

Running head: A REPETITION-SUPPRESSION ACCOUNT

A Repetition-Suppression Account of Between-Trial Effects in the Stroop Task

Ion Juvina

Carnegie Mellon University

Niels A. Taatgen

Carnegie Mellon University and University of Groningen

Author Note

This work was funded by the ONR grant no. N00014-06-1-0055. The authors thank Daniel Dickison for adjusting the first mechanism in ACT-R.

Correspondence concerning this article should be addressed to: Ion Juvina, Department of Psychology, Baker Hall, 336A, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, Tel. +1-412-268-2837, Email: [ijuvina@cmu.edu](mailto:ijuvina@cmu.edu).

### Abstract

Some of the influential theories and models of cognitive control are challenged when attempting to explain empirical data that replicate previously reported findings. These findings refer to four between-trial effects in the Stroop task, previously described in the literature on Stroop effects, negative priming, and inhibition of return. Existing theoretical accounts fail to explain all these four effects in an integrated way. A repetition-suppression mechanism is proposed in order to account for these data in an integrated way. Two computational cognitive models implementing this repetition-suppression mechanism are proposed. The values and limitations of each of these models, as well as their theoretical implications, are discussed.

Keywords: Stroop, Inhibitory control, Sequence effects, Cognitive modeling, Negative priming

PsycINFO classification code: 2340 Cognitive Processes

## A Repetition-Suppression Account of Between-Trial Effects in the Stroop Task

Recent advances in Cognitive Neuroscience have made it possible to better understand the nature and mechanisms of cognitive control. One of the typical functions of cognitive control is to compose task-relevant temporal sequences of actions. This paper attempts to prove with empirical and computational arguments that repetition-suppression is sometimes employed in temporal sequencing of actions, particularly when repetition of stimuli complicates the process of response selection.

It is currently under debate whether suppression (also referred to as *cognitive inhibition*) is one of the mechanisms of cognitive control. Some authors assert that cognitive inhibition is essential for cognitive control (Aron, 2007; Houghton & Tipper, 1996), others say that it is unnecessary (Egner & Hirsch, 2005; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003). In most (if not all) of the cases, the behavioral effects attributable to cognitive control can be accounted for equally well with or without a suppression mechanism.

In order to disentangle concurrent accounts we will impose several methodological constraints: (1) a viable theoretical account should be able to simultaneously explain a large range of effects, (2) it should be expressed in computational terms, and (3) it should be biologically plausible.

The classical Stroop task will be used because it is a landmark task for studying cognitive control (Miyake et al., 2000); an extensive body of literature has accumulated over many years to endorse the robustness of the Stroop task's behavioral effects as well as to aid with understanding the cognitive mechanisms responsible for these effects

(MacLeod, 1991). However, the theoretical arguments presented in this paper are not limited to the Stroop task; they have some bearing upon a range of tasks, particularly tasks involving processing temporal sequences of stimuli and selecting amongst potentially conflicting responses (what is sometimes referred to as “the continuous performance paradigm”).

The following section presents replications of known effects in the Stroop task and estimations of their relative magnitudes and time course. Section 2 shows how this empirical data challenge established theories and models of cognitive control and presents an alternative account. Section 3 presents two computational models that implement this alternative account. Section 4 discusses the contributions of each of these models to a coherent theory of cognitive control.

## 1. Empirical Studies

The empirical research presented here is part of an ongoing project aimed at contributing to a coherent theory of cognitive control. In one of the most comprehensive reviews of research on the Stroop task, MacLeod (1991) described a series of between-trial effects and proposed a suppression mechanism to account for all of them:

“When the irrelevant word on trial  $n-1$  is the name of the target ink color on trial  $n$ , interference with color naming will be enhanced temporarily; when the ink color on trial  $n-1$  matches the word on trial  $n$ , there will be some facilitation of color naming on trial  $n$ . If the word on trial  $n-1$  is repeated on trial  $n$ , then the word is already suppressed and will cause less interference in naming a different ink color on trial  $n$ . An interesting study would be to mix these two types of repetition effects in the same experiment, directly comparing their size.”  
“My own bias [...] is to invoke a suppression idea so that the facilitation and interference effects as a result of item sequence have a common grounding”  
(MacLeod, 1991, pp. 178)

It was our intention to conduct such an “interesting study” in which to replicate all of these between-trial effects and investigate whether a single integrated account can explain all of them as suggested by MacLeod in 1991. However, MacLeod has recently advocated against cognitive inhibition as an explanatory mechanism for attention and memory phenomena including negative priming and inhibition-of-return (MacLeod, 2007a; MacLeod, 2007b; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003). Thus, the research question we address here is whether this integrated account should be based on suppression (cognitive inhibition) or not. A subsequent goal will be to use neurally plausible computational modeling in order to understand how a cognitive control mechanism might operate in the human brain.

There were two empirical studies. The first study replicated all three between-trial effects described in the quotation above (for brevity, they will be referred to as Word-Color, Color-Word, and Word-Word, respectively), but it also revealed a significant Color-Color effect. This additional finding was also a replication of a known effect, although previously reported in different contexts (Law, Pratt, & Abrams, 1995; MacDonald & Joordens, 2000). It motivated a second study to replicate and increase the statistical power of the first study. Since the two studies yielded very similar results (see section 1.2.3.), they will be presented together.

## *1.1. Method*

### *1.1.1. Participants*

Fifty-three self-selected participants from Carnegie Mellon University's community were used in the first study. Participant age ranged from 18 to 59 years with an average of 24. There were 16 women and 37 men. In the second study, there were 39 participants who were self-selected from the same pool, with age ranging from 18 to 54 and averaging 25. There were 22 women and 18 men. All participants received a fixed amount of monetary compensation for their participation.

### *1.1.2. Design*

The standard Stroop task was adapted for screen-based administration and manual response. Stimuli were presented on a computer screen one at a time. There were three within-subject conditions: incongruent, congruent, and neutral. The three trial types corresponding to the three conditions were randomly mixed (non-blocked). Trial order was randomized for each participant. In the first study, every subject received 150 trials, 50 trials for each condition. In the second study, there were 240 trials per participant, 80 trials for each condition. The set size of neutral stimuli was decreased from 53 to 10. These changes were made in order to ensure that each participant encountered at least 30 repetitions for each of the between-trial effects. The location of stimuli on the screen was kept constant.

### *1.1.3. Materials*

A computer-based variant of the Stroop task was designed for the purpose of this research. Stimuli were color names (red, blue, yellow and green) and neutral words colored with one of the four colors denoted by the mentioned color names. The neutral

words were common English words unrelated semantically or phonologically to any of the color names. Stimuli were presented one at a time in the center of the screen and remained on the screen until the participant responded. Two response options were also displayed flanking the stimulus on its left and right sides. Response options were non-colored (i.e., in black) color names. One response option contained the correct answer and the other one an incorrect answer.

#### *1.1.4. Procedure*

Instead of verbally naming the color of the stimulus as in the classical Stroop task, participants were instructed to select as fast as possible the response option that matched the color of the stimulus from the two options presented on the left and right sides of the stimulus by pressing a key for each option. Previous research has shown that the Stroop task with manual response was sufficiently similar to the original Stroop task with vocal response (Nielsen, 1975). The session started with a short computer-guided tutorial that emphasized the correct response. During the task no feedback was provided.

#### *1.2. Results*

The data of two participants were excluded from analysis. One of these participants had very high reaction times, above 2000 ms on average, and this criterion had previously been used to exclude data from analysis (Miyake et al., 2000). The other participant that was excluded seemed to have misunderstood the task instructions. Given that he or she had an accuracy of zero (minimum) for the incongruent condition and one (maximum) for the congruent condition, we inferred that he or she reacted as if

responding to the word dimension instead of the color dimension of the stimulus. No other manipulation of the data was done.

Reaction time (RT) was used as a dependent measure. Sometimes when RT is used as a dependent measure it is log-transformed in order to correct for its skewed distribution. In our case, the results with and without the log-transformation of RT were similar. We decided to use the original (non-transformed) variable so that the magnitudes of all effects are always expressed in milliseconds. It is also a common practice to include only data from correct trials. In our data, RT was not significantly different when we excluded incorrect trials. Thus, we chose to present here the analyses performed on the whole dataset.

#### *1.2.1. Within-trial effects*

Accuracy data are consistent with previous studies, showing less than 2% errors for the congruent and neutral conditions and less than 10% errors for the incongruent condition (Table 1). Significant interference and facilitation effects were found (Table 1). Since within-trial effects were very consistent with those found in previous studies they will not be treated in more detail here.

(Table 1 about here)

#### *1.2.2. Between-trial effects*

These are effects related to a particular sequence of trials. They will be described one by one beginning with those mentioned by MacLeod (1991). Because we intend to

compare various accounts, we will not label these effects with terms that might suggest particular explanatory mechanisms (negative priming, inhibition-of-return etc.), as recommended by MacLeod (2003). Table 2 presents examples of all these effects.

