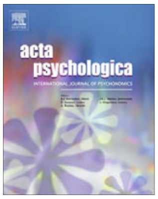
Acta Psychologica 131 (2009) 72–84

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



A repetition-suppression account of between-trial effects in a modified Stroop paradigm

Ion Juvina[a,*], Niels A. Taatgen[b,c]

[a] Department of Psychology, Baker Hall, 336A, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States
[b] Department of Psychology, Baker Hall, 345B, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States
[c] Department of Artificial Intelligence, University of Groningen, Nijenborgh 9, 9747 AG Groningen, Netherlands

ARTICLE INFO

Article history:
Received 20 September 2008
Received in revised form 27 February 2009
Accepted 10 March 2009

PsycINFO classification:
2340

Keywords:
Stroop
Cognitive inhibition
Sequence effects
Cognitive modeling
Negative priming

ABSTRACT

Theories that postulate cognitive inhibition are very common in psychology and cognitive neuroscience [e.g., Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, A. Towse, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford, University Press], although they have recently been severely criticized [e.g., MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 163–214). Elsevier Science]. This paper poses and attempts to answer the question whether a research program with cognitive inhibition as its main theoretical assumption is still worth pursuing. We present a set of empirical data from a modified Stroop paradigm that replicates previously reported findings. These findings refer to between-trial effects previously described in the literature on Stroop, negative priming, and inhibition-of-return. Existing theoretical accounts fail to explain all these effects in an integrated way. A repetition-suppression mechanism is proposed in order to account for these data. This mechanism is instantiated as a computational cognitive model. The theoretical implications of this model are discussed.

One of the typical functions of cognitive control is interference resolution – that is, protecting the execution of task-relevant sequences of actions against interference and distraction. It is currently under debate whether *cognitive inhibition* (also referred to as *cognitive suppression*) is one of the mechanisms of interference resolution. Some authors assert that cognitive inhibition is essential for cognitive control (Aron, 2007; Druey & Hubner, 2008; Hasher et al., 2007; Houghton & Tipper, 1996); others say that it is unnecessary (Egner & Hirsch, 2005; Hommel, Proctor, & Vu, 2004; MacLeod et al., 2003; Rothermund, Wentura, & De Houwer, 2005).

This paper attempts to disentangle these two concurrent theoretical positions regarding cognitive control. In this paper the theoretical stance postulating that cognitive suppression is not necessary for interference resolution will be referred to as “the no-suppression theory”. The theoretical stance postulating that cognitive suppression is essential for interference resolution will be referred to as “the suppression theory”. In order to disentangle these two concurrent theories we will impose two methodological constraints: (1) a viable theoretical account should be able to simultaneously explain a large range of effects, and (2) it should be expressed in computational terms and be able to make numerical predictions (Anderson, 2007; Christie & Klein, 2008; Meehl, 1990).

A modified Stroop paradigm will be used to test the verisimilitude of the two theories. The Stroop task 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). We have modified the classical Stroop paradigm by changing the response registration procedure. Details about the modified Stroop paradigm are presented together with the description of the first study.

The following section presents the first empirical study aimed at replicating the known between-trial effects in the Stroop task. Section 2 shows how these empirical data challenge established theories and models of cognitive control and presents an alternative account. Section 3 presents a computational model that implements this alternative account and makes detailed predictions for all the interactions among within-trial and between-trial effects observed in the first study. Section 4 presents the second empirical study aimed at testing model predictions. Section 5 concludes the paper with a discussion about the plausibility of cognitive inhibition as one of the mechanisms of cognitive control.

[*] Corresponding author. Tel.: +1 412 268 2837.
E-mail addresses: ijuvina@cmu.edu, ionjuvina@mac.com (I. Juvina).

0001-6918/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2009.03.002

1. First study

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, p. 178).

