed to first explain variance on the inverse scale. Thus, the variance that inverse previous RT explains in the LLPC effect does 'transform out' to the raw RT scale. Fifth, there are lines of converging evidence for a temporal-learning bias in the LLPC effect. While the literature as a whole paints an ambiguous picture as to whether the simple learning view is completely or only partially true, data do exist that seem problematic for the conflict-monitoring view.

In the concluding paragraph of Cohen-Shikora and colleagues (2019) the authors make a strong assertion:

> '[W]e cannot justify recommending that researchers adopt additional controls to account for temporal learning when investigating list-wide PC effects.'

Globally, I find it too strong to suggest that temporal-learning biases can be safely ignored on the basis of the extant data. It seems especially strong to favour results from one approach that produce largely null findings (GLMM) over another approach that produces relatively consistent evidence in favour of temporal learning (LME), especially without an explanation for why such a discrepancy in the results of two approaches exists in the first place. While I do agree that attempts to 'model away' temporal-learning biases with statistical models are challenging, compelling evidence for a temporal-learning bias does exist across a range of statistical modelling and experimental approaches, both inside and outside the attentional control domain. Of course, even if a temporal-learning bias does exist, said bias may or may not explain the entirety of the LLPC effect. Some early datasets are suggestive (esp., Schmidt, 2017) and others raise questions. I look forward to future results to further clarify this intriguing issue.

At a more general level, the present paper aimed to draw attention to two key considerations that are often overlooked in the literature. Firstly, decision-based processes, such as the setting of evidence accumulation criteria (e.g., as proposed by the temporal-learning account), have substantial influences on speeded response time behaviour. Unfortunately, such influences are often not orthogonal to manipulations of content. The present manuscript discussed the particular case of proportional manipulations of filler items in a LLPC manipulation, but the same concern can apply to other proportion/filler manipulations, along with other popular design types, such as with sequential manipulations (e.g., see Schmidt & Weissman, 2016). Indeed, as hinted at above by the warning of Kiger and Glass (1981), there is a very real danger that the same wheel will continue to be reinvented in numerous domains when the role of decision-related processes are eventually appreciated. Or, even more problematically, the role of decision-related processes may never be realized in many areas. To avoid such problems, more systematic consideration of decision-related processes seems warranted.

The second broader aim of the present work was to present a different view on data transformation. As I have argued in the present report, transformed data need not be viewed as a 'corruption' of a true raw effect. Depending on the research question, transformed data may be inappropriate in some cases, absolutely necessary in other cases, and in yet other cases the choice of whether to use raw or transformed data may be of little import. Though some have rightly pointed out scenarios in which transforms (such as inverse or log) are inherently problematic (Balota et al., 2013; Lo & Andrews, 2015), capable of inversing the direction of certain types of interactions, it is important to note that this concern is only applicable to certain scenarios.

Acknowledgements

This work was supported by the French 'Investissements d'Avenir' program, project ISITE-BFC (contract ANR15-IDEX-0003) to James R. Schmidt. I would like to thank Emily Cohen-Shikora for providing me the relevant R scripts and full datasets. R scripts for the reported analyses are available on the Open Science Framework (https://osf.io/4rgwa/).

Notes

1. Note this is somewhat of a simplification, as conflict might vary along a continuum (Yeung et al., 2011), though this description captures the rough idea.
2. This participant actually shows a below-average autocorrelation relative to the sample as a whole.
3. Readers interested in how these data were simulated can contact the author for more information.