(Table 2 about here)

Repetition priming data, cases in which the exact stimulus was repeated (for example, the word RED in red ink repeated as such), were excluded from the analyses of between-trial effects for the following reasons: (1) reactions to repeated stimuli were much faster than the other reactions and, arguably, a different type of response selection was involved, that is, a heuristic of the kind “if no stimulus change, then repeat the previous response” (Klein, 2004); (2) it would be impossible to establish where such a case belongs (for example, the congruent RED repetition mentioned above could be classified as any of the four between-trial effects); (3) it would artificially inflate or deflate the magnitudes of the between-trial effects depending upon whether these cases were included in the priming trials or in the baseline (Christie & Klein, 2001). The number of such cases was very small, that is, 4% of the total number of trials.

Frequencies for the between-trial effects were as follows: Word-Color 15%, Color-Word 14%, Word-Word 9%, and Color-Color 21% of the total number of trials. The absolute magnitudes of these effects will be presented as averaged differences between reaction times on priming versus non-priming trials (Table 3).

(Table 3 about here)

*1.2.2.1. The Word-Color effect (negative priming).* When the word on trial  $n-1$  names the color on trial  $n$ , reaction time increases. As shown in Table 3, the highest increase occurs in the incongruent condition (102ms), followed by the congruent condition (93ms) and neutral condition (40ms). This effect has been replicated many times, is very robust and fairly general (see Tipper, 2001, for a review).

*1.2.2.2. The Color-Word effect.* When the color on trial  $n-1$  is the same as the denotation of the word on trial  $n$ , reaction time decreases. This effect cannot occur in the neutral condition, because the neutral word does not denote any color. As shown in Table 3, the decrease in reaction time occurs solely in the incongruent condition (-74ms); in the congruent condition there is a very small (practically zero) increase in reaction time (8ms). This effect was first reported by Effler (1977) and replicated several times (Neill, 1978; Lowe, 1979; see also MacLeod, 1991, for a review).

*1.2.2.3. The Word-Word effect.* When the word on trial  $n-1$  is the same as the word on trial  $n$ , regardless of the colors of these words, reaction time decreases for the incongruent and neutral trials (-68ms and -51ms, respectively) and increases for the congruent trials (93ms) (Table 3). This effect is reversed in the congruent condition, that is, word repetition reduces both Stroop interference and facilitation. This result is a replication of the findings of Thomas (1977) and Effler (1980) reviewed by MacLeod (1991).

*1.2.2.4. The Color-Color effect.* When the color on trial  $n-1$  is the same as the color on trial  $n$ , regardless of the words of these stimuli, reaction time increases. As shown in Table 3, the highest increase occurs in the incongruent condition (79 ms),

followed by the neutral condition (23 ms); the increase is almost zero in the congruent condition (8 ms). This effect has not been reported in the context of the Stroop task (to our knowledge) but it was reported in the literatures on negative priming (MacDonald & Joordens, 2000) and inhibition-of-return (Law, Pratt, & Abrams, 1995). It is important to mention here that many authors report a decrease in reaction time for Target-Target repetitions (Lowe, 1979; Tipper, 1985). We obtained this decrease only in the case of repetition priming trials, that is, trials in which both stimulus features are repeated. As mentioned above, the repetition priming trials were excluded from this analysis.

### *1.2.3. Time Course and Relative Magnitudes of Between-Trial Effects*

In order to check for statistical significance these data were submitted to a Linear Mixed Effects (LME) model. This type of analysis was chosen because it allows controlling for individual differences and accurately determining the magnitudes of small effects in hierarchically nested data. Thus, the data points corresponding to the within- and between-trial effects are not independent across subjects; they are nested within subjects. It is known that priming effects are rather small in magnitude, often around 20 ms (MacLeod & Bors, 2002). Ignoring the inherent nesting characteristic of the data would diminish or eliminate some of the small-size effects.

A first analysis was intended to check if there are significant differences between the two studies. A LME model with reaction time (RT) as the dependent variable and *study*, *condition* and the four *between-trial effects* as independent variables was run. Results showed a main effect of *study* ( $t=2.038$ ,  $p=0.045$ ) and an interaction between *study* and *condition* ( $t=2.41$ ,  $p=0.016$ ). Reaction time was smaller overall in the second

study as compared with the first study, and particularly in the congruent condition. Differences in average reaction time between different studies have also been reported by MacLeod and Bors (2002). There were no significant interactions between *study* and any of the four *between-trial effects*. The Pearson correlation between the two studies, calculated on both within- and between-trial effects, was  $r=0.99$ . For these reasons the data of the two studies were combined for analysis. A similar procedure of combining two studies, the second being a replication of the first, was employed by MacLeod and Bors (2002).

Additional variables were included in the model to study the time course of between trial effects. Thus, for each effect, not only the *1-back* version was included but also the *2-* and *3-back* versions. For example, the *Word-Color-2back* represents the case where the word on trial  $n-2$  names the color on trial  $n$ . The dependent variable is *Reaction Time*. Independent variables are *Condition* (Incongruent, Congruent, and Neutral), and the four between-trial effects together with their *2-back* and *3-back* correspondents. The interaction between *Condition* and the *Word-Word* effect was also included in the model as an independent variable based on the observation that the *Word-Word* effect is reversed in the Congruent condition (see section 1.2.2.3. and Table 3).

The results of the initial LME model are presented in Table 4. Notice that all of the four effects are significant ( $\alpha = 0.05$ ). Some of the *2-back* effects and the interaction between *Condition* and the *Word-Word* effect are also significant.

(Table 4 about here)

The coefficients of this LME model can be used as estimates of the magnitudes of the between-trial effects. For example, for the *Word-Color* effect, the value of the LME coefficient indicates an increase in reaction time for the priming trials as compared to the non-priming trials of 50.7 ms, if all of the other variables in the model are kept constant. The coefficients for all the between-trial effects are plotted in Figure 1.

(Figure 1 about here)

A stepwise procedure was employed in order to find the best LME model. This procedure starts with the largest model (Table 4), computes the Akaike Information Criterion (AIC) for each factor, and drops one factor at a time until no further factors can be dropped without a significant decrease in AIC. The best model found for our data is summarized in Table 5.

(Table 5 about here)

The best model has dropped all of the *3-back* versions of the between-trial effects but has retained the *2-back* version of the *Color-Color* effect even though it was only marginally significant. It is noteworthy to observe that the estimates of the between-trial effects are very robust, that is, they do not considerably change their value from the initial to the last model.

Since most of the *2-back* versions of the between-trial effects were maintained in the best model, the time course of the between-trial effects can be estimated as twice the

average duration of a trial (1069 ms), that is 2138 ms. Other studies have reported a longer time course of between-trial effects (Erickson & Reder, 1998; Tipper, Weaver, Cameron, Brehaut, & Bastedo, 1991).

A comprehensive exploratory analysis was also performed, starting with a LME model that included all possible variables and their interactions, and then progressing toward the best model in a stepwise manner as described above. The best model obtained this way was not significantly different than the one presented above.

With regard to the magnitudes of the between-trial effects, the estimates used for model simulations presented in section 3 will be those presented in Figure 1 and taken from the initial LME model (Table 4). This was done in order to constrain the models to simulate the original data in its entirety, that is, not only the statistically significant results but also the results that indicate statistically nil effects. In addition, the estimates presented in Figure 1 show a pattern of the data that might be indicative for the underlying mechanism of the between-trial effects. Thus, Figure 1 suggests that between-trial effects are maximal at 1-back and progressively decrease in magnitude at 2- and 3-back. At 3-back their magnitude is statistically zero, but it seems that this is a result of a gradual decay.

## 2. Theoretical Accounts

Besides the well-known within-trial Stroop effects (increased and decreased RT in the incongruent and congruent conditions, respectively, as compared to the neutral condition), four significant between-trial effects have been found. Although not as well

known as the within-trial effects, these between-trial effects have been documented (Law, Pratt, & Abrams, 1995; MacDonald & Joordens, 2000; MacLeod, 1991). We have only replicated them and estimated their relative magnitudes and durations while controlling for the within-trial effects and individual differences. Although these findings have been known for a while, to our knowledge, there is no integrated account for all four between-trial effects. They are described by different authors and interpreted in isolation. For example, MacDonald & Joordens (2000) explained the Color-Color effect (they call it “negative priming in attended repetition trials”) by means of the selection-feature mismatch account, without constraining their account to simultaneously explain the Color-Word effect. Our research is the first attempt that we know of to analyze these four effects together in a study and explain them with a single account. Lowe (1979, 1985) made an attempt to study all sequence effects in the Stroop task in a single study but subsequently retained only two of them (negative priming and repetition priming) for which he provided an integrated account. An integrated suppression-based account has been suggested in a review by MacLeod (1991), but recently the same author has expressed strong criticism for any suppression-based account of attention and memory phenomena (MacLeod, 2007a; MacLeod, 2007b; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003).