It was our intention to conduct such an "interesting study" to replicate all these between-trial effects and to 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, 2007b, 2003). Thus, the research question we address here is whether this integrated account should be based on suppression (cognitive inhibition) or not. This question has inspired a plethora of recent empirical research and theoretical analyses (e.g., Aron, 2007; Christie & Klein, 2008; Druey & Hubner, 2008; Hasher et al., 2007; MacLeod, 2007a, 2007b; to name just a few).

1.1. Method

1.1.1. Participants

Fifty-three participants were recruited from Carnegie Mellon University's community via a website advertisement. Participant age ranged from 18 to 59 years with an average of 24. There were 16 women and 37 men. They received a fixed amount of monetary compensation for their participation.

1.1.2. Design

There were three within-subject conditions: incongruent, congruent, and neutral. Every participant received 150 trials, 50 trials for each condition. The three trial types corresponding to the three conditions were randomly mixed (non-blocked). Trial order was randomized for each participant. Between-trial conditions occurred from this random sequencing of the three trial types.

1.1.3. Apparatus and procedure

The standard Stroop task was adapted for screen-based administration and manual response. 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 53 common English words unrelated semantically or phonologically to any of the color names. They were selected from the most frequent nouns in English and their relatedness to the color names was judged by the experimenters. Stimuli were presented one at a time in the center of the screen and they remained on the screen until the participant responded. A fixation cross was presented in the middle of the screen for 1.5 s before the onset of a new stimulus. 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. In the incongruent condition the incorrect answer was identical to the distractor word. The location of stimuli on the screen was kept constant.

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. The reason for altering the standard response registration procedure is presented in the following paragraph.

This task was part of a larger study aimed at investigating the cognitive control aspects of multi-tasking. We were interested in interference control in tasks that involve perceptual, cognitive, and motor components; the vocal component was not of interest for us in this project. For this reason we considered using a manual version of the task. However, the typical manual Stroop task, in which each color is mapped on a unique manual response, has been shown to produce reduced levels of interference and fast decrease in interference with practice (see MacLeod (1991), for a review). The reduced interference is probably caused by the direct association that is formed with practice between the perception of colors and the associated manual responses. The mapped key presses loose their dimensional overlap with color concepts (Kornblum, 1994) because the retrieval of a color name is likely to be bypassed. When memory retrievals are bypassed, the main source of interference in the Stroop task, that is reading and retrieving color names, no longer exists. By asking participants to select the right answer from two options given on the screen, we reintroduced the words as source of interference. This way, naming a color involves going through a verbal step. Thus, having to select names of colors presented on screen makes the manual Stroop task more compatible with the standard (vocal) Stroop task, by bringing back its semantic and linguistic components. Interference arises from the possibility to retrieve an incorrect color name as in the vocal variant of the task. Each response option has an equal probability to appear on the left or right side of the stimulus, thus, preventing the selection process from becoming automated.

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 one participant were excluded from the analysis, because the reaction times exceeded 2000 ms on average (this criterion had previously been used to exclude data from analysis in Miyake et al., 2000). A number of trials (5.12%) were excluded from the analysis because they had very low (lower than 300 ms) or very high (higher than 2000 ms) reaction times.

Sometimes when the reaction time 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 meaningful units (s). No other manipulation of the data was done.

1.2.1. Within-trial effects

Accuracy data were 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). Reaction time data were also consistent with other studies in the Stroop literature, showing Stroop interference in the incongruent condi-

Table 1
Within-trial effects.

	Incongruent	Congruent	Neutral
Accuracy	0.922	0.996	0.988
Mean RT (s)	1.160	0.973	1.047

tion and Stroop facilitation in the congruent condition (Table 1). Since within-trial effects were very consistent with those found in previous studies they will not be treated in more detail here.

