View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Ghent University Academic Bibliography

# Removing the Influence of Feature Repetitions on the Congruency Sequence Effect: Why Regressing Out Confounds From a Nested Design Will Often Fall Short

James R. Schmidt,[1] Maarten De Schryver,[1] & Daniel H. Weissman[2]

[1]Department of Experimental Clinical and Health Psychology, Ghent University, Belgium

[2]Department of Psychology, University of Michigan

## Author Note

James R. Schmidt is a postdoctoral researcher of the Research Foundation – Flanders (FWO – Vlaanderen). Maarten De Schryver is supported by Methusalem Grant BOF09/01M00209 of Ghent University. We would like to thank Wim Notebaert and Tom Verguts for their input on this paper and for providing us with added data, which allowed us to check our fitting procedure (though we opted not to calculate a lack-of-fit test for these data for the reasons described in Footnotes 1 and 3). We would also like to thank J. Toby Mordkoff for providing us with extra data. Correspondence concerning this article can be addressed to James R. Schmidt, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. E-mail: james.schmidt@ugent.be.

## Abstract

This paper illustrates a shortcoming of using regression to control for confounds in nested designs. As an example, we consider the congruency sequence effect (CSE), which is the observation that the congruency effect in distracter interference (e.g., Stroop) tasks is smaller following incongruent as compared to congruent trials. The CSE is often interpreted as indexing conflict adaptation: a relative increase of attention to the target following incongruent trials. However, feature repetitions across consecutive trials can complicate this interpretation. To control for this confound, the standard procedure is to delete all trials with a stimulus or response repetition and analyze the remaining trials. Notebaert and Verguts (2007) present an alternative method that allows researchers to use all trials. Specifically, they employ multiple regression to model conflict adaptation independent of feature repetitions. We show here that this approach fails to account for certain feature repetition effects. Further, modelling these additional effects is typically not possible due to an upper bound on the number of degrees of freedom in the experiment. These findings have important implications for future investigations of conflict adaptation and, more broadly, for all researchers who attempt to regress out confounds in nested designs.

**Keywords:** congruency sequence effects, conflict adaptation, feature repetitions, regression, lack-of-fit tests, nesting

# Introduction

For almost every major finding in experimental psychology there exist multiple accounts. What is not always clear is how to best distinguish among these accounts. For instance, some have argued that the debate over exemplar versus abstraction accounts of category representation is essentially unresolvable, because any effect that appears to support one account one could just as easily be explained by the other (Barsalou, 1990). Some debates may appear easier to disentangle. However, we argue in the present article that certain approaches to distinguishing competing accounts are not as definitive as they initially appear.

For example, consider the all-too-frequent situation in which an experimental psychologist wishes to determine whether an experimental effect is better explained by a confounding variable in the study design than by the variable of interest. In this situation, a psychologist might devise an experiment in which the proposed confound can no longer influence the effect of interest. If the effect is still observed, then it cannot be explained by the (now removed) confound. An alternative approach employs regression to model and remove the influence of the confounding variable in the biased (i.e., confounded) data set. The present article highlights some limitations of this second, regression approach for nested designs using an example from the literature on congruency sequence effects (CSEs; Notebaert & Verguts, 2007). In the General Discussion, we review the broader implications of our findings for researchers in a variety of domains, including those studying working memory updating, size estimation, and resting-state functional connectivity as measured by functional magnetic resonance imaging (fMRI).

As illustrated in Figure 1, the CSE is the observation that the congruency effect in distracter-interference tasks is reduced when the previous trial was incongruent relative to congruent (Gratton, Coles, & Donchin, 1992). The CSE is often interpreted as indexing *conflict adaptation*, a process whereby the distribution of attention to the target and/or distracter is adjusted after incongruent trials (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001). Specifically, after experiencing heightened response conflict in a previous incongruent trial (relative to a previous congruent trial), participants increase attention to the target and/or decrease attention to the distracter, leading to a smaller congruency effect in the current trial.

**(Figure 1)**

Some researchers, however, argue that feature repetition confounds may explain the CSE better than conflict adaptation (Mayr, Awh, & Laurey, 2003; Hommel, Proctor, & Vu, 2004; for a review, see Schmidt, 2013a). Feature repetitions occur when the target and/or distracter repeat across consecutive trials. In the Stroop task, for example, the target colour may repeat from one trial to the next (*target-target repetition*), the distracter word may repeat from one trial to the next (*distracter-distracter repetition*), the distracter word on the previous trial may match the target colour on the current trial (*distracter-target repetition*), or the target colour on the previous trial may match the distracter word on the current trial (*target-distracter repetition*). *Complete repetition trials*, in which both the target and distracter repeat, are linked to relatively fast performance and occur only when the previous and current trial are both congruent (cC trials) or both incongruent (iI trials). In contrast, *partial repetition* trials, in which the target from the previous trial repeats while the distracter alternates, or vice-versa, are linked to relatively slow performance and occur frequently when the congruency of the previous trial does not match the congruency of the current trial (cI and iC trials). Thus, an unequal distribution of different types of feature repetitions across cC, cI, iC, and iI trials may account for the CSE better than conflict adaptation (Mayr et al., 2003; Hommel et al., 2004).

A key question, then, is whether conflict adaptation plays any role in producing the CSE independent of feature repetition biases. To answer this question, an unbiased measure of the CSE is needed that controls for feature repetition effects. The most common approach is to calculate the CSE after both *complete repetition* and *partial repetition* trials have been deleted, leaving only *complete alternation* trials, in which neither the target nor the distracter is repeated from the previous trial. Obtaining a significant CSE when employing this "repetition deletion technique" could not be explained by feature repetitions and would therefore suggest that conflict adaptation may indeed contribute to the CSE.

Findings from the repetition deletion technique have been mixed. On the one hand, several researchers employing this approach have reported significant CSEs (e.g., Akçay & Hazeltine, 2007; Freitas, Bahar, Yang, & Bahar, 2007; Kerns, Cohen, MacDonald, Cho, Stenger, & Carter, 2004; Notebaert, Gevers, Verbruggen, & Liefooghe, 2006; Verbruggen, Notebaert, Liefooghe, & Vandierendonck, 2006). On the other hand, most of the tasks employed by these researchers contained contingency learning confounds (see Schmidt, 2013 for a review), and subsequent studies have shown that removing both feature repetition and contingency learning confounds usually eliminates the CSE (e.g., Mayr et al., 2003; Mordkoff, 2012; Schmidt & De Houwer, 2011). Thus, some researchers have suggested that the CSE reflects learning and memory processes related to feature repetitions and contingency learning biases rather than conflict adaptation (e.g., Schmidt, 2013).

Notebaert and Verguts (2007), however, argue that the repetition deletion technique is problematic for two reasons. First, statistical power is reduced. This is a valid point, given that a very large proportion of trials must be deleted to remove all feature repetitions. This problem can be counteracted with longer experiments and/or more participants, but these procedures are obviously suboptimal. Second, the CSE is assessed on just one type of transition: *complete alternation* trials. Since the conflict adaptation account should apply equally to all types of stimulus transitions, the importance of this point is less clear to us. Nevertheless, we agree with the first point that the reduction of power is problematic.