A first attempt to interpret these between-trial effects directs us toward a repetition-suppression account: representations pertaining to just-completed trials are suppressed in order to prevent them from interfering with future trials. However, we will defer for now to advance such an account. As recommended by MacLeod (2003), we will first analyze the available inhibition-free accounts.

The episodic retrieval account (Neill, 1997) holds that the to-be-named feature of the current stimulus triggers an automatic retrieval (Logan, 1990) of the most recent episode, in which the concept corresponding to that feature has been used, and the associated reaction. For example, assuming the word “red” re-occurred as the color *red*, it would trigger the retrieval of an episode composed of the concept “red” and the reaction “do-not-respond”. Since the reaction derived from the retrieved episode is not adequate for the current stimulus, an additional retrieval is required to generate the proper reaction, which explains the time delay. This account predicts longer reaction times when the previous word feature re-occurs as the current color feature, that is, the Word-Color effect found in our data, also known as the negative priming effect. In the case of Color-Word repetition, this account would not predict a decrease in reaction time, as observed in our data; the color feature of the preceding trial has the reaction “respond” associated with it; when it comes back as the word feature of the current stimulus, it should increase interference because the irrelevant feature has now a “respond” reaction associated with it. In the case of the previous color feature re-occurring as the current color feature (the Color-Color case), this account would predict no increase in reaction times; the most recent episode involving the current color contains exactly the reaction needed for the current stimulus; thus, there would be no reason for an increase in reaction time as observed in our data. In the case of Word-Word repetition, this account would probably predict a decrease in reaction time (as observed in our data) because the previous word feature has a “do-not-respond” associated with it, which could make it easier to reject the same irrelevant word in the current trial. Thus, the episodic retrieval account accurately

predicts two of the four between-trial effects (Word-Color and Word-Word) but makes wrong predictions for the other two effects (Color-Word and Color-Color).

Another inhibition-free account suggested by MacLeod, Dodd, Sheard, Wilson, and Bibi (2003) is the feature mismatch account (Lowe, 1979; Park & Kanwisher, 1994). This account posits that when the repetition is accompanied by a feature mismatch, additional time is taken to resolve this conflict. For example, when the “red” word precedes the *red* color, redness is repeated but it occurs in conflicting features (word vs. color). This account predicts an increase in reaction time for the Word-Color effect but also for the Color-Word effect because there is a feature mismatch in this case as well. In the case of Color-Color repetitions, this account would not predict an increase in reaction time because the feature of the repeated entity does not change. There is also no feature mismatch in Word-Word repetitions, thus an increase in reaction time would not be predicted. However, the feature mismatch account cannot explain why there is a decrease in reaction time for the Word-Word repetitions. In summary, the feature mismatch account explains only the Word-Color effect, makes wrong predictions for the Color-Word and Color-Color effects, and does not explain the Word-Word effect.

Since the Color-Color effect has been reported in the inhibition-of-return literature (Law, Pratt, & Abrams, 1995) and interactions between Stroop effects and inhibition-of-return effects have been documented (Fuentes, Boucart, Vivas, Alvarez, & Zimmerman, 2000; Vivas & Fuentes, 2001), we have also considered the inhibition-free account of inhibition-of-return suggested by MacLeod, Dodd, Sheard, Wilson, and Bibi (2003) and called the *attentional momentum* account (Pratt, Spalek, & Bradshaw, 1999). According to this account, attention can be oriented toward locations along the direction

of orientation faster than to locations that require a change in the direction of orientation. This theory could not be applied to our case as such because all the stimuli appear at the same location. However, object-based and semantic inhibition-of-return effects have been documented (Fuentes, Vivas, & Humphreys, 1999; Tipper, Weaver, Jerreat, & Burak, 1994) and the attentional momentum theory can be extended to comprise objects and even abstract mental concepts such as *redness*. Provided we had such an extended theory of attentional momentum able to account for all inhibition-of-return effects, would it explain our between-trial effects? Interestingly, this theory might account for them all, if there were cues to redirect attention between the Stroop trials as in the classical inhibition-of-return task (i.e., the cue-target paradigm). However, the Stroop task was administered in a continuous target-target paradigm. There were no intervening stimuli (or cues) to redirect attention between two consecutive trials. For example, in the Color-Color case, attention needs to be directed toward the same concept as in the preceding trial, and its momentum should facilitate its orientation. Similarly, this account fails to explain the other three effects as well: since there is no change in the direction of attention with repetition, there is no reason for difficulty in re-orienting attention toward a just-attended entity. This account could only explain the 2- and 3-back versions of the between-trial effects, because in these cases there are intervening trials between repetitions.

If the existing inhibition-free theories cannot account for the presented data in an integrated way, then can the existing accounts based on cognitive inhibition do so? One of the most influential suppression-based accounts is *the selective inhibition* account (Houghton, Tipper, Weaver, & Shore, 1996; Neill & Westberry, 1987), which is

implemented as a computational model. This account posits an initial bottom-up activation of both features (word and color) followed by a top-down activation of the to-be-named feature (color) and inhibition of the to-be-ignored feature (word) of the current stimulus. When the inhibited feature returns as the to-be-named feature of the next stimulus, its inhibition has to be overridden by re-activation. This account predicts longer reaction times when the previous word feature re-occurs as the current color feature (i.e., the Word-Color effect), and shorter reaction time when the word repeats (i.e., the Word-Word effect). However, since only the to-be-ignored feature is inhibited (i.e., inhibition is selective), this account predicts that reaction time will not increase when the previous color re-occurs as the current color. In fact, in the Color-Color case, reaction times should decrease, since the to-be-named feature (color) has just been activated in the previous trial. For the same reason, this account cannot explain the Color-Word effect. The color of the preceding stimulus has been activated, thus when it re-occurs as the word of the current stimulus, it has a higher potential to interfere with the color naming of the current stimulus, thus causing reaction time to increase. In summary, the selective inhibition account predicts well only two of the between-trial effects and it makes wrong predictions for the remaining two. Thus, this account does not do any better than the inhibition-free accounts. It seems that this account fails when it tries to explain between-trial effects as by-products of within-trial effects, that is, when it posits that inhibition acts selectively at a trial level in order to prevent the distractor from interfering with the target. Lowe (1979, 1985) was among the first to challenge the selective inhibition account and to argue that between-trial effects in the Stroop task are to be attributed to other cognitive processes (strategic) than those causing within-trial effects. Milliken and

Joordens (1996) demonstrated that selection of targets from distractors was not necessary for the negative priming effect to occur. Therefore, we chose to treat the between-trial effects as independent of within-trial effects.

The account we propose, called *repetition suppression*, posits a control mechanism dedicated to between-trial interference. It is the temporal sequencing of trials for which this control mechanism is used rather than the selection of targets from distractors. Temporal sequencing of actions as a function of cognitive control is as important as the function of distinguishing the relevant from irrelevant information (Houghton & Tipper, 1996).

For the rest of the article we will use the term *suppression* instead of cognitive inhibition in order to avoid the confusion between cognitive and neural inhibition, as recommended by MacLeod (2007a). First, we will describe how repetition suppression can explain all of the between-trial effects in an integrated way. In the next section, we will present two computational models that implement this account and we will discuss how the brain could perform such a control function.

The repetition suppression account posits that at the end of a trial all representations that have been used to make a decision in that trial are suppressed in order to prevent their interference with the next trial. This suppression decreases in strength as the time passes and it can be detected in behavior only when repetitions occur. In fact, this account makes a stronger prediction: traces of this between-trial suppression *should* occur at any time when repetitions occur. Thus, in the Word-Color effect, the concept denoted by the word feature of the stimulus on the preceding trial re-occurs as the color feature of the current stimulus. Since the representation of this concept has been

suppressed, it takes longer to name the color than in trials without this kind of repetition. In the Color-Word effect, the word on the current trial has less potential to interfere with or to facilitate color naming because its corresponding concept has been suppressed. This fact causes reduced interference in the incongruent condition (i.e., decrease in reaction time) but also reduced facilitation in the congruent condition (i.e., increase in reaction time) (see Table 3). In the Color-Color effect, reaction time increases because the concept has just been suppressed and it needs reactivation to be used in the current trial. In the Word-Word effect, the word on the current trial has less potential to interfere with or to facilitate color naming in the incongruent and congruent conditions, respectively. This explains why reaction time decreases in the incongruent condition and increases in the congruent condition (see Table 3 and the interaction between condition and Word-Word in the LME model).