1.2.2. Between-trial effects

These are effects related to a particular sequence of trials. As it will be shown below, all three between-trial effects described by MacLeod (1991) were replicated. They will be referred to as Word–Color, Color–Word, and Word–Word, respectively. In addition, this study revealed a significant Color–Color effect. This additional finding was also a replication of a known effect, although previously reported in different contexts (Christie & Klein, 2001; Law, Pratt, & Abrams, 1995; MacDonald & Joordens, 2000). 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 et al. (2003). Table 2 presents examples of all these effects. Of the total number of trials, 17% were Word–Color, 16% were Color–Word, 11% were Word–Word, and 25% were Color–Color. These frequencies are unequal because they are proportional to their possibility of occurrence, as recommended by Christie and Klein (2008). We have only constrained the numbers of trials in the three conditions (incongruent, congruent, and neutral) to be equal.

In order to estimate the magnitudes of these effects, the data were submitted to a Linear Mixed Effects (LME) analysis. 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-sized effects. This analysis is also known to be robust to unbalanced designs (Garson, n.d.).

An LME analysis was run with *reaction time* (RT) as dependent variable, *condition* (incongruent, congruent, and neutral) and the four *between-trial effects* as fixed factors, and *subject* as a grouping factor. The results of this analysis are presented in Table 3. Between-trial effects are presented in the last four rows of Table 3.

The coefficients of the LME model are 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 repetition trials as compared to the non-repetition trials of 45 ms, when all of the other variables in the model are kept constant. In the following subsections each of these between-trial effects will be discussed.

Table 2
Examples of between-trial effects.

	Preceding trial	Current trial	Reaction time
Word–Color	RED (blue)	GREEN (red)	Increase
Color–Word	YELLOW (red)	RED (green)	Decrease
Word–Word	BLUE (green)	BLUE (red)	Decrease
Color–Color	RED (green)	BLUE (green)	Increase

Note. The color of the stimulus is presented in brackets.

Table 3
The results of the LME analysis for the first study.

	Value	Std. error	DF	*t*-Value	*p*-Value
Intercept	1.170	0.027	7079	44.004	0.000
Congruent	−0.191	0.008	7079	−25.256	0.000
Neutral	−0.133	0.008	7079	−16.550	0.000
Word–Color	0.045	0.009	7079	5.065	0.000
Color–Word	−0.049	0.010	7079	−5.149	0.000
Color–Color	0.029	0.008	7079	3.743	0.000
Word–Word	−0.011	0.011	7079	−1.038	0.299

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 with 45 ms ($t = 5.06$, $p = 0.000$). This effect has been replicated many times, and 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 with 49 ms ($t = -5.15$, $p = 0.000$). This effect was first reported by Effler (1977) and was replicated several times (Lowe, 1979; Neill, 1978; see also MacLeod (1991), for a review).

1.2.2.3. 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 with 29 ms ($t = 3.74$, $p = 0.000$). 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 and inhibition-of-return (Christie & Klein, 2001; Law et al., 1995; MacDonald & Joordens, 2000). It is important to mention here that other authors report a decrease in reaction time for Target–Target repetitions (Lowe, 1979; Tipper, 1985).

1.2.2.4. 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 with 11 ms. Although this decrease was not significant in our first study ($t = -1.04$, $p = 0.299$), other authors report a significant decrease in reaction time when the distractor is repeated (Christie & Klein, 2001; Effler, 1980; MacLeod, 1991; Rothermund et al., 2005).

1.3. Discussion of the first study

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 between-trial effects have been described. Although not as well known as the within-trial effects, these between-trial effects have been documented (e.g., Christie & Klein, 2001; Druey & Hubner, 2008; Law et al., 1995; MacDonald & Joordens, 2000; MacLeod, 1991). We have only replicated them and estimated their relative magnitudes while controlling for the within-trial effects and individual differences.

One aspect that was not addressed in this study was the interaction among within-trial and between-trial effects as well as the interactions of the between-trial effects with one another. These interactions could not be studied here because of the sparse nature of the data and the low number of trials in each case; they will be addressed in the second study presented in Section 4.

2. Theoretical accounts

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 and 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 approach is 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, 2007b, 2003).

The 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 et al. (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 and reported by other authors) 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).