To address the two shortcomings above, Notebaert and Verguts (2007) introduced a multiple regression approach (henceforth, NV regression) for estimating the contribution of conflict adaptation to reaction time (RT) independent of feature repetition confounds and without deleting trials. In this approach, different variables that may influence RT are coded by distinct binary regressors. For example, the regression they present employs eight regressors: (1) target-target repetitions, (2) distracter-distracter repetitions, (3) feature integration (complete repetitions and complete alternations versus partial repetitions), (4) negative priming (distracter-target repetitions, wherein both the previous and the current trial are incongruent), (5) target-distracter repetitions, (6) previous congruency, (7) current congruency, and (8) the CSE (i.e., the previous congruency x current congruency interaction). The aim of the NV regression is to determine whether the regression coefficient associated with the CSE regressor is significant while controlling for the other variables. Notebaert and Verguts reported a significant regression coefficient for the CSE regressor and concluded (with some reservations) that conflict adaptation exists.

There is a potential problem, however, with the statistical assumptions made by the NV regression approach. Each regressor, or variable, in the model represents a simple binary effect. The model therefore assumes that the effect of each variable (e.g., target-target repetitions) is the same regardless of the level of any other variable (e.g., previous congruency, current congruency, or other types of feature repetitions). Put another way, the model does not allow for interactions between different variables. This is a problem of *nesting*. Nesting occurs in a statistical model when only a restricted version of the full factorial model is tested. In the case of the NV regression, the statistical model is nested because it only tests for a few of the simpler main effects, and does not consider the potential interactions between the variables. The term “nesting” is also used to describe study designs in which a full factorial crossing of study factors produces empty cells (i.e., impossible or unobserved combinations of factors). Nesting in the design makes it impossible to test a full factorial model, ultimately requiring nesting in the statistical model. As we explain next, the NV regression approach may not always be valid, due to a nested design problem.

In particular, the presence of interactions involving two or more variables in the NV regression (e.g., a target-target repetition effect that is larger if there is also a target-distracter repetition) could complicate the interpretation of the CSE regressor. Specifically, the CSE regressor could “steal” variance in RT that originates from un-modelled interactions between different types of feature repetitions if it is correlated with these interactions. A significant beta parameter for the CSE regressor in such a situation could then potentially reflect interactions between different types of feature repetitions, rather than conflict adaptation. In sum, although the NV regression aims to assess the CSE independent of feature repetition confounds, it could fail to do so if un-modelled interactions involving feature repetitions contribute to variance in RT.

This problem is only worrisome when the CSE is positively correlated with un-modelled interactions involving feature repetitions. While some un-modelled interactions might be uncorrelated with the CSE (in which case there is no confound) and others might reduce the CSE, on the whole feature repetitions tend to engender a positive CSE (e.g., Mayr et al., 2003). Thus, if they are positively correlated with the CSE, then un-modelled interactions involving feature repetitions will likely increase (rather than decrease) the CSE, making it appear as though conflict adaptation contributes to the CSE when it does not. The present findings provide some preliminary evidence for this possibility.

Before continuing our discussion of Notebaert & Vergut’s (2007) approach, we would like to note that the issues we are discussing are not specific to the CSE. Rather, they are more general concerns that apply whenever a nested model is employed to eliminate the influence of a confounding variable that is only incompletely modelled. Thus, the implications of our analysis are important for any study that employs a similar regression approach. To illustrate this point, we describe in the General Discussion how our findings are relevant to other subfields of experimental psychology as well as to research in cognitive neuroscience. But, first, we return to the potential limitation of the NV regression approach we have identified.

Is there a way to avoid this limitation? The only way is to model all possible interactions among the four types of feature repetitions, two levels of previous congruency, and two levels of current congruency. That is, one could model all the main effects and interactions between each of the four types of feature repetitions, previous congruency, and current congruency in a 2 x 2 x 2 x 2 x 2 x 2 (and thus 64-cell) design. Sixty-four regressors would be required to model all of the possible combinations of these six binary factors. These regressors would include 1 intercept, 6 main effects, 15 two-way interactions, 20 three-way interactions, 15 four-way interactions, 6 five-way interactions, and 1 six-way interaction. By including all 64 regressors, a researcher could ensure that the conflict adaptation regressor is not influenced by feature repetition biases. However, this is not possible for two reasons.

First, some of the interactions involving feature repetitions cannot be observed in a real experiment. For instance, a target-target repetition and a distracter-distracter repetition cannot simultaneously occur in a single iC trial. Similarly, a distracter-target repetition and a distracter-distracter repetition cannot simultaneously occur in a single incongruent trial. As Table 1 illustrates, only 15 of the 64 possible combinations of feature repetitions, previous congruency, and current congruency can be observed in an experiment. In statistical terms, this means that the design is nested. In effect, the NV regression approach estimates the mean RT in each of these 15 conditions with the 8 regressors that Notebaert and Verguts (2007) include in their model. Indeed, every observation in the data corresponds to one of these 15 points on the regression line[1] because, as we stated earlier, only 15 of the 64 cells in the design can be observed in a real experiment (we revisit this important point later). Returning to the main issue, though, 49 (i.e., 64 – 15) of the cells are *not* observable, meaning that the data required to compute most of the interactions involving feature repetitions are missing.

**(Table 1)**

Second, it is impossible to include all 64 main effect and interaction regressors in the model without exceeding the experimental degrees of freedom. Given that only 15 unique combinations of the regressors are present in the data from an actual experiment, the model has only 14 degrees of freedom before *any* regressors are included. Thus, to preserve at least one degree of freedom, a maximum of 13 regressors can be included,[2] which is far less than 64. If more than 13 regressors are included, then the regression model will be able to maximize on random error and perfectly estimate the means for all 15 unique trial types. This is the case because each regressor in a model can connect, at a minimum, two points (in this case, conditions). Thus, if the number of regressors in a model is one less than the number of points to estimate (i.e., if there are 14 regressors to estimate 15 "points" or conditions), then the regression will, by definition, be able to perfectly estimate the means for all 15 conditions (i.e., connect all the points), even if the regressors included in the model are meaningless. In statistical terms, this type of regression model is called a *saturated model*. Three conclusions follow from this line of reasoning: (1) it is impossible for more than 14 regressors to explain variance when there are only 15 cell means to estimate, even if more than 14 factors play a real role in producing these means, (2) if more than 14 regressors are included, then all between-condition variance will be explained, but this will likely be the arbitrary result of maximization on random error, and (3) if more than 14 regressors are included, then a test of model misspecification, described shortly, is meaningless because the model will always be able to explain all between-condition variance (a result of the previous points).