The repetition suppression account somewhat resembles the inhibition account of inhibition-of-return (Posner & Cohen, 1984; Tipper, Weaver, Jerreat, & Burak, 1994). What is different is that it operates at a semantic level. What is suppressed is not the “return” of a particular representation of an object or location but rather the represented concepts related to perceived features of stimuli, regardless of whether these features are targets or distractors (i.e., to be selected or to be ignored features). Repetition suppression is a memory-based account explaining effects that occur in a continuous target-target paradigm, and not an attention-based account explaining effects that occur in a cue-target paradigm.

A similar control mechanism dealing with past information that has become irrelevant for the current context is mentioned elsewhere and called *Resistance to*

*proactive interference* (Friedman & Miyake, 2004). While we acknowledge that Friedman and Miyake's term is relevant because it refers to the purpose of such control mechanism, we decided to use the term *Repetition suppression* because it refers strictly to the behavioral effects we have observed.

In summary, the repetition suppression account seems to be able to explain the between-trial effects better than concurrent accounts and it does so in an integrated way. The next self-imposed methodological constraint was to implement this theoretical account in a neurally plausible computational model.

### 3. Mechanisms of Repetition Suppression

As it is often the case, there is not a unique way to implement a theory at a computational level. One possible way to constrain a computational model of a particular task is to observe how the brain performs the same task. A review of the literature on inhibitory control revealed two neurally plausible ways to implement a repetition suppression account in a computational model. They will be respectively called *bottom-up* and *top-down* suppression. A computational cognitive model will be presented for each of these accounts. These models were developed with the aid of the last version of the ACT-R<sup>1</sup> cognitive architecture (Anderson, 2007). These two models are identical with regard to how they implement the within-trial effects (see section 3.1) and they only differ with regard to their implementation of the between-trial effects (sections 3.2 and 3.3).

---

<sup>1</sup> Adapted Control of Thought - Rational

### 3.1. The “Relative Automaticity” Account

Many models of the within-trial effects have been developed (Cohen, Dunbar, & McClelland, 1990; Herd, Banich, & O'Reilly, 2006; Lovett, 2005) and there seems to be a large consensus that a “relative automaticity” account best explains these effects (MacLeod & MacDonald, 2000). This account originates with an old finding in psychology, namely that reading words is more automatic than naming colors (Cattell, 1886; Fraisse, 1969). This is explained by the fact that human adults have vastly greater practice at reading words than at naming colors.

Our model of within-trial effects implements this difference in automaticity as a difference in strengths of association between representations of stimulus features and memory elements associated with these features. Thus, when a stimulus is perceived, its *word* and *color* dimensions are represented in the *goal buffer* – a storage structure used to maintain information that is important for controlling the flow of actions involved by the task at hand. For brevity, these representations will be referred to as *control units*. For example, if the current stimulus is the word “blue” in *red* ink (incongruent condition), the two control units are representations of the word “blue” and the color *red*, respectively. The two control units spread activation toward associated memory elements, thus biasing their retrieval. The word “blue” spreads activation toward the concept of blueness, while the color *red* spreads activation toward the concept of redness. The amount of activation spreading from the goal buffer (called *source activation*, for brevity) is limited and is equally shared by the two control units. The amount of activation received by a memory

element is a function of the amount of activation spread toward it and its strength of association with the corresponding control unit. In our model, words have larger strengths of association than colors, reflecting the difference in practice between reading words and naming colors. As a result, when a stimulus is presented, the concept associated with its word dimension is more active than the concept associated with its color dimension. In our example, *blueness* will be more active than *redness*. In order to name the color of the current stimulus, a memory retrieval request is being made and the concept of blueness is retrieved. At this point, if memory retrieval were sufficient for performing an action, the model would commit an error, responding *blue* instead of *red*. However, the behavior of an ACT-R model is guided not only by perception and memory retrievals but also by firing of production rules of the kind “if condition, then action.” In this case, a production rule detects the wrong retrieval and requests a new retrieval directed at the right color concept. Since memory retrievals take time, responses to incongruent stimuli take longer than responses to neutral stimuli. In the congruent condition, both control units spread activation toward the same concept in memory, thus increasing its activation and speeding up its retrieval (Figure 2). In addition, for congruent stimuli, the first retrieval is sufficient for generating a correct response, even when it is guided solely by the word dimension of the stimulus. Thus, facilitation is not simply the reverse of interference, as mentioned by MacLeod and MacDonald (2000). This way of modeling within-trial effects is similar to other models of the Stroop task (Altmann & Davidson, 2001; Cohen, Dunbar, & McClelland, 1990; Herd, Banich, & O'Reilly, 2006; Lovett, 2005; Roelofs, 2003). Notice that it does not require a mechanism of suppression of the more automatic response in favor of the less automatic, but task-relevant, response. Such a suppression

mechanism is only needed to account for between-trial effects, as shown in the following sections.

(Figure 2 about here)

### *3.2. Bottom-up Suppression (Reactive Inhibition)*

Repetition avoidance has been extensively studied in cognitive control tasks such as the task of generating sequences of random numbers (RNG). It seems that the process is relatively automatic and does not rely on a limited capacity resource (Baddeley, Emslie, Kolodny, & Duncan, 1998; Shallice, 2004). In addition, models of cognitive control in sequential behavior often postulate a biphasic pattern of activation and suppression. In short, this biphasic pattern consists of early activation followed by late suppression, which should allow activation at novel locations, objects, etc. (Klein, 2004; Pratt, Hillis, & Gold, 2001; Tipper, Weaver, Jerreat, & Burak, 1994) According to this idea, suppression follows activation in order to allow proper composition of sequences of actions (Houghton & Tipper, 1996). Reactive inhibition is a related concept which claims that inhibition is greater to the extent that a distractor is expected to intrude. Reactive inhibition seems to be an after-effect of processing that is not usually intended (Logan, 1994). The adaptive function of a bottom-up suppression mechanism is best explained in the following quotation:

“Many natural systems reflect a tendency for positive priming, such that an item that has recently occurred is more readily accessed, and therefore if the system is to avoid becoming locked in a positive feedback cycle of perseveration, there

needs to be some form of short-term and automatic inhibition or negative priming.” (Baddeley et al. 1998, pp. 846)

ACT-R uses a form of inhibitory tagging (Fuentes, 1999; Jonides & Smith, 1997) to implement inhibition of return effects in vision and to prevent perseverative retrieval in memory tasks. An attended location in a visual display is tagged as *attended* and the search for a new location to attend is biased toward locations that have not been tagged as *attended*. Similarly, a retrieved memory element is tagged as *recently-retrieved* and a new retrieval is biased toward memory elements that have not been tagged as *recently-retrieved*. Tags are attached to memory elements for a while and eliminated after a certain time has passed. This mechanism is called FINST (fingers of instantiation) and its principles are borrowed from Pylyshyn (2000).

For our purposes, the memory-finst mechanism seems appropriate to model bottom-up repetition suppression, because what is repeated is the semantic concept that underlies the visual features of the stimuli. For example, in the Color-Word effect, there is no repetition of any visual feature of the stimuli, but there is repetition of the concept that is instantiated first as a color and then as a word. Representations of concepts in memory are activated when the stimuli are perceived and one of these representations is needed for naming the color of the current stimulus. Retrieving the correct concepts from memory is the key toward generating correct responses.

A small adjustment to the standard memory-finst mechanism of ACT-R was necessary. First, all-or-none tags attached to memory elements, as it is the case in the standard ACT-R, were counterproductive. They completely blocked a recently retrieved memory element from being re-retrieved. However, the observed between-trial effects

suggest that re-retrieval is delayed but not completely blocked. Thus, the finsts have been assigned a continuous value instead of an all-or-none value. This value is subtracted from the activation of the corresponding memory element and thus it slows down its retrieval, instead of blocking it. Second, the empirical data suggest a gradual decay of this continuous value (see Figure 1), instead of a constant delay, as it is the case in the standard ACT-R. This adjustment will be referred to as “the decaying finst mechanism”.

By adding a decaying finst mechanism to the model described in section 3.1, the pattern of between-trial effects shown in Figure 3 has been obtained. Thus, in the Word-Color (W-C) trials, the concept corresponding to the word feature of the preceding stimulus has been retrieved and *finsted*<sup>2</sup> (i.e., its activation has been discounted). When the same concept needs to be re-retrieved to name the color of the current stimulus, retrieval takes longer than in control cases. When the repetition is not immediate, but rather occurs over two or three trials (W2-C and W3-C, respectively), the effect magnitude gradually diminishes because the finst has decayed.