The stimulus–response integration account (Hommel et al., 2004; Rothermund et al., 2005) is a variant of the episodic retrieval account. It postulates that the previous stimulus and its associated response form an integrated episode (even file) that is automatically retrieved when a new stimulus is processed. If the new stimulus requires a different response than the one in the event file, a conflict occurs causing delay in reaction time. Any change in response is a potential cause of such delay, for example, a change in response location. In our task, the location of the correct response is randomized, thus, changes in response location are very frequent. If this account is correct, changes in response location can cause increased reaction times in the Word–Color and Color–Color cases, as observed in our data, but they would also cause increased reaction times for the Color–Word and Word–Word cases, which would be at odds with our data.

Another inhibition-free account suggested by MacLeod et al. (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 et al., 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 et al. (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 which is able to account for all inhibition-of-return effects, it would also explain our between-trial effects.

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 reactivation. 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 time 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 naming the color of the current stimulus, thus, causing reaction time to increase. In summary, the selective inhibition account predicts 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), Lowe (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 to 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 information 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 and parsimonious way. In the next section, we will present a computational model that implements this account and predicts more detailed data about the interactions among within-trial and between-trial effects.

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 can be detected in behavior only when repetitions occur. In fact, this account makes a stronger prediction: traces of this between-trial suppression should *always* occur 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 time 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 color naming because its corresponding concept has been suppressed. This fact causes reduced interference (i.e., decrease in reaction time) in these trials. 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 color naming causing reaction time to decrease.

The repetition-suppression account somewhat resembles the inhibition account of inhibition-of-return (Posner & Cohen, 1984; Tipper et al., 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 is not an attention-based account explaining effects that occur in a cue-target paradigm. The semantic nature of repetition effects was also emphasized by Druey and Hubner (2008); they found similar effects for response repetitions and response category repetitions, which suggests that the core mechanism operates at the semantic level.

A similar control mechanism dealing with past information that has become irrelevant for the current context is mentioned elsewhere and is 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. Hubner and Druey (2006) propose a similar inhibition account for repetition effects in task-switching studies (see also Druey and Hubner (2008)). As in our account, the functional role of inhibition is to mitigate between-trial interference. As they put it, "each response is inhibited in order to prevent its accidental re-execution" (Druey & Hubner, 2008, p. 515).

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 and parsimonious way. The next self-imposed methodological constraint was to implement this theoretical account in a computational model that is able to make numerical predictions.

3. A computational model of repetition suppression

This model was developed with the aid of the latest version of the ACT-R[1] cognitive architecture (Anderson, 2007). ACT-R is a hybrid (symbolic and sub-symbolic) cognitive architecture used to develop cognitive models of various tasks. The architecture is composed by specialized modules (vision, memory, motor, etc.) coordinated by productions rules. The symbolic elements of the architecture (procedural rules and declarative memories) have associated sub-symbolic quantities (activations and utilities) that govern their availability and their manifestation in the model's behavior.

3.1. Modeling within-trial effects

Many models of the within-trial effects have been developed (Altmann & Davidson, 2001; 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; Frasse, 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 (link weights) between representations of stimulus features and memory elements associated with these features. This way of modeling the difference in automaticity is comparable to the one used by Cohen et al. (1990). When a stimulus is perceived, its *word* and *color* dimensions are represented in the *imaginal buffer* – a short-term storage structure used to maintain information that is important for the task at hand. For example, if the current stimulus is the word "blue" in *red* ink (incongruent condition), the two representations in the imaginal buffer are the word "blue" and the color *red*. The two representations 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 imaginal buffer is limited and is equally shared by the two representations. The amount of activation received by a memory element is a function of the amount of activation that spreads toward it and its strength of association

[1] Adaptive Control of Thought-Rational. The ACT-R6 modeling software is available at http://act-r.psy.cmu.edu/.

with the corresponding representation. 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 made and the concept of blueness is retrieved. At this point, if memory retrievals 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. This rule fires when the retrieved concept and the representation of the color feature of the stimulus do not match. The same mechanism of detecting a wrong retrieval is implemented in other models of the Stroop task (Altmann & Davidson, 2001; Lovett, 2005). There are potentially better ways to implement competition between memories at retrieval (e.g., Van Maanen & Van Rijn, 2007), but we have chosen to reuse one of the mechanisms from the previous models. Since memory retrievals take time, responses to incongruent stimuli take longer time than responses to neutral stimuli. In the congruent condition, both representations spread activation toward the same concept in memory, thus, increasing its activation and speeding up its retrieval. 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 in principle to other models of the Stroop task (Altmann & Davidson, 2001; Cohen et al., 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.