In contrast, note that with the repetition deletion technique there is no nesting. As can be seen in Table 1, the cells retained in this analysis (grey) vary only in current congruency and previous congruency, both of which are fully modelled. One might therefore conclude that the repetition deletion analysis is superior, because nothing is left un-modelled. As mentioned earlier, however, the NV regression approach provides greater statistical power. Further, the preceding criticisms of NV regression are moot if complex feature repetition biases do not actually exist. That is, if: (a) interactions involving feature repetitions have no effect on the data, and (b) the variables coded for in the NV regression capture all of the *actual* feature repetition biases, then the NV regression is completely valid. Thus, the goal of the present study was to determine whether or not the NV regression is ever invalid. Such a result would indicate that caution should be exercised when employing this approach to assess the contribution of conflict adaptation to the CSE.

To test the validity of the NV regression approach, we first applied Notebaert and Verguts' (2007) model to the correct RT data from Experiments 1 and 2 of Schmidt and De Houwer (2011) and to the 25% congruency (i.e., contingency-free) condition from Mordkoff (2012).[3] Since our main aim was to test the statistical assumptions made by the model, we conducted a *lack-of-fit test* (not to be confused with a goodness-of-fit test), which assesses whether a model misestimates observations in a systematic way (e.g., see Faraway, 2004). If it does, then the model is *misspecified*, meaning that one or more additional regressors are needed to explain systematic un-modelled variance in the data. Of course, some degree of error in condition estimates is expected even in a correctly-specified model due to random noise. However this error should be random. Critically, a lack-of-fit test can tell the difference between random and systematic error by separating the degree of model misspecification from the between-participant error (Faraway, 2004). This is achieved with a statistical test that determines whether the degree to which conditional means are inaccurately predicted (termed *lack-of-fit variance*) exceeds the degree expected based on random variation between participants (termed *pure error variance*).

Whether or not the lack-of-fit test is significant depends on whether the model is correctly specified. If the model *is* correctly specified, then the *F*-value for the lack-of-fit test should be about 1 and, hence, not significant. This result should be obtained if the NV regression accounts for all relevant feature repetition effects separately from the CSE regressor. Of course, a null result does not indicate the model is correct. It merely indicates that no evidence for an error in model specification was observed. In contrast, if there is significantly more misspecification than expected based on random noise between participants, then the *F*-test will be significant. This result should only be obtained if one or more important effects involving feature repetitions has been excluded from the model.

How might a lack-of-fit test be conducted on the NV regression model? Recall that there are 15 experimentally-producible combinations out of the 64 possible combinations of previous congruency, current congruency, and the four repetition types. For this reason, one can use the regression equation from the NV approach, which contains eight regressors, to estimate the cell means in each of these 15 conditions (the condition-specific weighting for each of the eight regressors is presented in Table 2). Subtracting the model-estimated mean RT in each condition from each participant's mean RT in that same condition will reveal how much each participant's RT differed from the model's RT. Finally, a repeated-measures one-way ANOVA on these difference scores will reveal whether they differ from a flat line.

**(Table 2)**

The lack-of-fit test will lead to one of two outcomes. First, if the NV regression is correctly-specified, then the mean difference scores across participants in each of the 15 conditions above will not differ from zero more than would be expected from random, between-participant error.[4] Thus, if the 15 means are submitted to a one-way ANOVA, the ANOVA should return an *F*-value of roughly 1, so long as an appropriate correction to the degrees of freedom is made to account for the fact that eight regression parameters are employed to estimate the fifteen condition means.[5] Second, if the NV regression is not correctly specified, then at least some of the mean difference scores will differ from zero (and, hence, from each other) more than would be expected based on random, between-participant error. Thus, if the 15 means are submitted to a one-way ANOVA, the ANOVA should return a significant *F*-value, consistent with a significant degree of model misspecification. To our knowledge, the present use of a lack-of-fit test on repeated measures data is unique. However, this approach is analogous to previous uses of lack-of-fit tests on other types of data (e.g., see, Faraway, 2004; see also, Footnote 4 for a demonstration that this approach works as intended).

## Method and Results

### NV regression

We used three different data sets for our analyses. We refer to these as Experiment 1 (Schmidt & De Houwer, 2011, Experiment 1), Experiment 2 (Schmidt & De Houwer, 2011, Experiment 2), and Experiment 3 (Mordkoff, 2012, 25% congruency condition). All three experiments employed four alternative-forced-choice (4-AFC) tasks in which each distracter was presented equally often with each target to avoid contingency learning biases (for a discussion of this issue, see Schmidt, 2013). Experiment 1 employed a Stroop task, Experiment 2 employed an Eriksen flanker task, and Experiment 3 employed a Simon task. Of importance, the CSE was absent in all three tasks after employing the repetition deletion technique, which deletes all trials with feature repetitions from the analyses before calculating the CSE.

We conducted the NV regression in a trivially-different fashion than Notebaert and Verguts (2007). These authors conducted a regression on the correct RT data for each participant and then averaged the resulting estimated regression coefficients for each condition across participants (c.f., Lorch & Myers, 1990). In contrast, we conducted a linear mixed effects (LME) regression model (using the MIXED procedure in SPSS) on the correct RT data for each participant, which is essentially identical and typically preferred because it can provide greater flexibility (Van den Noortgate & Onghena, 2006). In this approach, participants are added as random factors and a single regression coefficient is calculated for each factor across all participants. The eight fixed factors, or regressors, that we included in the LME for each experiment are listed in Table 2.

It is important to note that, while not reported, we also performed the regression analysis identically to Notebaert & Verguts (2007). Critically, we observed no notable differences between the regression coefficients or statistical tests yielded by the NV regression approach and those yielded by the LME approach that we report in the present article, including the critical lack-of-fit test. In fact, the two approaches produced group-averaged regression coefficients that differed by only fractions of a millisecond. The condition estimates used for the lack-of-fit test were therefore also nearly indistinguishable.

Returning to how we conducted the LME regression, like Notebaert and Verguts (2007), we excluded (a) error trials and (b) correct trials that were preceded by error trials. Unlike Notebaert and Verguts, however, we included all eight regressors of the NV regression in a single step, rather than adding the previous congruency and CSE regressors in a second step. A one-step regression assigns variance to all eight regressors simultaneously, while a two-step regression assigns variance to the regressors in the first step before assigning variance to the regressors in the second step. Of importance, this methodological choice does not influence the outcome of the critical lack-of-fit test, because one- and two-step regressions provide the same end fit to a data set. The condition estimates for the lack-of-fit test were therefore identical with both approaches.

The regression coefficients and statistical tests from the LME models of Experiments 1, 2, and 3 are presented in Table 3. In all three experiments, there was a significant regression coefficient for the *congruency effect*, indicating faster responses for congruent relative to incongruent trials. There were also significant regression coefficients for *target-*

*target* and *distracter-distracter repetitions*, indicating faster responses for repetitions relative to alternations. Finally, there was a significant regression coefficient for the *feature repetition* regressor, indicating that performance averaged across complete repetitions and complete alternations was faster than performance for the remaining trials (i.e., partial repetitions).