(Figure 3 about here)

In the Color-Word trials, the concept corresponding to the color feature of the preceding stimulus has been retrieved and finsted. The same concept is associated with the word feature of the current stimulus. Normally, this concept would have high activation due to its high strength of association with the *word control unit* from the goal buffer and would interfere with color naming on the current trial. However, because it has

---

<sup>2</sup> This term is also used by the author of the FINST concept (Pylyshyn, 2000)

been finsted, this concept is less likely to be retrieved in the current trial, that is, it has less potential to interfere with color naming on this trial. This explains the decrease in reaction time for Color-Word trials.

In the Color-Color trials, the concept corresponding to the color feature of the preceding stimulus has been retrieved and finsted. The same concept is associated with the color feature of the current stimulus and it takes longer to be re-retrieved in order to be used in naming the color. In the Word-Word trials, the concept corresponding to the word feature of the preceding stimulus has been retrieved and finsted. The same concept is associated with the word feature of the current stimulus. Normally, this concept would interfere with color naming on the current trial, but due to its finsted activation it is less likely to be retrieved, thus allowing a faster color naming than in control trials. However, in the congruent condition, retrieval of this concept would normally produce facilitation, which is prevented from occurring because the concept has been finsted. This explains the increase in reaction time in the Word-Word trials in the congruent condition as observed in the empirical data (see table 3).

This model produces a reasonably good fit to the empirical data (Correlation = 0.924, Mean deviation = 15.4 ms) by implementing a relatively simple mechanism – decaying finst. This mechanism seems to implement well the main characteristics of a bottom-up repetition suppression account, that is, it automatically applies to all memories, it is a short-term after-effect of activation, and it serves the function of preventing positive feedback and perseveration. However, as shown in Figure 3, the model produces smaller magnitudes for the Word-Color effect and larger magnitudes for the Color-Word and Color-Color effects than observed in the empirical data. These

deviations are caused by an intrinsic characteristic of the decaying-finst mechanism, that is, it acts after retrieval. In terms of Logan (1994), the suppression modeled by the decaying-finst mechanism is an after-effect of activation. In the terms of our ACT-R model, a memory element gets finsted only if and immediately after it has been retrieved. Because the finst decays, the exact moment of retrieval determines the magnitude of the aftereffect. Thus in Word-Color trials, the concept corresponding to the word feature of the preceding stimulus was retrieved before the concept corresponding to the color feature of the preceding stimulus, because of its higher strength of association with the word control unit. By the time when the same concept needs to be re-retrieved to name the color of the current stimulus, the finst has already partially decayed. This is why the Word-Color effects are smaller in magnitude than expected. When the concept corresponding to the color feature of the preceding stimulus repeats, as in the Color-Word and Color-Color effects, the effects are larger because the retrieval and the subsequent finst have happened more recently than in the case of repeating the concept corresponding to the word feature. Thus, these local misfits are caused by the sequential order of processing for the word and color dimensions of the stimulus. However, there is evidence that the two dimensions are processed in parallel (MacLeod & Bors, 2002). If the two concepts were simultaneously retrieved and then finsted, these misfits would probably not occur.

The neural substrate of bottom-up (reactive) suppression could be the circuitry between locus coeruleus (LC) and cortex. Neural activity is increased in LC and its projections to cortex during processing of task-relevant stimuli. After processing of a particular stimulus, LC inhibits itself via local connections, creating a refractory period in

which the cortex is unable to process the same stimulus or similar stimuli (Nieuwenhuis, Gilzenrat, Holmes, & Cohen, 2005).

### *3.3. Top-Down Suppression (Active Inhibition)*

Performance in the Stroop task usually correlates with performance in tasks that are known to involve top-down suppression (Miyake et al., 2000), such as the antisaccade task (Hallett, 1978) and stop-signal task (Logan, 1994). Stroop effects have been shown to reverse, diminish or disappear in a number of circumstances (Harrison & Espelid, 2004; MacLeod, 1991; MacLeod & Sheehan, 2003; Metzler & Parkin, 2000). In particular, negative priming (the Word-Color effect) occurs only under specific task instructions (MacLeod, 1991) after extensive practice with a small and homogeneous set of stimuli (Weger & Inhoff, 2006). These findings suggest that between-trial effects might be caused by the same top-down suppression mechanism involved in stop signal tasks (Aron et al., 2007; Levy & Anderson, 2008; Verbruggen, Logan, Liefoghe, & Vandierendonck, in press).

A model implementing a top-down suppression mechanism has been developed as an alternative to the decaying first model. This model postulates a control structure that is external to the memory retrieval process and only influences it in particular circumstances. This structure, called “the suppression buffer,” is analogous to the goal buffer in that it has a biasing influence on memory retrieval, except that this influence is opposite in sign. The activation spread from the goal buffer has an excitatory influence on the targeted memory elements, whereas the activation spread from the suppression

buffer has an inhibitory influence on the targeted memory elements. The goal buffer and the suppression buffer are independent of each other, that is, they can operate simultaneously on the same memory element. In other words, suppression is not an aftereffect of activation. Another important difference with the decaying first model is that suppression is not a property of a memory element (a quantity that is attached to it) but it is an external influence that lasts as long as the source of this influence lasts.

This model assumes that top-down suppression is employed in order to prevent information from preceding trials from interfering with the processing required by the current trial. After a trial has been completed, its associated stimulus is represented in the suppression buffer. The two features of this stimulus (word and color) spread negative activation (i.e., suppression) toward their associated concepts represented in memory. When the suppressed concepts need to be retrieved to assist with the processing of the current stimulus, the known between-trial effects are manifested. The suppression buffer can accommodate up to three preceding stimuli and their source of suppression decays gradually. Buffers that accommodate multiple representations with different decaying sources of spreading activation are not supported by the standard ACT-R architecture. Negative spreading activation is usually not used in ACT-R models (although, see Van Maanen and Van Rijn, 2007, for an exception). These features were created ad-hoc for this model in order to illustrate the theoretical value of a top-down suppression account.

Running the top-down suppression model produced the pattern of results presented in Figure 4. This model also produces a good fit to the data (Correlation = 0.949, Mean deviation = 11.4 ms), even a slightly better fit than the decaying first model. The four between-trial effects have now comparable magnitudes. However, the Color-

Color effect in the empirical data is not as large as the model estimates it. Perhaps the magnitude of the Color-Color effect is also modulated by other factors that we have not considered in our model.

The neural substrate of top-down suppression could be the fronto-subthalamic circuitry, also called the indirect pathway (Aron, 2007; Chambers et al., 2006). The top-down suppression signal is originated in the right inferior frontal cortex and activates the subthalamic nucleus (STN). STN excites globus pallidum and this inhibits thalamus. As a result, thalamus stops activating specific cortical areas.

(Figure 4 about here)

#### 4. General Discussion and Conclusion

Criticism has recently been expressed with reference to psychological theories that postulate suppression (cognitive inhibition) as an explanatory mechanism for observed behavioral effects (MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003). Since we agreed that many of these points of criticism were justified, we seriously considered them while conducting the research reported here. One of these criticisms states that the term cognitive inhibition is misleading because it creates confusion with the phenomenon of neural inhibition. In response to this criticism, we have adopted the term “suppression” instead of “cognitive inhibition.” Another criticism points at a tendency to postulate suppression for any findings showing decreases in performance before alternative suppression-free accounts have been considered. We have addressed known behavioral

effects showing both increases and decreases in performance and tried to explain them with an integrated account. Before postulating a suppression account, we have analyzed the existing suppression-free accounts, and showed that they fail to explain all effects in an integrated way. The suppression mechanism proposed here is also different from the classical selective inhibition account because it addresses between-trial interferences as opposed to within-trial interference. Friedman and Miyake (2004) suggested that *Resistance to proactive interference* (what we have called *Repetition suppression*) might be a distinct dimension of inhibitory control, separate from *Prepotent response inhibition* or *Resistance to distractor interference*.

Although the between-trial effects presented here have been known for a long time, we have replicated all of them in the same study and have been able to estimate their time course and relative magnitudes while controlling for within-trial effects, as suggested by MacLeod (1991). To our knowledge, an integrated account for these effects has not been previously proposed. We have shown that a repetition suppression account can explain all of these effects and we have demonstrated two neurally plausible ways to implement this account in a computational cognitive model.

The repetition suppression account has been shown to explain the presented data better than alternative accounts. However, we have no grounds to generalize the outcome of this comparative analysis beyond the task and the dataset presented here. In other words, we do not imply that the theoretical accounts shown to fail here are fundamentally invalid. Neither do we imply that the repetition suppression account presented here should explain all of the data in the Negative Priming and Inhibition-of-Return literatures. We do acknowledge that the sequence effects in the Stroop task presented

here and the hypothesized inhibitory control mechanisms might not be reproducible in different tasks under different circumstances. It is characteristic of inhibitory control mechanisms to be employed only in specific circumstances (Neill & Westberry, 1987; Lowe, 1985; Tipper & Cranston, 1985; Weger & Inhoff, 2006), related to task difficulty, information load, amount of interference, emphasis on speed vs. accuracy, practice, size of the set of stimuli, probability of trial types, etc.