Since the focus of this paper is on the between-trial effects, our way of modeling within-trial effects is rather minimal and not original. Our model is similar in principle to other ACT-R models of the Stroop task (Altmann & Davidson, 2001; Lovett, 2005), only simpler, because we did not intend to model the semantic gradient or the effect of practice as in the cited models. We only aim to adequately model the relative differences in reaction time between incongruent, congruent, and neutral conditions. A trace of the model run for a congruent trial is presented in Table 4. A fit of our model to within-trial effects is presented in Section 4.2 and Fig. 2.

3.2. Modeling between-trial effects

Repetition avoidance has been extensively studied in cognitive control tasks such as the task of generating sequences of random numbers. 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 et al., 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 which is not usually intended (Logan, 1994). The adaptive function of a 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, p. 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 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. 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, called *FINST activation* instead of an all-or-none value. The FINST activation of a memory element is subtracted from its existing activation and, thus, it slows down its retrieval, instead of blocking it. This adjustment will be referred to as "the decaying-FINST mechanism".

FINST activation is computed based on the following formula:

$$fa = \frac{mfa}{2^{\frac{fd}{fhl}}}$$

where: fa is the FINST activation, mfa is the max FINST activation parameter, fd is the FINST delay – the time elapsed since the FINST was set, fhl is the FINST half-life parameter.

The max FINST activation parameter (starting value) and the FINST half-life parameter (controlling the decay rate) were set by fitting the model against the data from the first study. The values of these parameters were also tested in an ACT-R model of the "free recall" task.[2]

By adding a decaying-FINST mechanism to the model described in Section 3.1, the pattern of between-trial effects shown in Fig. 1 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*[3] (i.e., its activation has been dis-

[2] Unpublished but available at http://www.andrew.cmu.edu/user/ijuvina/Publications.htm.

[3] This term is also used by the author of the FINST concept (Pylyshyn, 2000).

Table 4
A simplified trace of the model run for a congruent stimulus.

Time (s)	Actions
0.000	A goal is set to name the color of the stimuli
0.000	A CONGRUENT stimulus appears in the middle of the screen
0.125	The stimulus has been perceived
0.130	The stimulus features have been represented in the Imaginal buffer and are ready to spread activation toward their corresponding memory elements (word = RED; color = red)
0.180	A retrieval of a color concept has been initiated
0.190	The concept REDNESS is retrieved because it has the highest activation. REDNESS has been activated by both the color and the word representations of the stimulus. After being retrieved, REDNESS is FINSTed, so that it negatively primes the next stimulus
0.240	The stimulus representation has been transferred in a short-term storage in order to guide the selection of the appropriate name prompt on the screen. This representation will also positively prime the next stimulus.
0.270	One of the prompt locations is attended (left or right)
0.385	The prompt's content has been perceived
0.435	A rule has fired signaling that the encoded color name (response) does not match the prompt. Thus, the other prompt is attended.
0.550	The content of the other prompt has been perceived
0.600	A rule has recognized that the response was found on the screen. A key press is initiated
0.820	The key press has been completed (END OF THE CONGRUENT TRIAL)

Note. The duration of each action may be different for different trials because of the intrinsic noise involved in most processes.