Some effects were significant in only a subset of the experiments. In Experiment 1, there was a significant regression coefficient for *previous congruency*, indicating slower responses to trials following a congruent trial. In Experiment 2, there was a marginal regression coefficient for *target-distracter repetitions*, indicating marginally slower responses on repetition trials. In Experiment 3, there was a significant regression coefficient for *negative priming*, indicating slower responses for negative priming trials. Most relevant for present purposes, the *CSE* regressor was marginally significant in Experiment 1 and significant in Experiment 3, indicating a larger congruency effect after a congruent trial than after an incongruent trial. These latter findings contrast with the null CSEs reported with the repetition deletion procedure, which also yielded smaller numerical estimates of the CSE (i.e., with trims, the effect was only 1 ms in Experiment 1 and 6 ms in Experiment 3).[6] These contrasting findings are consistent with two possible interpretations: (1) the NV regression provided more statistical power than the repetition deletion procedure for detecting a CSE or (2) the NV regression was misspecified.

**(Table 3)**

## Lack-of-fit tests

To investigate whether the NV regression model was misspecified, we conducted a lack-of-fit test in each of the three experiments. To this end, we first used the regression equations described in the previous section to generate a model-estimated mean RT for each of the 15 relevant conditions in Table 2. The conditional mean RTs for participants, model-estimated mean RTs, and differences between the two are shown in Table 4, separately for each experiment. Positive and negative difference scores, respectively, indicate that model-estimated mean RT was greater or less than participant mean RT. As can be seen from the 15 difference scores, the model misestimates the conditional means by an average of 12 ms in Experiment 1, 3 ms in Experiment 2, and 7 ms in Experiment 3. Next, for each of the 15 conditions separately, we subtracted each participant's mean RT from the model-estimated mean RT. Finally, we conducted a repeated-measures one-way ANOVA on these 15 difference score variables. The ANOVA revealed large violations of sphericity in all three experiments. We therefore employed MANOVA to conduct the lack-of-fit test as it makes no assumptions about sphericity (see O'Brian & Kaiser, 1985).[7]

The MANOVA revealed that the lack-of-fit test was significant in Experiment 1, $F(6,9) = 34.846$, *Wilk's* $\Lambda = .041$, $p < .001$, marginal in Experiment 2, $F(6,9) = 2.982$, *Wilk's* $\Lambda = .359$, $p = .069$, and significant in Experiment 3, $F(6,9) = 6.613$, *Wilk's* $\Lambda = .201$, $p = .006$. Thus, in general, the degree to which the model did not fit the data exceeded the amount expected from random error. This result suggests the model was *misspecified*, meaning that one or more additional regressors would be needed to explain systematic un-modelled variance in the data. Thus, the significant CSE yielded by the NV regression approach in Experiment 3, which contrasts with the null CSE that were observed with the repetition deletion technique, was likely due to un-modelled feature repetition effects.

**(Table 4 about here)**

## Discussion

One goal of the present commentary was to investigate whether the NV regression approach to isolating a conflict adaptation effect is ever invalid. To this end, we applied a LME model to the data from Experiments 1 and 2 of Schmidt and De Houwer (2011) and the 25% congruency (i.e., contingency-free) condition of Mordkoff (2012). A lack-of-fit test demonstrated that the NV regression approach significantly misestimated the conditional means in both Experiment 1 (Schmidt & De Houwer, 2011, Experiment 1) and Experiment 3 (the 25% congruency condition of Mordkoff, 2012) and marginally misestimated the conditional means in Experiment 2 (Schmidt & De Houwer, 2011, Experiment 2). Further, the CSE estimated with the NV regression approach was significant in Experiment 3 and marginally significant in Experiment 1, in contrast to the non-significant CSEs previously reported in these experiments by researchers employing the repetition deletion technique. These findings suggest that the NV regression approach did not code for all possible effects of feature repetitions on RT. Thus, the CSE regressor had the opportunity to "steal" un-modelled, correlated RT variance stemming from interactions involving different types of feature repetitions, thereby increasing the probability that it would achieve significance.

One might wonder whether the NV regression approach simply provides greater statistical power for detecting a CSE than the repetition deletion technique. In other words, perhaps only the NV regression approach was powerful enough to detect small conflict adaptation effects that were truly present in our data sets. This possibility appears unlikely for two reasons. First, the estimates of CSE magnitude were numerically larger in Experiments 1 and 3 when using the NV regression approach relative to the repetition deletion technique. However, it is unclear why this should be the case from the perspective of statistical power. Indeed, complete alternation trials are generally the slowest of all trials (see Table 4). Thus, the conflict adaptation effect should scale up to a *larger* size with the repetition deletion technique (which uses only complete alternation trials) than with the NV regression approach (which makes use of all trials). Second, the lack-of-fit results indicate that the (nested) regression approach is less valid than the (non-nested) repetition deletion technique, in the sense that the CSE regressor has the opportunity to "steal" variance in RT from un-modelled repetition effects only with the former approach. This finding may explain why estimates of CSE magnitude were higher with the NV regression approach as compared to the repetition deletion technique. Given these considerations, it appears unlikely that the present results were driven by increased statistical power for detecting a CSE with the NV regression approach as compared to the repetition deletion technique.

## Implications for prior findings with the NV regression approach

By showing that the NV regression approach does not account for all effects of feature repetitions on RT in the present data set, our findings suggest that certain prior claims of CSEs independent of feature repetitions may need to be re-evaluated (e.g., Blais & Verguts, 2012; Braem, Verguts, Roggeman, & Notebaert, 2012; Eichele, Juvodden, Ullsperger, & Eichele, 2010; Steinhauser et al., 2012). For example, Braem and colleagues (2012) reported that the CSE was modulated by reward after controlling for feature repetitions with the NV regression approach and concluded that reward modulates conflict adaptation. Given the present findings, however, it is possible that reward modulated feature repetition effects that were not coded in the model. Consistent with this possibility, target repetition effects were significantly stronger with reward than without in their experiment. Additional studies could be conducted to investigate this alternative interpretation of Braem and colleagues' finding.

A second finding that may deserve further scrutiny is that the CSE is greater with small stimulus sets than with large stimulus sets (Blais & Verguts, 2012). To explain this result, Blais and Verguts presented a variant of the adaptation-by-binding account (see Verguts and Notebaert, 2009), in which conflict-modulated learning occurs most strongly for recently encountered stimuli. Based on this variant, Blais and Verguts argued that conflict-modulated learning, and hence the size of the CSE, should be larger with small stimulus sets than with large ones, because each feature occurs more frequently (and, hence, recently) with small stimulus sets. Notably, however, the CSE was significantly greater with small than with large stimulus sets when the authors employed the NV regression approach to control for immediate feature repetitions, but *not* when the authors removed trials with immediate feature repetitions from the analysis. Given our findings suggesting that the NV regression approach does not always “regress out” all possible immediate feature repetition effects, this discrepancy suggests an alternative interpretation of Blais and Verguts’ findings. Specifically, the CSE yielded by the NV regression approach might have increased as the set size became smaller simply because the number of immediate stimulus repetitions increased as the set size became smaller. Future studies could be aimed at testing this hypothesis.