The computational mechanism that we have used to model bottom-up suppression (decaying-finst) is in line with the way the ACT-R theory models suppression in memory and vision phenomena. We have only added a decaying characteristic to the classical finst mechanism of ACT-R in order to account for the observed gradual decay of between-trial effects. The mechanism we have used to model top-down suppression (decaying negative spreading activation) is relatively novel and it needs more research in order to be fully validated. Although evidence in favor of this account from fMRI and TMS studies tends to accumulate (Aron, 2007; Chambers et al., 2006), there are many aspects that need further investigation. It is not possible at this point to differentiate between the top-down and bottom-up suppression models. Both are plausible and fit the data equally well. Alternative accounts remain to be explored. For instance, the top-down mechanism could be conceived of as a series of accumulating bottom-up influences that are not tied to the retrieval process but rather to the activation process. In an upcoming study, we will test this possible account. We will try to separate the activation process from the retrieval process and check if the between-trial effects described here will occur under the activation-only condition.

In conclusion, we have attempted to demonstrate that a repetition-suppression mechanism is a viable theoretical account for explaining a range of between-trial effects in an integrated way. More research is needed to adequately characterize this account and, in particular, to answer the question whether repetition suppression is an intrinsic property of memory activation or it is the effect of a controlling influence that is external to the activation process and it only acts in specific circumstances.

References

- Altmann, E. M., & Davidson, D. J. (2001). *An integrative approach to Stroop: Combining a language model and a unified cognitive theory* Paper presented at the Twenty-Third Annual Conference of the Cognitive Science Society, Hillsdale, NJ
- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.
- Aron, A. R. (2007). The Neural Basis of Inhibition in Cognitive Control *Neuroscientist*, 13(3), 214-228.
- Aron, A. R., Durston, S., Eagle, D. M., Logan, G. D., Stinear, C. M., & Stuphorn, V. (2007). Converging Evidence for a Fronto-Basal-Ganglia Network for Inhibitory Control of Action and Cognition. *The Journal of Neuroscience* 27(44), 11860-11864.
- Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random Generation and the Executive Control of Working Memory. *The Quarterly Journal of Experimental Psychology: Section A*, 51(4), 819-852.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11, 63-65.
- Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., et al. (2006). Executive "Brake Failure" following Deactivation of Human Frontal Lobe. *Journal of Cognitive Neuroscience*, 18(3), 444 - 455.
- Christie, J., & Klein, R. M. (2001). Negative priming for spatial location? *Can J Exp Psychol*, 55(1), 24-38.

- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332-361.
- Effler, M. (1977). The influence of serial factors on the Stroop test. *Psychologische Beitrage*, *19*, 189-200.
- Effler, M. (1980). Processes in naming Stroop-stimuli: An analysis with word repetition effects. *Archiv fur Psychologie*, *133*, 249-262.
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(1784-1790).
- Erickson, M. A., & Reder, L. M. (1998). *The Influence of Repeated Presentations and Intervening Trials on Negative Priming* Paper presented at the The Twentieth Annual Conference of the Cognitive Science Society Mahwah, NJ
- Fraisse, P. (1969). Why is naming longer than reading? . *Acta psychol.*, *30*, 96-103.
- Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General*, *133*(1), 101-135.
- Fuentes, L. J. (1999). Inhibitory Tagging of Stimulus Properties in Inhibition of Return: Effects on Semantic Priming and Flanker Interference *The Quarterly Journal of Experimental Psychology: Section A*, *52*(1), 149 - 164.
- Fuentes, L. J., Boucart, M., Vivas, A. B., Alvarez, R., & Zimmerman, M. A. (2000). Inhibitory tagging in inhibition of return is affected in schizophrenia: evidence from the stroop task. *Neuropsychology*, *14*(1), 134-140.

- Fuentes, L. J., Vivas, A. B., & Humphreys, G. W. (1999). Inhibitory mechanisms of attentional networks: Spatial and semantic inhibitory processing *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1114-1126.
- Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18, 1279–1296.
- Harrison, Y., & Espelid, E. (2004). Loss of negative priming following sleep deprivation. *The Quarterly Journal of Experimental Psychology: Section A*, 57(3), 437-446.
- Herd, S. A., Banich, M. T., & O'Reilly, R. C. (2006). Neural Mechanisms of Cognitive Control: An Integrative Model of Stroop Task Performance and fMRI Data *J. Cognitive Neuroscience MIT Press*, 18(1 ), 22-32
- Houghton, G., & Tipper, S. P. (1996). Inhibitory Mechanisms of Neural and Cognitive Control: Applications to Selective Attention and Sequential Action *Brain and Cognition*, 30, 20-43.
- Houghton, G., Tipper, S. P., Weaver, B., & Shore, D. I. (1996). Inhibition and interference in selective attention: Some tests of a neural network model. *Visual Cognition*, 3(119–164).
- Jonides, J., & Smith, E. E. (1997). The architecture of working memory. In M. D. Rugg (Ed.), *Cognitive Neuroscience* (pp. 243-276). Sussex, England: Psychology Press.
- Klein, R. M. (2004). Orienting and Inhibition of Return. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences III*. Cambridge, Massachusetts: MIT Press.
- Law, M. B., Pratt, J., & Abrams, R. A. (1995). Color-based inhibition of return. *Percept Psychophys.*, 57(3), 402-408.

- Levy, B. J., & Anderson, M. C. (2008). Individual differences in the suppression of unwanted memories: The executive deficit hypothesis. *Acta Psychologica, 127*, 623-635.
- Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22*, 1-35.
- Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language*. San Diego: Academic Press.
- Lovett, M. C. (2005). A Strategy-Based Interpretation of Stroop. *Cognitive Science(29)*, 493-524.
- Lowe, D. G. (1979). Strategies, context, and the mechanism of response inhibition *Memory and Cognition, 7(5)*, 382-389.
- Lowe, D. G. (1985). Further investigations of inhibitory mechanisms in attention. *Memory and Cognition, 13(1)*, 74-80.
- MacDonald, P. A., & Joordens, S. (2000). Investigating a memory-based account of negative priming: support for selection-feature mismatch. *Journal of Experimental Psychology: Human Perception and Performance, 26(4)*, 1478-1496.
- MacLeod, C. M. (1991). Half a Century of Research on the Stroop Effect: An Integrative Review *Psychological Bulletin, 109(2)*, 163-203.
- MacLeod, C. M. (2007a). Cognitive inhibition: Elusive or illusion? In I. H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 301-305). New York: Oxford University Press.

- MacLeod, C. M. (2007b). The concept of inhibition in cognition. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 3-23 ). Washington, DC: American Psychological Association.
- MacLeod, C. M., & Bors, D. A. (2002). Presenting two color words on a single Stroop trial: Evidence for joint influence, not capture *Memory and Cognition*, *30*(5), 789-797.
- MacLeod, C. M., & MacDonald, P. A. (2000). Inter-dimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383-391.
- MacLeod, C. M., & Sheehan, P. W. (2003). Hypnotic control of attention in the Stroop task: A historical footnote. *Consciousness and Cognition* *12*(3), 347-353.
- MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In Opposition to Inhibition. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 43, pp. 163-214): Elsevier Science.
- Metzler, C., & Parkin, A. J. (2000). Reversed negative priming following frontal lobe lesions. *Neuropsychologia*, *38*, 363-379.
- Milliken, B., & Joordens, S. (1996). Negative priming without overt prime selection. *Canadian Journal of Experimental Psychology*, *50*(4), 333-346.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.

- Neill, W. T. (1978). Decision processes in selective attention: Response priming in the Stroop color-word task. *Perception and Psychophysics*, *23*, 80-84.
- Neill, W. T. (1997). Episodic Retrieval in Negative Priming and Repetition Priming. *Journal of Experimental Psychology; Learning, Memory, and Cognition* *23*(6), 1291-1305.
- Neill, W. T., & Westberry, R. L. (1987). Selective attention and the suppression of cognitive noise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 327-334.
- Nielsen, G. D. (1975). The locus and mechanism of the Stroop color word effect (Doctoral dissertation, University of Wisconsin-Madison, 1974). *Dissertation Abstracts International*, *35*, 5672-B.
- Nieuwenhuis, S., Gilzenrat, M. S., Holmes, B. D., & Cohen, J. D. (2005). The Role of the Locus Coeruleus in Mediating the Attentional Blink: A Neurocomputational Theory. *Journal of Experimental Psychology: General*, *134*(3), 291–307.
- Park, J., & Kanwisher, N. (1994). Negative priming for spatial locations: Identity mismatching, not distractor inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 613-623.
- Posner, M. I., & Cohen, Y. (1984). Components of visual attention. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention & Performance* (Vol. 10, pp. pp. 531 - 556). Hillsdale, NJ: Erlbaum.
- Pratt, J., Hillis, J., & Gold, J. M. (2001). The effect of the physical characteristics of cues and targets on facilitation and inhibition. *Psychonomic Bulletin & Review*, *8*(3), 489-495.