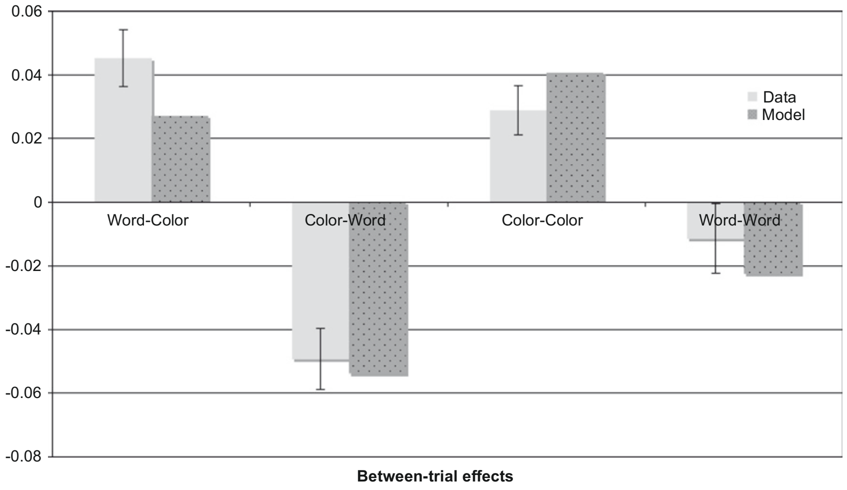


Fig. 1. LME coefficients for the between-trial effects in the first study for the model and the empirical data.

counted). When the same concept needs to be re-retrieved to name the color of the current stimulus, retrieval takes longer time than in control cases.

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 imaginal buffer and would interfere with color naming in the current trial. However, because it has 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 in 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 time 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 in 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. A trace of a model run for an incongruent trial illustrating the FINST mechanism is presented in Table 5.

This model produces a reasonably good fit to the empirical data (correlation = 0.957; mean deviation = 0.011 s) by implementing a relatively simple mechanism – decaying-FINST. This mechanism seems to implement well the main characteristics of a 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 Fig. 1, the model produces smaller magnitudes for the Word–Color effect and larger magnitudes for the Color–Color effect 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 after-effect. 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

Table 5
A simplified trace of the model run for an incongruent stimulus.

Time (s)	Actions
0.000	A goal is set to name the color of the stimuli
0.000	An INCONGRUENT stimulus appears in the middle of the screen
0.125	The stimulus has been perceived
0.130	The stimulus features have been represented in the Imaginal buffer and are ready to spread activation toward their corresponding memory elements (word = BLUE; color = red)
0.180	A retrieval of a color concept has been initiated
0.211	The concept BLUENESS is retrieved because it has the highest activation BLUENESS has been activated by the representation of the word feature of the stimulus REDNESS has been activated by the color representation, which has a smaller strength (weight) than the word representation, reflecting higher automaticity with words than with colors
0.311	The wrong retrieval is recognized and a new retrieval is initiated, with the color red as the retrieval cue
0.352	The concept REDNESS is retrieved because it matches the retrieval cue REDNESS was also previously retrieved and FINSTed. Thus, its activation has been discounted, causing its retrieval to take slightly longer time than in control cases (when it was not FINSTed). This kind of retrieval delay causes the between-trial effects
0.402	The stimulus representation has been transferred in a short-term storage in order to guide the selection of the appropriate name prompt on the screen
0.432	One of the prompt locations is attended (left or right)
0.547	The prompt's content has been perceived
0.597	A rule has fired signaling that the encoded color name (response) does not match the prompt. Thus, the other prompt is attended
0.712	The content of the other prompt has been perceived
0.762	A rule has recognized that the response was found on the screen. A key press is initiated
0.982	The key press has been completed (END OF THE INCONGRUENT TRIAL)

Note. The duration of each action may be different for different trials because of the intrinsic noise involved in most processes.

name the color of the current stimulus, the FINST is 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[4] and then FINSTed, these misfits would probably not occur.