4. Note that the negative sign is merely to preserve the original direction of the distribution and the 1000 in the numerator is simply to remove some of the decimal places from the inverted RTs. Changes to the numerator are not relevant for actual model fit (e.g., 1/RT is mathematically equivalent).
5. This was misreported in Schmidt (2013c) as 230, likely due to counting unique participant numbers on a participant list provided by Hutchison, which contained a different number of participants than the actual dataset for unknown reasons.
6. A dataset posted on the Open Science Framework (https://osf.io/b9zyv/) seemingly has 95 participants, but I used the same data as Cohen-Shikora and colleagues (2019). It is uncertain where the two extra participants (in Experiment 1b) come from.
7. As normally plotted with PC as the *x*-axis categories, the bars/lines for congruent and incongruent trials do not touch each other, but the interaction is still crossover because the lines do cross when switching congruency and PC in the plots (see Loftus, 1978, for further explanation).
8. Which transform to use, of course, depends on the specific assumptions of how general slowing impacts observed effects. The choice of transforms may, therefore, be ambiguous, unless the model of cognitive slowing directly implies a specific transform.
9. For instance, I only consider drift rate to a fixed boundary, and do not discuss nondecision time, starting points, etc.
10. Note that the Gamma family corrects the statistical assumptions of the regression, but the identity link function does explicitly specify that previous and current RT should be related to each other linearly, as depicted in the scatterplots.
11. The difference in correlations is even larger for this dataset (27%) if the Q–Q plots are better normalized with a 375 ms trim, but I have stuck with a 300 ms trim for consistency with Cohen-Shikora and colleagues (2019).
12. See an Excel document in the OSF link with an example demonstration of this.
13. Giacomo Spinelli suggested some simulated data to illustrate this point, which I have reproduced and extended (see Excel document in the OSF repository). In particular, it is possible to create a situation (albeit somewhat artificial) in which there is an autocorrelation in RTs and a main effect of congruency that are additive, which results in a more underadditive interaction after an inverse transform. On the other hand, these artificial situations do not produce the large modulations of the LLPC effect when analyzed like in the analyses to follow.

14. It might also be worth mentioning that the GLMM consistently failed to converge in all analyses including previous RT, albeit less severely in random intercept models. In contrast, the model converges with inverse-transformed data in all LME models.

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychol. Bull.*, *142*, 693–728. doi: 10.1037/bul0000047.

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *J. Exp. Psychol. Learn. Mem. Cogn.*, *38*, 152–163. doi: 10.1037/a0024953.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *J. Exp. Psychol. Learn. Mem. Cogn.*, *39*, 1563–1571. doi: 10.1037/a0032186.

Blais, C., & Besner, D. (2006). Reverse Stroop effects with untranslated responses. *J. Exp. Psychol. Hum. Percept. Perform.*, *32*, 1345–1353. doi: 10.1037/0096-1523.32.6.1345.

Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *J. Cogn. Neurosci.*, *22*, 2758–2767. doi: 10.1162/jocn.2009.21365.

Borowsky, R., & Besner, D. (2006). Parallel distributed processing and lexical–semantic effects in visual word recognition: are a few stages necessary? *Psychol. Rev.*, *113*, 181–193. doi: 10.1037/0033-295X.113.1.181.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.*, *108*, 624–652. doi: 10.1037/0033-295x.108.3.624.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychol. Rev.*, *114*, 539–576. doi: 10.1037/0033-295X.114.3.539.

Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *J. Exp. Psychol. Learn. Mem. Cogn.*, *40*, 567–587. doi: 10.1037/a0035032.

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

Bugg, J. M., & Crump, M. J. C. (2012). In support of a distinction between voluntary and stimulus-driven control: a review of the literature on proportion congruent effects. *Front. Psychol.*, *3*, 367. doi: 10.3389/fpsyg.2012.00367.

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Mem. Cogn.*, *36*, 1484–1494. doi: 10.3758/MC.36.8.1484.

Bugg, J. M., McDaniel, M. A., Scullin, M. K., & Braver, T. S. (2011). Revealing list-level control in the Stroop task by uncovering its benefits and a cost. *J. Exp. Psychol. Hum. Percept. Perform.*, *37*, 1595–1606. doi: 10.1037/a0024670.

Bugg, J. M., Diede, N. T., Cohen-Shikora, E.., & Selmeczy, D. (2015). Expectations and experience: Dissociable bases for cognitive control? *J. Exp. Psychol. Learn. Mem. Cogn.*, *41*, 1349–1373. doi: 10.1037/xlm0000106.

Chateau, D., & Lupker, S. J. (2003). Strategic effects in word naming: Examining the route-emphasis versus time-criterion accounts. *J. Exp. Psychol. Hum. Percept. Perform.*, *29*, 139–151. doi: 10.1037/0096-1523.29.1.139.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Can. J. Psychol.*, *40*, 343–367. doi: 10.1037/h0080103.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *Am. J. Sociol.*, *100*, 1261–1293. doi: 10.1086/230638.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychol. Rev.*, *97*, 332–361. doi: 10.1037/0033-295X.97.3.332.

Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect. *J. Exp. Psychol. Learn. Mem. Cogn.*, *45*, 1703–1723. doi: 10.1037/xlm0000670.

Entel, O., Tzelgov, J., & Bereby-Meyer, Y. (2014). Proportion congruency effects: Instructions may be enough. *Front. Psychol.*, *5*, 1108. doi: 10.3389/fpsyg.2014.01108.

Forrin, B. (1975). Naming latencies to mixed sequences of letters and digits. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V* (pp. 345–356). New York, NY, USA: Academic Press.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychol. Rev.*, *118*, 614–636. doi: 10.1037/a0025255.

Funes, M. J., Lupiáñez, J., & Humphreys, G. (2010). Analyzing the generality of conflict adaptation effects. *J. Exp. Psychol. Hum. Percept. Perform.*, *36*, 147–161. doi: 10.1037/a0017598.

Gilden, D. L. (1997). Fluctuations in the time required for elementary decisions. *Psychol. Sci.*, *8*, 296–301. doi: 10.1111/j.1467–9280.1997.tb00441.x.

Gilden, D. L. (2001). Cognitive emissions of 1/*f* noise. *Psychol. Rev.*, *108*, 33–56. doi: 10.1037/0033-295X.108.1.33.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise in human cognition. *Science*, *267*, 1837–1839. doi: 10.1126/science.7892611.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *J. Exp. Psychol. Hum. Percept. Perform.*, *8*, 875–894. doi: 10.1037/0096-1523.8.6.875.

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

Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychol. Rev.*, *75*, 359–373. doi: 10.1037/h0026287.

Grice, G. R., & Hunter, J. J. (1964). Stimulus intensity effects depend upon the type of experimental design. *Psychol. Rev.*, *71*, 247–256. doi: 10.1037/h0047547.

Grosjean, M., Rosenbaum, D. A., & Elsinger, C. (2001). Timing and reaction time. *J. Exp. Psychol. Gen.*, *130*, 256–272. doi: 10.1037/0096-3445.130.2.256.

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

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychol. Bull.*, *109*, 340–347. doi: 10.1037/0033-2909.109.2.340.

Holway, A. H., & Pratt, C. C. (1936). The Weber ratio for intensitive discrimination. *Psychol. Rev.*, *43*, 322–340. doi: 10.1037/h0059748.

Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *J. Exp. Psychol. Learn. Mem. Cogn.*, *37*, 851–860. doi: 10.1037/a0023437.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *J. Exp. Psychol. Gen.*, *132*, 47–70. doi: 10.1037/0096-3445.132.1.47.

Kiger, J. I., & Glass, A. L. (1981). Context effects in sentence verification. *J. Exp. Psychol. Hum. Percept. Perform.*, *7*, 688–700. doi: 10.1037/0096-1523.7.3.688.

Kinoshita, S., & Lupker, S. J. (2003). Priming and attentional control of lexical and sublexical pathways in naming: A reevaluation. *J. Exp. Psychol. Learn. Mem. Cogn.*, *29*, 405–415. doi: 10.1037/0278-7393.29.3.405.

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

Kinoshita, S., Forster, K. I., & Mozer, M. C. (2008). Unconscious cognition isn't that smart: Modulation of masked repetition priming effect in the word naming task. *Cognition*, *107*, 623–649. doi: 10.1016/j.cognition.2007.11.011.

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. *J. Exp. Psychol. Gen.*, *140*, 622–636. doi: 10.1037/a0024230.

Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Vis. Cogn.*, *18*, 655–681. doi: 10.1080/13506280902986058.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *J. Exp. Psychol. Hum. Percept. Perform.*, *20*, 219–234. doi: 10.1037/0096-1523.20.2.219.

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

Loftus, G. R. (1978). On interpretation of interactions. *Mem. Cogn.*, *6*, 312–319. doi: 10.3758/BF03197461.

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. *Mem. Cogn.*, *7*, 166–174. doi: 10.3758/BF03197535.

Logan, G. D., Zbrodoff, N. J., & Williamson, J. (1984). Strategies in the color-word Stroop task. *Bull. Psychon. Soc.*, *22*, 135–138. doi: 10.3758/BF03333784.

Los, S. A. (1994). Procedural differences in processing intact and degraded stimuli. *Mem. Cogn.*, *22*, 145–156. doi: 10.3758/BF03208886.

Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychol.*, *94*, 145–188. doi: 10.1016/0001-6918(95)00050-X.

Los, S. A. (1999a). Identifying stimuli of different perceptual categories in mixed blocks of trials: Evidence for cost in switching between computational processes. *J. Exp. Psychol. Hum. Percept. Perform.*, *25*, 3–23. doi: 10.1037/0096-1523.25.1.3.

Los, S. A. (1999b). Identifying stimuli of different perceptual categories in pure and mixed blocks of trials: evidence for stimulus-driven switch costs. *Acta Psychol.*, *103*, 173–205. doi: 10.1016/S0001-6918(99)00031-1.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Can. J. Psychol.*, *36*, 684–700. doi: 10.1037/h0080661.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *J. Exp. Psychol. Learn. Mem. Cogn.*, *23*, 570–590. doi: 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. *J. Mem. Lang.*, *49*, 556–575. doi: 10.1016/S0749-596X(03)00094-9.

Masson, M. E. J., & Kliegl, R. (2013). Modulation of additive and interactive effects in lexical decision by trial history. *J. Exp. Psychol. Learn. Mem. Cogn.*, *39*, 898–914. doi: 10.1037/a0029180.

Mozer, M. C., Kinoshita, S., & Davis, C. (2004). Control of response initiation: Mechanisms of adaptation to recent experience. In K. Forbus, D. Gentner, & T. Regier (Eds), *Proceedings of the Twenty Sixth Conference of the Cognitive Science Society* (pp. 981–986). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Niemi, P. (1981). Constant vs. variable stimulus intensity and visual simple reaction time. *Percept. Mot. Skills*, *53*, 615–619. doi: 10.2466/pms.1981.53.2.615.

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

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.*, *85*, 59–108. doi: 10.1037/0033-295X.85.2.59.

Ridderinkhof, K. R., Vandermolen, M. W., & Bashore, T. R. (1995). Limits on the application of additive factors logic: Violations of stage robustness suggest a dual-process architecture to explain flanker effects on target processing. *Acta Psychol.*, *90*, 29–48. doi: 10.1016/0001-6918(95)00031-O.

Risko, E. F., Blais, C., Stolz, J. A., & Besner, D. (2008). Nonstrategic contributions to putatively strategic effects in selective attention tasks. *J. Exp. Psychol. Hum. Percept. Perform.*, *34*, 1044–1052. doi: 10.1037/0096-1523.34.4.1044.

Robidoux, S. (2017). *Geometry and harmony: The consequences of transforming data*. Blog post retrieved from http://serjerobidoux.blogspot.com/2017/03/geometry-and-harmony-consequences-of.html.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds), *The handbooks of aging. Handbook of the psychology of aging* (2nd ed., pp. 400–426). New York, NY, USA: Van Nostrand Reinhold.

Sanders, A. F. (1977). Structural and functional aspects of the reaction process. In S. Dornic (Ed.), *Attention and Performance VI* (pp. 3–25). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Schmidt, J. R. (2013a). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered. *Psychon. Bull. Rev.*, *20*, 615–630. doi: 10.3758/s13423-012-0373-0.

Schmidt, J. R. (2013b). The Parallel Episodic Processing (PEP) model: Dissociating contingency and conflict adaptation in the item-specific proportion congruent paradigm. *Acta Psychol.*, *142*, 119–126. doi: 10.1016/j.actpsy.2012.11.004.

Schmidt, J. R. (2013c). Temporal learning and list-level proportion congruency: Conflict adaptation or learning *when* to respond? *Plos One*, *8*, e0082320. doi: 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. *J. Cogn. Psychol.*, *26*, 373–385. doi: 10.1080/20445911.2014.896367.

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

Schmidt, J. R. (2016b). Proportion congruency and practice: A contingency learning account of asymmetric list shifting effects. *J. Exp. Psychol. Learn. Mem. Cogn.*, *42*, 1496–1505. doi: 10.1037/xlm0000254.

Schmidt, J. R. (2017). Time-out for conflict monitoring theory: Preventing rhythmic biases eliminates the list-level proportion congruent effect. *Can. J. Exp. Psychol.*, *71*, 52–62. doi: 10.1037/cep0000106.

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychon. Bull. Rev.*, *26*, 753–771. doi: 10.3758/s13423-018-1520-z.