- Pratt, J., Spalek, T. M., & Bradshaw, F. (1999). The time to detect targets at inhibited and noninhibited locations: Preliminary evidence for attentional momentum. *Journal of Experimental Psychology: Human Perception and Performance*, 25(730–746).
- Pylyshyn, Z. W. (2000). Situating vision in the world *Trends in Cognitive Sciences*, 4(5), 197-207.
- Roelofs, A. P. A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task *Psychological review*, 110, 88-124.
- Shallice, T. (2004). The fractionation of supervisory control. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences III*. Cambridge, Massachusetts MIT Press.
- Thomas, J. K. (1977). *Stroop interference with word or hue pre-exposure*. University of Michigan.
- Tipper, S. P. (1985). The negative priming effect: Inhibitory effects of ignored primes. *Quarterly Journal of Experimental Psychology*, 37A, 571-590.
- Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *The Quarterly Journal of Experimental Psychology: Section A*, 54A, 321–343.
- Tipper, S. P., Weaver, B., Cameron, S., Brehaut, J. C., & Bastedo, J. (1991). Inhibitory mechanisms of attention in identification and localization tasks: Time course and disruption. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(4), 681-692.
- Tipper, S. P., Weaver, B., Jerreat, L. M., & Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual attention. *J Exp Psychol Hum Percept Perform*, 20(3), 478-499.

- Van Maanen, L., & Van Rijn, H. (2007). An Accumulator Model of Semantic Interference. *Cognitive Systems Research*, 8(3), 174-181.
- Verbruggen, F., Logan, G. D., Liefoghe, B., & Vandierendonck, A. (in press). Short-term aftereffects of response inhibition: Repetition priming or between-trial control adjustments? *Journal of Experimental Psychology: Human Perception and Performance*.
- Vivas, A. B., & Fuentes, L. J. (2001). Stroop interference is affected in inhibition of return. *Psychon Bull Rev*, 8(2), 315-323.
- Weger, U. W., & Inhoff, A. W. (2006). Semantic inhibition of return is the exception rather than the rule. *Perception and Psychophysics*, 68, 244-253.

Table 1

*Typical Within-Trial Effects*

	Incongruent	Congruent	Neutral
Accuracy (%)	0.92	0.99	0.98
Mean RT (ms)	1183.6	953.1	1055

*Note.* The reaction time (RT) in the incongruent and congruent conditions is higher and lower, respectively, than in the neutral condition (interference and facilitation, respectively). The inverse pattern occurs for accuracy data.

Table 2

*Examples of between-trial effects*

	Preceding trial	Current trial	Reaction time
Word-Color	RED	GREEN	Increase
Color-Word	YELLOW	RED	Decrease
Word-Word	BLUE	BLUE	Decrease
Color-Color	RED	BLUE	Increase

Table 3

*The Absolute Magnitudes (ms) of the Between-Trial Effects per Condition*

	Incongruent	Congruent	Neutral
Word-Color	102	93	40
Color-Word	-74	8	---
Word-Word	-68	93	-51
Color-Color	79	8	23

Table 4

*Results of the Initial LME Model*

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1197.1	27.4	16140	43.7	0.000
Congruent Condition	-235.9	10.5	16140	-22.4	0.000
Neutral Condition	-156.3	10.9	16140	-14.3	0.000
Word-Color	50.7	12.9	16140	3.9	0.000
Color-Word	-37.1	12.5	16140	-3.0	0.003
Color-Color	31.7	10.1	16140	3.1	0.002
Word-Word	-39.7	19.3	16140	-2.1	0.039
Word-Color-2back	29.1	11.3	16140	2.6	0.010
Color-Word-2back	-26.3	12.0	16140	-2.2	0.028
Color-Color-2back	16.4	9.5	16140	1.7	0.085
Word-Word-2back	-16.7	13.7	16140	-1.2	0.222
Word-Color-3back	8.0	11.1	16140	0.7	0.473
Color-Word-3back	-7.4	12.0	16140	-0.6	0.539
Color-Color-3back	7.4	9.7	16140	0.8	0.445
Word-Word-3back	2.4	13.7	16140	0.2	0.863
Congr*Word-Word	99.6	33.3	16140	3.0	0.003
Neutral*Word-Word	25.9	55.8	16140	0.5	0.642

Table 5

*Results of the Best Fitting LME Model*

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1197.7	27.0	16145	44.3	0.000
Cond-congruent	-236.2	10.5	16145	-22.5	0.000
Cond-neutral	-153.5	10.4	16145	-14.7	0.000
Word-Color	50.9	12.9	16145	4.0	0.000
Color-Word	-37.4	12.5	16145	-3.0	0.003
Color-Color	31.6	10.1	16145	3.1	0.002
Word-Word	-39.6	19.3	16145	-2.1	0.040
Word-Color-2back	24.9	10.7	16145	2.3	0.020
Color-Word-2back	-30.1	11.5	16145	-2.6	0.009
Color-Color-2back	17.2	9.5	16145	1.8	0.070
Congr*Word-Word	99.0	33.3	16145	3.0	0.003
Neutral*Word-Word	26.4	55.8	16145	0.5	0.636

## Figure Captions

*Figure 1.* LME Coefficients for the between-trial effects. Stars above some of the bars indicate significant effects ( $\alpha=0.05$ ).

*Figure 2.* Within-trial effects as shown in the empirical data and in the two models presented in sections 3.2 and 3.3.

*Figure 3.* The pattern of between-trial effects as estimated from empirical data and simulated by the “decaying first” model. On the horizontal axis the four between-trial effects and their corresponding 2- and 3-back variants are deployed. The vertical axis shows the magnitudes of these effects: positive values indicate increase and negative values decrease in reaction time as compared to control trials. The error bars indicate standard errors of the means.

*Figure 4.* The pattern of between-trial effects as estimated from empirical data and simulated by the “top-down suppression” model. On the horizontal axis the four between-trial effects and their corresponding 2- and 3-back variants are deployed. The vertical axis shows the magnitudes of these effects: positive values indicate increase and negative values decrease in reaction time as compared to control trials. The error bars indicate standard errors of the means.

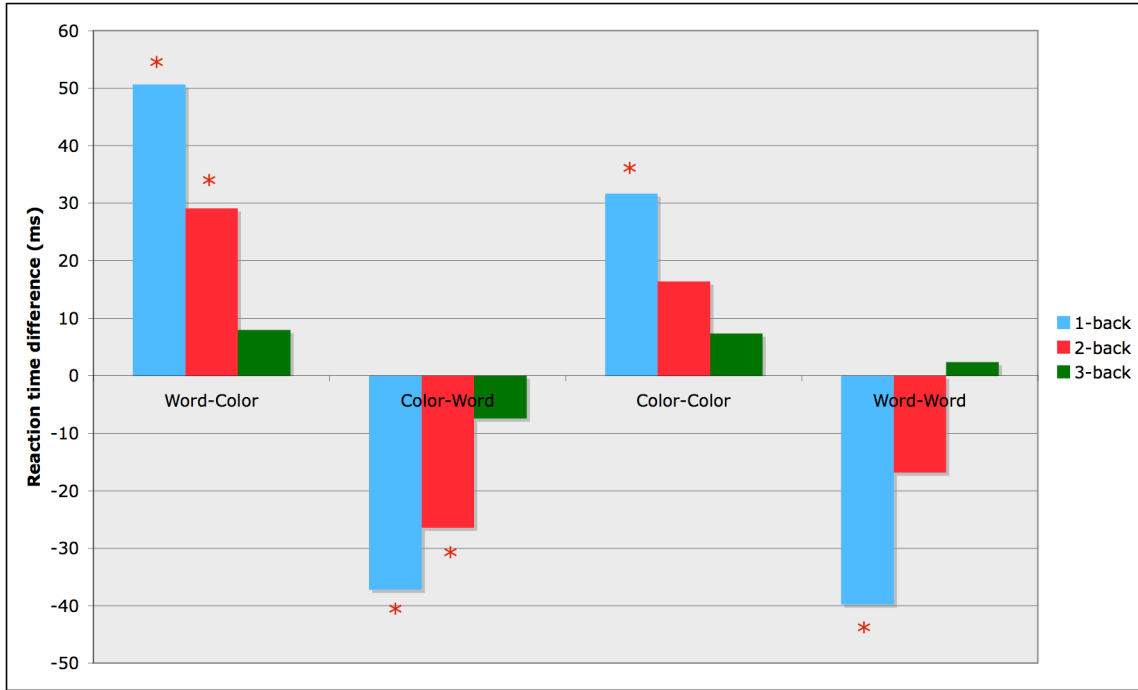


Figure 1. LME Coefficients for the between-trial effects. Stars above some of the bars indicate significant effects ( $\alpha=0.05$ ).