## Implications for future studies with the NV regression approach

Future studies might potentially identify a differently-specified model that does not violate the lack-of-fit test. Although such a development would be encouraging, it is important to note that while a significant lack-of-fit test indicates that a model is incorrect, a non-significant lack-of-fit test does not indicate that a model is correct. This follows the logic of any null statistic: even when evidence for the alternative hypothesis is lacking, the alternative hypothesis may nevertheless be true (i.e., there may be a Type 2 error). Critically, detected or not, error in the structure of a regression model makes the regression coefficients difficult to interpret. In such cases, a regressor in the model (e.g., the CSE regressor) can “steal” variance produced by a correlated but un-modelled variable (e.g., interactions between different types of feature repetitions), resulting in a significant regression coefficient for that regressor in the absence of the theoretical process of interest (e.g., conflict adaptation). These considerations suggest that NV regression may not provide unequivocal evidence of CSEs independent of feature repetition confounds, even when a lack-of-fit test does not achieve significance.

Given the limitations of the NV regression approach, future researchers investigating the CSE might consider other approaches that (a) delete trials with feature repetitions “after the fact” (Kunde & Wühr, 2006; Mayr et al., 2003; Mordkoff, 2012; Schmidt & De Houwer, 2011) or (b) prevent feature repetitions from occurring in the original trial sequence without introducing contingency learning biases (Jiménez & Méndez, 2012; Mayr et al., 2003; Schmidt & Weissman, 2014; Weissman, Jiang, & Egner, in press). Since these approaches estimate the CSE solely from performance in complete alternation trials, there should be no concerns about the effects of un-modelled feature repetition effects on the CSE.

In sum, we have highlighted an important problem associated with "regressing out" the influence of a confounding variable in a nested design. As with more typical regression approaches, regressing out the influence of a confounding variable in a nested design will fail to the extent that the appropriate regressors do not completely capture variance associated with the confounding variable. However, this problem is magnified with nested designs because there are no regressors to code for various higher-order interactions involving a confounding variable. Thus, un-modelled variance due to such interactions may influence regression-derived estimates of the variable(s) of interest. Since it is never an experimental psychologist's goal to partially (rather than fully) control for a confounding variable, our findings indicate the need for caution when trying to regress out the influence of a confounding variable in a nested design.

## Broader implications

We now turn to the second main goal of our commentary: to illustrate that the problem we have identified with regressing out the influence of a confounding variable in a nested design exists in many areas of psychology and neuroscience. We now consider three examples of the "nested design problem" from the literatures on working memory updating, size estimation, and resting-state functional connectivity. As will become clear, the "nested design problem" complicates the interpretation of data in multiple domains.

First, consider a study from the literature on working memory updating. Kessler and Oberauer (2014) presented participants with four items followed by another four items in each trial and manipulated (1) the number of old items that were repositioned or changed to new items, (2) the number of new combinations of items, (3) the number of changed sequences of successive numbers, and (4) the number of times the list switched from an old to a new item when read from left-to-right. This was a nested design, because fully crossing these four factors was impossible (e.g., it is impossible to have a new sequence of numbers without introducing new or repositioned items). It is therefore possible that the effect of one factor (e.g., the number of new or repositioned items) was actually driven by un-modelled interactions between two other factors (e.g., the number of new combinations of items and the number of new sequences of items). Whether or not such an alternative explanation of these data is plausible remains uncertain. However, a lack-of-fit test conducted on such data would help to determine whether evidence of misspecification exists in the authors' best-fitting model of the data. Though not conclusive, a non-significant lack-of-fit test would be consistent with the authors' interpretation of the results.

Second, consider a study by Kirsch, Königstein, and Kunde (2014) who were interested in the roles of motor performance and task feedback on judgments of target size. Participants were asked to move a (disappearing) cursor toward a target circle, after which they were to estimate the target's size. The authors reported that size estimations were influenced by whether a participant "hit" or "missed" the target. This influence might have been due to participants' knowledge of their actual motor accuracy, as measured by the degree to which the final location of their movement deviated from the center of the circle. Alternatively, this influence might have been due to the feedback participants received about their accuracy.

To distinguish between these two potential explanations, the authors gave participants "hit" feedback in some trials wherein the target was barely missed and "miss" feedback in some trials wherein the target was barely hit. The authors then employed regression to determine whether accuracy feedback influenced the results independent of actual motor accuracy. The results of the regression supported this view by showing that size judgments were influenced by the feedback regressor after controlling for motor accuracy. However, this use of regression is potentially just as problematic as in our earlier example above because the experimental design was once again nested. For example, while trials in which the target was either just missed or just hit could receive either “hit” or “miss” feedback, clear misses were always given “miss” feedback and clear hits were always given “hit” feedback. For this reason, very accurate responses only contributed to the estimate for a hit and very inaccurate responses only contributed to the estimate for a miss. It is therefore possible that size estimates were only affected by motor performance, and that the feedback regressor merely capitalized on un-modelled variance from trials with very accurate and/or very inaccurate responses, which was not captured by the strictly linear motor accuracy regressor. Although this alternative interpretation of the data may not be correct, a lack-of-fit test could reveal whether evidence of model misspecification exists. Alternatively, deleting the very accurate and very inaccurate responses and restricting the analyses to moderately-inaccurate trials would eliminate the nesting problem, similar to the repetition deletion technique for the CSE.

Third, consider work from the resting-state functional connectivity literature. In this literature, functional MRI is employed to assess the degree to which the blood-oxygenated level-dependent (BOLD) signal is correlated between different brain regions across time while study participants lie still without performing a task. Numerous researchers have reported that resting-state functional connectivity varies across different subject populations (e.g., Van Dijk, Sabuncu, & Buckner, 2011). Recently, however, it has been shown that some of this variance can be explained by un-modelled, higher-order head motion artifacts (e.g., Lemieux et al., 2007; Satterthwaite et al., 2012), which are not completely “regressed out” by incorporating linear estimates of motion into regression analyses of resting-state fMRI data (e.g., Power et al., 2014; Satterthwaite et al., 2012). Linear motion regressors probably fail to capture all of the motion effect because they do not model all of the ways that motion and resting state activation interact. The regression model typically employed is thus a nested version of a more complex, correctly-specified model that would include more regressors. Determining the correctly-specified model, however, is no easy task. Thus, researchers in this field now employ approaches analogous to the repetition deletion technique, such as deleting time points at which motion artifacts occur. They also employ several other approaches that do not rely on regression to correct for head motion artifacts (Fair et al., 2013). This example further illustrates the potential problems associated with trying to "regress out" confounds and shows that these problems extend even beyond the experimental psychology literature.