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *J. Exp. Psychol. Learn. Mem. Cogn.*, *34*, 514–523. doi: 10.1037/0278-7393.34.3.514.

Schmidt, J. R., & De Houwer, J. (2016). Time course of colour–word contingency learning: Practice curves, pre-exposure benefits, unlearning, and relearning. *Learn. Motiv.*, *56*, 15–30. doi: 10.1016/j.lmot.2016.09.002.

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

Schmidt, J. R., Cheesman, J., & Besner, D. (2013). You can't Stroop a lexical decision: Is semantic processing fundamentally facilitative? *Can. J. Exp. Psychol.*, *67*, 130–139. doi: 10.1037/a0030355.

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. *Cogn. Psychol.*, *91*, 82–108. doi: 10.1016/j.cogpsych.2016.10.004.

Schmidt, J. R., Augustinova, M., & De Houwer, J. (2018). Category learning in the colour–word contingency learning paradigm. *Psychon. Bull. Rev.*, *25*, 658–666. doi: 10.3758/s13423-018-1430-0.

Shor, R. E. (1975). An auditory analog of the Stroop test. *J. Gen. Psychol.*, *93*, 281–288.

Silver, N. C., Hittner, J. B., & May, K. (2004). Testing dependent correlations with nonoverlapping variables: a Monte Carlo simulation. *J. Exp. Educ.*, *73*, 53–69. doi: 10.3200/JEXE.71.1.53-70.

Smid, H. G. O. M., Lamain, W., Hogeboom, M. M., Mulder, G., & Mulder, L. J. M. (1991). Psychophysiological evidence for continuous information-transmission between visual search and response processes. *J. Exp. Psychol. Hum. Percept. Perform.*, *17*, 696–714. doi: 10.1037/0096-1523.17.3.696.

Spinelli, G., & Lupker, S. J. (in press). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and color–word correlation confounds. *J. Exp. Psychol. Learn. Mem. Cogn.* doi: 10.1037/xlm0000820.

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. *J. Exp. Psychol. Hum. Percept. Perform.*, *45*, 995–1014. doi: 10.1037/xhp0000656.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychol.*, *30*, 276–315. doi: 10.1016/0001-6918(69)90055–9.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Stevens, A., Schwarz, J., Schwarz, B., Ruf, I., Kolter, T., & Czekalla, J. (2002). Implicit and explicit learning in schizophrenics treated with olanzapine and with classic neuroleptics. *Psychopharmacology*, *160*, 299–306. doi: 10.1007/s00213-001-0974-1.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.*, *18*, 643–662. doi: 10.1037/h0054651.

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *J. Exp. Psychol. Learn. Mem. Cogn.*, *27*, 117–138. doi: 10.1037/0278-7393.27.1.117.

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

Urry, K., Burns, N. R., & Baetu, I. (2015). Accuracy-based measures provide a better measure of sequence learning than reaction time-based measures. *Front. Psychol.*, *6*, 1158. doi: 10.3389/fpsyg.2015.01158.

Van Duren, L. L., & Sanders, A. F. (1988). On the robustness of the additive factors stage structure in blocked and mixed choice reaction designs. *Acta Psychol.*, *69*, 83–94. doi: 10.1016/0001-6918(88)90031–5.

Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychol. Rev.*, *114*, 830–841. doi: 10.1037/0033-295X.114.3.830.

Wendt, M., Luna-Rodriguez, A., & Jacobsen, T. (2012). Conflict-induced perceptual filtering. *J. Exp. Psychol. Hum. Percept. Perform.*, *38*, 675–686. doi: 10.1037/a0025902.

West, R., & Baylis, G. C. (1998). Effect of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychol. Aging*, *13*, 206–217. doi: 10.1037//0882-7974.13.2.206.

Woehrle, J. L. & Magliano, J. P. (2012). Time flies faster if a person has a high working-memory capacity. *Acta Psychol.*, *139*, 314–319. doi: 10.1016/j.actpsy.2011.12.006.

Wühr, P., Duthoo, W., & Notebaert, W. (2015). Generalizing attentional control across dimensions and tasks: Evidence from transfer of proportion-congruent effects. *Q. J. Exp. Psychol.*, *68*, 779–801. doi: 10.1080/17470218.2014.966729.

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