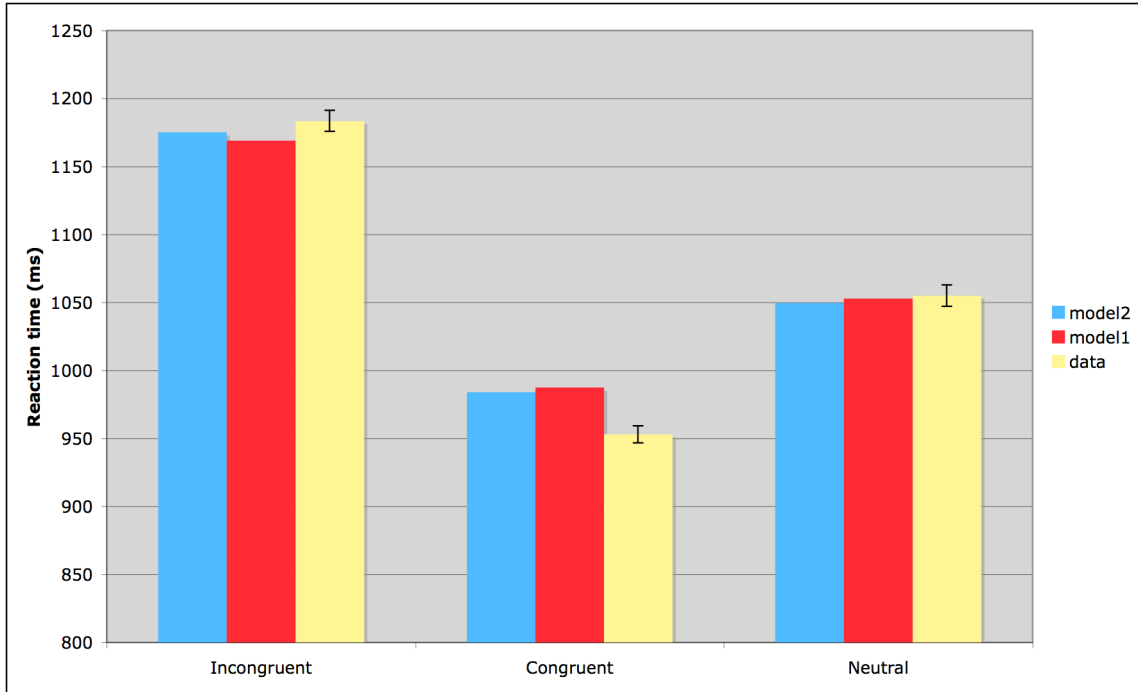


Figure 2. Within-trial effects as shown in the empirical data and in the two models presented in sections 3.2 and 3.3.

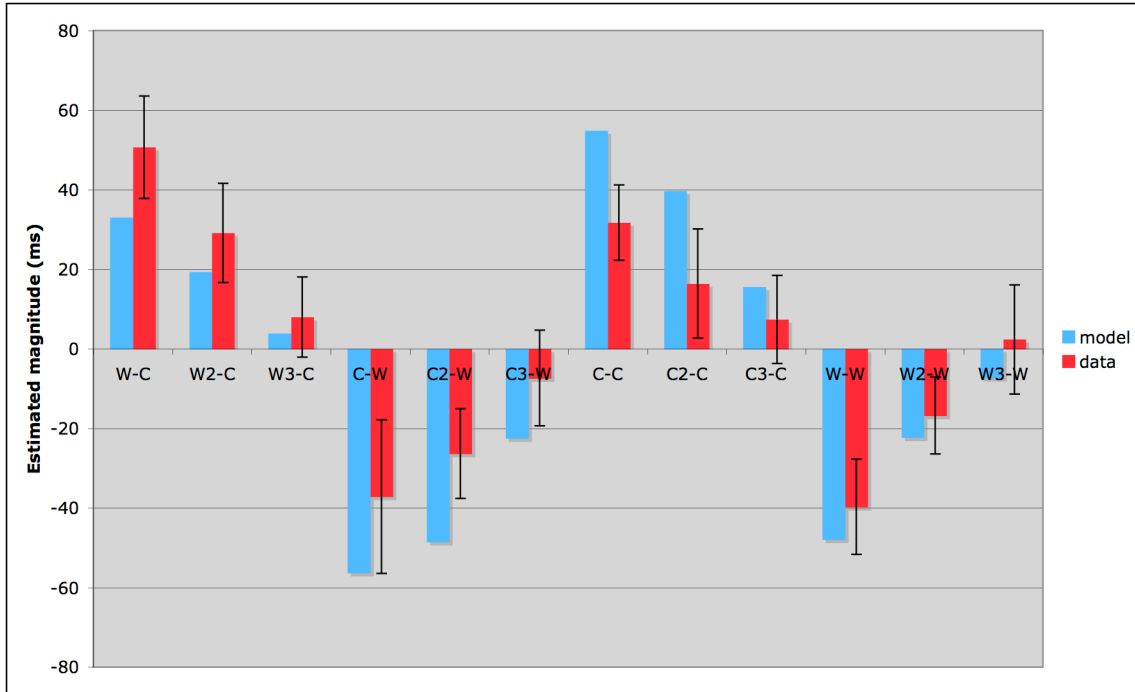


Figure 3. The pattern of between-trial effects as estimated from empirical data and simulated by the “decaying first” model. On the horizontal axis the four between-trial effects and their corresponding 2- and 3-back variants are deployed. The vertical axis shows the magnitudes of these effects: positive values indicate increase and negative values decrease in reaction time as compared to control trials. The error bars indicate standard errors of the means.

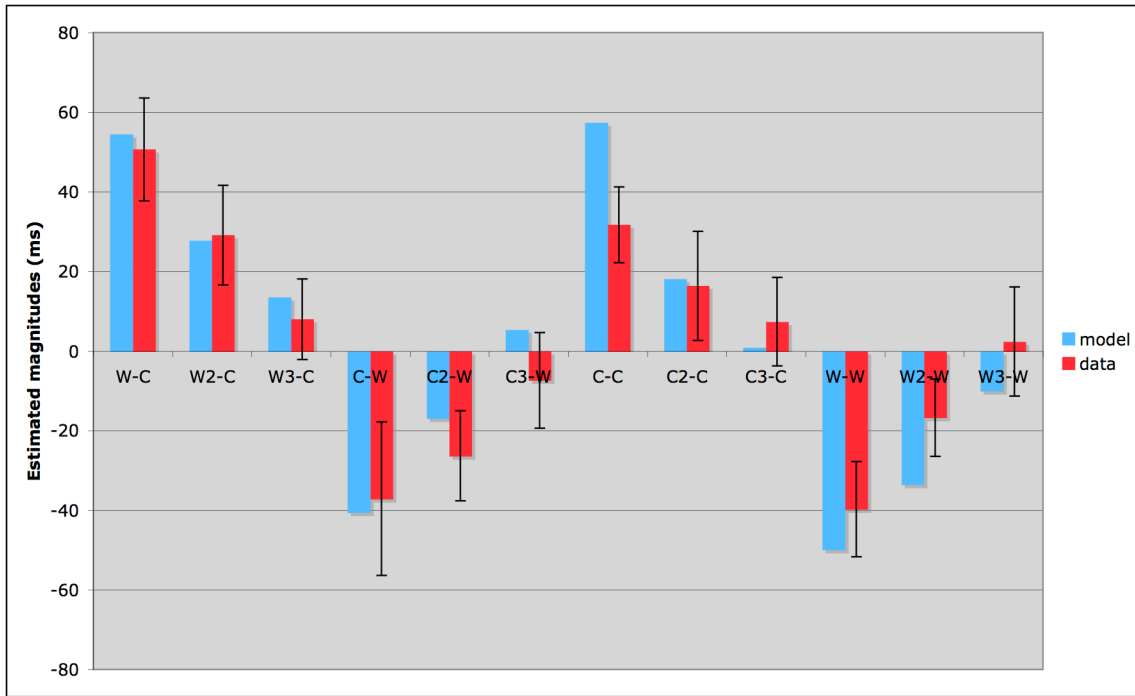


Figure 4. The pattern of between-trial effects as estimated from empirical data and simulated by the “top-down suppression” model. On the horizontal axis the four between-trial effects and their corresponding 2- and 3-back variants are deployed. The vertical axis shows the magnitudes of these effects: positive values indicate increase and negative values decrease in reaction time as compared to control trials. The error bars indicate standard errors of the means.