Although the experiments discussed above illustrate that the "nested design problem" is a pervasive one, we do not mean to suggest that all of the conclusions drawn from these experiments are incorrect. For instance, it appears quite reasonable to conclude that participants are influenced by feedback when making size judgments (Kirsch et al., 2014). Our point is simply to show that there are clear misconceptions over the effectiveness of regression approaches to controlling for confounds, particularly when nested designs are employed in which higher-order interactions involving a confound cannot be modelled and may therefore continue to influence the variable(s) of interest. In such situations, including a regressor to code for a confound implies that an effect of interest has been isolated in a "confound-free" manner, even though this is unlikely to be the case. In short, while ruling out confounds is an important aim in experimental psychology, regression is not necessarily the ideal way to accomplish this goal, particularly when nested designs are employed. Further, while lack-of-fit tests can be employed to assess whether there is systematic un-modelled variance in a data set that likely emanates from un-modelled confounds, such tests are not a perfect solution to the "nested design problem." As mentioned before, while a significant lack-of-fit test gives clear evidence that the model is incorrectly specified, a non-significant lack-of-fit test is ambiguous.

Given the discussion above, one might conclude that regression should never be employed to control for confounds. However, this is not the case. In some situations, employing a regression approach to "regress out" the influence a confounding variable may not introduce a nesting problem because, unlike with the CSE, it may be possible to collect data from all of the cells in the factorial design. In other situations, a regression approach, even if imperfect (i.e., for the reasons discussed in the present paper), may be the approach that allows the highest level of control over confounds. For instance, if the feature repetition deletion technique could not be employed to assess the CSE, then the NV regression approach would be the best option available. For these reasons, regression will in many cases provide a very useful data analysis approach. We only aim to caution that the caveats of this approach should be carefully considered and that easier-to-interpret analysis techniques (e.g., the repetition deletion technique) should be sought out and preferred wherever possible.

## Conclusion

Notebaert and Verguts (2007) correctly concluded that "the explanatory value of a factor depends on the other factors included in the regression" (p. 1259). Further, they correctly acknowledged that the significance of the CSE regressor in their analysis may have been driven by feature repetition effects that were not coded in the regression model. Consistent with this possibility, the present findings suggest that the NV regression approach may not effectively isolate the CSE from feature repetition confounds. We therefore suggest that investigating the CSE in complete alternation trials is the best approach to controlling for feature repetition effects. A second goal of this report was to highlight the broader problem of attempting to "regress out" the influence of a confounding variable in a nested design. This practice is widespread in experimental psychology and cognitive neuroscience and can lead to false confidence that a variable of interest influences a dependent measure independent of (incompletely-modelled) confounding variables. When clearer dissociation procedures exist, as they do with the CSE, then regression should be employed with greater caution or completely avoided.

## References

Akçay, Ç., & Hazeltine, E. (2007). Conflict monitoring and feature overlap: Two sources of sequential modulations. *Psychonomic Bulletin & Review, 14,* 742–748.

Blais, C., & Verguts, T. (2012). Increasing set size breaks down sequential congruency: Evidence for an associative locus of cognitive control. *Acta Psychologica, 141,* 133–139.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

Braem, S., Verguts, T., Roggeman, C., & Notebaert, W. (2012). Reward modulates adaptations to conflict. *Cognition, 125,* 324–332.

Barsalou, L.W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T.K. Srull & R.S. Wyer (Eds.), *Advances in social cognition, Volume III: Content and process specificity in the effects of prior experiences* (pp. 61–88). Hillsdale, NJ: Lawrence Erlbaum Associates.

Eichele, H., Juvodden, H. T., Ullsperger, M., & Eichele, T. (2010). Mal-adaptation of event-related EEG responses preceding performance errors. *Frontiers in Human Neuroscience, 4,* Article 65.

Fair, D. A., Nigg, J. T., Iyer, S., Bathula, D., Mills, K. L., Dosenbach, N. U. F., et al. (2013). Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. *Frontiers in Systems Neuroscience,* 6, Article 80.

Faraway, J. J. (2004). *Linear models with R.* London: Chapman and Hall.

Freitas, A. L., Bahar, M., Yang, S., & Bahar, R. (2007). Contextual adjustments in cognitive control across tasks. *Psychological Science, 18,* 1040–1043.

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information:

Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121,* 480–506.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus–response episodes. *Visual Cognition, 5,* 183–216.

Hommel, B., Proctor, R. W., & Vu, K.-P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research, 68,* 1–17.

Jiménez, L., & Méndez, A. (2012). It is not what you expect: Dissociating conflict adaptation from expectancies in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 271–284.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303,* 1023–1026.

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 738–754.

Kirsch, W., Königstein, E., & Kunde, W. (2014). Action feedback affects the perception of action-related objects beyond actual action success. *Frontiers in Psychology, 5,* Article 17.

Kunde, W., & Wühr, P. (2006). Sequential modulations of correspondence effects across spatial dimensions and tasks. *Memory & Cognition, 34,* 356–367.

Lemieux, L., Salek-Haddadi, A., Lund, T. E., Laufs, H., Carmichael, D. (2007). Modelling large motion events in fMRI studies of patients with epilepsy. *Magnetic Resonance Imaging, 25,* 894–901.

Lorch, R. F., & Myers, J. L. (1990). Regression-analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition, 16,* 149–157.

Mayr, U., & Awh, E. (2009). The elusive link between conflict and conflict adaptation. *Psychological Research, 73,* 794–802.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience, 6,* 450–452.

Mordkoff, J. T. (2012). Observation: Three reasons to avoid having half of the trials be congruent in a four-alternative forced-choice experiment on sequential modulation. *Psychonomic Bulletin & Review, 19,* 750-757.

Nieuwenhuis, S., Stins, J. F., Posthuma, D., Polderman, T. J. C., Boomsma, D. I., & de Geus, E. J. (2006). Accounting for sequential trial effects in the flanker task: Conflict adaptation or associative priming? *Memory & Cognition, 34,* 1260–1272.

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

Notebaert, W. & Verguts, T., (2007). Dissociating conflict adaptation from feature integration: A multiple regression approach. *Journal of Experimental Psychology: Human Perception and Performance, 33, 12*56–1260.

O'Brian, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97,* 316–333.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage, 84,* 320–341.

Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., Gur, R. C., & Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage, 60,* 623–632.

Schmidt, J. R. (2013). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review, 20,* 615–630.

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

Schmidt, J. R., & Weissman, D. H. (2014). Congruency sequence effects without feature integration or contingency learning confounds. *PLOS ONE, 9,* e102337.

Steinhauser, M., Eichele, H., Juvodden, H. T., Huster, R. J., Ullsperger, M., & Eichele, T. (2012). Error-preceding brain activity reflects (mal-)adaptive adjustments of cognitive control: A modeling study. *Frontiers in Human Neuroscience, 6,* Article 97.

Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: It's not just priming. *Cognitive, Affective, & Behavioral Neuroscience, 5,* 467–472.

Van den Noortgate, W., & Onghena, P. (2006). Analysing repeated measures data in cognitive research: A comment on regression coefficient analyses. *European Journal of Cognitive Psychology, 18,* 937–952.

Van Dijk, K. R. A., Sabuncu, M. R., & Buckner, R. L. (2011). The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage, 59,* 431–438.

Verbruggen, F., Notebaert, W., Liefooghe, B., & Vandierendonck, A. (2006). Stimulus- and response-conflict-induced cognitive control in the flanker task. *Psychonomic Bulletin & Review, 13,* 328–333.

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

Weissman, D. H., Jiang, J., & Egner, T. (in press). Determinants of congruency sequence effects without learning and memory confounds. *Journal of Experimental Psychology:*

*Human Perception and Performance.*

Wendt, M., Kluwe, R. H., & Peters, A. (2006). Sequential modulations of interference evoked by processing task-irrelevant stimulus features. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 644–667.

## Footnotes

[1] Actually, in the datasets used by Notebaert and Verguts (2007) there were *less* than 15 cells to estimate with their 8 regressors. If a task has less than four response options, then the number of possible conditions is reduced (for further explanation, see Schmidt & De Houwer, 2011).

[2] An anonymous reviewer suggested that this claim is false because, in each participant, there are multiple observations (i.e., trials) for each of the 15 trial types. However, the number of observations per trial type is irrelevant. What is relevant is that there exist only 15 points on the regression line for the model to estimate. Therefore, only 14 regressors (plus the intercept) are needed to perfectly estimate these 15 points. This fact can be demonstrated by adding 6 random (but non-redundant) regressors to the eight that are already included in the NV regression model. As we confirmed in further (unreported) analyses of the data from Experiment 1, this procedure resulted in all of the variance between the 15 conditions being explained. Therefore, the value of the F-statistic corresponding to the lack-of-fit test was exactly zero.

[3] Related to the first footnote, these three data sets are most desirable, because all 15 of these combinations are possible in these four-choice tasks, allowing for maximal degrees of freedom in the present analysis. These experiments are also contingency-unbiased, preventing other potential complications with a contingency learning confound. In contrast, the data from Notebaert and Verguts' (2007) three-choice task, for which we did not compute a lack-of-fit statistic, is missing one of the 15 aforementioned combinations (condition 9 in Tables 2 & 4). This is not a trivial issue for two reasons: (1) degrees of freedom are already limited, and (2) the combination lost is a particularly interesting one, representing complete alternations for one of the four cells of the CSE design. A final interesting point is that all three of these data sets provided no significant evidence for a

CSE independent of feature repetition biases.

[4] An anonymous reviewer suggested that this lack-of-fit test is biased toward producing a significant *F*-value. The reason is that the subset of the 64 combinations of previous congruency, current congruency, and the four repetition types that can be observed in a real experiment is not random. Contrary to the reviewer's suggestion, however, we would argue that if the model is correctly-specified, then it should be able to correctly reproduce the real-world parameters that were used to create the data set, regardless of whether the missing cells are randomly or non-randomly selected. To determine which view is correct, we conducted a simulation, suggested by the reviewer, which involved computing the cell values for the 64 combinations of our six factors described above using the NV regression parameters from Experiment 1. We added some random normal error for each observation (*SE* = 10) and included 100 simulated participants. When applied to this simulated data, the NV regression did an exceptional job of replicating the regression parameters that were used to generate the data, whether it was applied to all 64 combinations or to just the 15 combinations that are observable in actual participants. Further, when a lack-of-fit test was applied to the difference scores between the simulated and model-estimated RTs for the 15 observable conditions, the test was not significant, $F(6,86) = .992$, *Wilk's Λ* = .935, $p = .474$. These values are exactly what one would expect with no model misspecification, and thus indicate that the lack-of-fit test we employ is not biased. In fact, the only way the reviewer's view could be correct is if non-randomly selecting a subset of trial types caused the NV regression to inaccurately compute the regression coefficients for the eight factors in the model. If this were the situation, which our simulation shows is not the case, then it would be just as detrimental to the validity of the NV regression as the limitation we have identified.

[5] Each of the eight regressors will increase the ability of the model to capitalize on random

error. Thus, the degrees of freedom for the numerator should equal 15 (conditions) – 8 (regressors) – 1 (intercept) = 6. Without this correction, Type 2 errors will be artificially inflated. Appropriately, this means that the degrees of freedom for the numerator will be 0 if there are 14 or more regressors. Since all conditions would be perfectly estimated in this scenario, there are no degrees of freedom (see Footnote 2).

[6] In the ideal situation, one would determine whether estimates of the CSE derived from the NV regression and repetition deletion approaches differ from one another. However, the sample size required to make this comparison with high statistical power is often prohibitive. For instance, if the true CSE magnitude in Experiment 1 was zero, then detecting a difference between 0 ms (estimated with the repetition deletion technique) and 7 ms (estimated with the NV regression approach) would require a sample size of around 300 participants to achieve a relatively high power of .8. This is due, in part, to the fact that the NV regression approach produces a significant CSE only because a relatively small parameter estimate for the CSE is associated with a very (probably artificially) small estimate of the error variance. In contrast, the repetition deletion technique estimates the CSE to be around 0 ms and is associated with a relatively high estimate of the error variance. Thus, detecting a significant difference between these only slightly different estimates of the CSE is difficult, in part, because one is associated with a much higher estimate of the error variance than the other.

[7] Given a major violation of sphericity and an acceptable sample size, MANOVA is generally a more powerful approach for dealing with violations of sphericity.

***Table 1.*** The 15 experimentally-observable combinations of current congruency, previous congruency, and the four types of feature repetitions.

| distracter-distracter: | Repetition (R) | | | | | | | | Alternation (A) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| target-target: | R | | | | A | | | | R | | | | A | | | |
| distracter-target: | R | | A | | R | | A | | R | | A | | R | | A | |
| target-distracter: | R | A | R | A | R | A | R | A | R | A | R | A | R | A | R | A |
| Congruent-Congruent | ✓ | | | | | | | | | | | | | | | ✓ |
| Congruent-Incongruent | | | | | | | ✓ | | | ✓ | | | | | | ✓ |
| Incongruent-Congruent | | | | | | ✓ | | | | | ✓ | | | | | ✓ |
| Incongruent-Incongruent | | | | ✓ | | | | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ |

*The four shaded cells are those used in the repetition deletion analysis.

***Table 2.*** Different linear combinations of eight regressors were employed to predict mean RT in 15 unique trial types.

| Trial Type | Congruency | Target-Target | Distracter-Distracter | Feature Integration | Negative Priming | Target-Distracter | Previous Congruency | Conflict Adaptation |
|---|---|---|---|---|---|---|---|---|
| *Congruent–Congruent* | | | | | | | | |
| (1) BLUE$_{\text{blue}}$ → RED$_{\text{red}}$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| (2) BLUE$_{\text{blue}}$ → BLUE$_{\text{blue}}$ | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| *Congruent–Incongruent* | | | | | | | | |
| (3) BLUE$_{\text{blue}}$ → RED$_{\text{green}}$ | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| (4) BLUE$_{\text{blue}}$ → BLUE$_{\text{red}}$ | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| (5) BLUE$_{\text{blue}}$ → RED$_{\text{blue}}$ | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Incongruent–Congruent* | | | | | | | | |
| (6) RED$_{\text{blue}}$ → GREEN$_{\text{green}}$ | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| (7) RED$_{\text{blue}}$ → RED$_{\text{red}}$ | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| (8) RED$_{\text{blue}}$ → BLUE$_{\text{blue}}$ | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| *Incongruent–Incongruent* | | | | | | | | |
| (9) RED$_{\text{blue}}$ → GREEN$_{\text{yellow}}$ | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| (10) RED$_{\text{blue}}$ → RED$_{\text{green}}$ | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| (11) RED$_{\text{blue}}$ → GREEN$_{\text{blue}}$ | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| (12) RED$_{\text{blue}}$ → RED$_{\text{blue}}$ | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| (13) RED$_{\text{blue}}$ → GREEN$_{\text{red}}$ | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| (14) RED$_{\text{blue}}$ → BLUE$_{\text{green}}$ | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| (15) RED$_{\text{blue}}$ → BLUE$_{\text{red}}$ | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |

*Although example Stroop trial types from Experiment 1 are provided in the Table, the same analysis was employed with analogous flanker and Simon trial types, respectively, in Experiments 2 and 3.

***Table 3.*** Regression model results.

| Regressor | Estimate | *t* | *p* |
|---|---|---|---|
| Experiment 1 | | | |
| *Intercept* | 506 | 27.149 | <.001 |
| *Current congruency* | -57 | -14.393 | <.001 |
| *Target-target* | 209 | 53.651 | <.001 |
| *Distracter-distracter* | 31 | 7.960 | <.001 |
| *Feature integration* | 15 | 3.648 | <.001 |
| *Negative priming* | 1 | .141 | .888 |
| *Target-distracter* | -2 | -.453 | .651 |
| *Previous congruency* | 9 | 2.327 | .020 |
| *CSE* | 7 | 1.700 | .089 |
| | | | |
| Experiment 2 | | | |
| *Intercept* | 543 | 36.448 | <.001 |
| *Current congruency* | -36 | -13.126 | <.001 |
| *Target-target* | 101 | 37.301 | <.001 |
| *Distracter-distracter* | 14 | 5.355 | <.001 |
| *Feature integration* | 17 | 5.949 | <.001 |
| *Negative priming* | -3 | -.973 | .331 |
| *Target-distracter* | -4 | -1.885 | .059 |
| *Previous congruency* | -1 | -.504 | .614 |
| *CSE* | 2 | .692 | .489 |
| | | | |
| Experiment 3 | | | |
| *Intercept* | 515 | 30.990 | <.001 |
| *Current congruency* | -57 | -14.480 | <.001 |
| *Target-target* | 100 | 23.166 | <.001 |
| *Distracter-distracter* | 18 | 4.090 | <.001 |
| *Feature integration* | 18 | 4.053 | <.001 |
| *Negative priming* | -13 | -2.606 | .009 |
| *Target-distracter* | -2 | -.617 | .537 |
| *Previous congruency* | -4 | -.995 | .320 |
| *CSE* | 9 | 2.163 | .031 |

***Table 4.*** Actual and model-estimated mean RTs in milliseconds (data from Schmidt & De Houwer, 2011; Mordkoff, 2012).

| Trial | Experiment 1 | | | Experiment 2 | | | Experiment 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Type | Obs | Est | Diff | Obs | Est | Diff | Obs | Est | Diff |
| *cC* | | | | | | | | | |
| (1) | 687 | 698 | 11 | 618 | 613 | -5 | 554 | 557 | 4 |
| (2) | 486 | 459 | -27 | 492 | 503 | 10 | 462 | 441 | -20 |
| *cI* | | | | | | | | | |
| (3) | 776 | 761 | -15 | 655 | 651 | -4 | 630 | 623 | -7 |
| (4) | 740 | 746 | 6 | 652 | 658 | 5 | 620 | 625 | 6 |
| (5) | 554 | 567 | 13 | 565 | 567 | 2 | 536 | 540 | 4 |
| *iC* | | | | | | | | | |
| (6) | 684 | 695 | 12 | 615 | 617 | 1 | 560 | 570 | 10 |
| (7) | 673 | 679 | 7 | 617 | 619 | 2 | 586 | 570 | -15 |
| (8) | 529 | 503 | -26 | 543 | 537 | -6 | 494 | 490 | -4 |
| *iI* | | | | | | | | | |
| (9) | 765 | 745 | -20 | 645 | 650 | 6 | 630 | 618 | -12 |
| (10) | 737 | 729 | -8 | 657 | 653 | -4 | 615 | 618 | 3 |
| (11) | 543 | 551 | 8 | 566 | 566 | 0 | 531 | 536 | 5 |
| (12) | 494 | 505 | 11 | 538 | 535 | -2 | 493 | 500 | 7 |
| (13) | 750 | 744 | -6 | 654 | 654 | 0 | 634 | 631 | -3 |
| (14) | 736 | 747 | 11 | 657 | 655 | -2 | 618 | 620 | 2 |
| (15) | 739 | 746 | 7 | 656 | 658 | 2 | 629 | 633 | 4 |

*Obs = observed (participant); Est = estimated (model); Diff = difference.

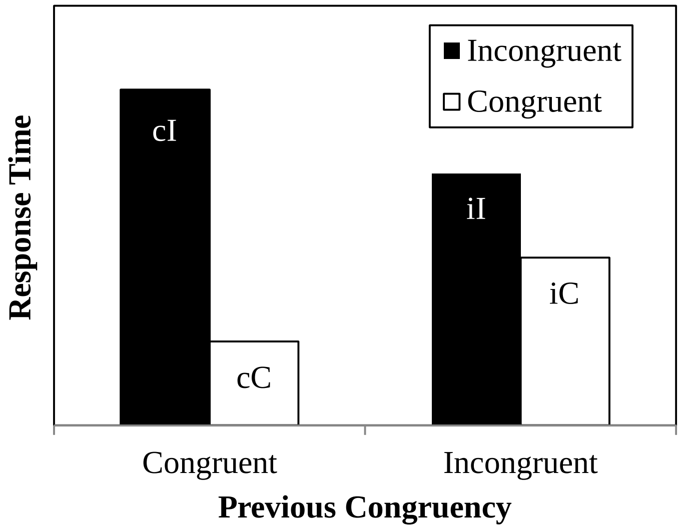

***Figure 1.*** Example congruency sequence effect.