4. Second study

As mentioned above, the first study did not address the interactions among within-trial and between-trial effects as well as the interactions of between-trial effects with one another. Accounting for all these interactions could provide a powerful argument for the verisimilitude of our theory and give us principled reasons for ruling out alternative accounts. An example of such interaction is the "switch" condition in which the Word–Color and the Color–Word effects occur simultaneously (Christie & Klein, 2008). Another aspect that was not fully addressed in the first study was the baseline for the between-trial effects and its dependence on the within-trial effects. For example, the Word–Color effect could occur in a sequence of incongruent trials or in a sequence of incongruent and neutral trials; it is important to use the proper baseline for each case and not aggregate across different cases. The objective of the second study was to study these interactions as a means to test our "suppression theory" and its associated computational model.

4.1. Method

4.1.1. Participants

Thirty-nine participants were recruited from Carnegie Mellon University's community via a website advertisement. Participant age ranged from 18 to 54 and averaged 25. There were 21 women and 18 men. They received a fixed amount of monetary compensation for their participation.

4.1.2. Design

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. Every participant received 240 trials, 80 trials for each condition. These changes were made in order to ensure that each participant encountered a sufficient number of repetitions of each type.

4.1.3. Apparatus and procedure

Apparatus and procedure that are same as those used for the first study were used in the second study. The only change was the set size of neutral stimuli, which was decreased from 53 to 10. The following is a list of all neutral stimuli used in this study: Action, Case, Fact, Form, General, Matter, Number, Part, Present, and System.

4.2. Results

The data of one participant were excluded from the analysis. This participant seemed to have misunderstood the task instructions. Given that he had an accuracy of zero (minimum) for the incongruent condition and one (maximum) for the congruent condition, we inferred that he reacted as if responding to the word dimension instead of the color dimension of the stimulus. As in the first study, trials were excluded from analysis if reaction time was lower than 0.3 s and higher than 2 s (5.1% of all trials).

In order to ensure that each effect was compared with the proper baseline, the data were coded as follows:

- Between-trial effects and their combinations were identified. For example, the Word–Color effect can occur alone or in combination with each of the other effects, thus, giving rise to the following cases: Word–Color, Word–Color & Color–Word, Word–Color & Color–Color, and Word–Color & Word–Word. For convenience, they will be indicated by their initials: WC, WC&CW, WC&CC, WC&WW.

[4] The ACT-R architecture only allows one retrieval at a time; changing this basic architectural constraint would only produce a minor improvement in the model fit.

Table 6
All between-trial effects and their interactions.

No.	Repetition ID	Previous trial type	Current trial type	Previous stimulus	Current stimulus
1	WC-inc-inc	Incongruent	Incongruent	YELLOW (red)	GREEN (yellow)
2	WC-inc-neu	Incongruent	Neutral	BLUE (green)	DESK (blue)
3	CW-inc-inc	Incongruent	Incongruent	BLUE (yellow)	YELLOW (red)
4	CW-neu-inc	Neutral	Incongruent	SIDE (blue)	BLUE (green)
5	CC-inc-inc	Incongruent	Incongruent	GREEN (yellow)	BLUE (yellow)
6	CC-inc-neu	Incongruent	Neutral	BLUE (yellow)	PART (yellow)
7	CC-neu-inc	Neutral	Incongruent	TABLE (green)	BLUE (green)
8	CC-neu-neu	Neutral	Neutral	SCREEN (yellow)	ACTION (yellow)
9	WW-inc-inc	Incongruent	Incongruent	BLUE (green)	BLUE (yellow)
10	WW-neu-neu	Neutral	Neutral	ORDER (green)	ORDER (blue)
11	WC-CW-inc-inc	Incongruent	Incongruent	YELLOW (red)	RED (yellow)
12	WC-CC-cgr-inc	Congruent	Incongruent	RED (red)	BLUE (red)
13	WC-CC-cgr-neu	Congruent	Neutral	GREEN (green)	LOOK (green)
14	WC-WW-inc-cgr	Incongruent	Congruent	BLUE (yellow)	BLUE (blue)
15	CW-CC-inc-cgr	Incongruent	Congruent	GREEN (blue)	BLUE (blue)
16	CW-CC-neu-cgr	Neutral	Congruent	TABLE (green)	GREEN (green)
17	CW-WW-cgr-inc	Congruent	Incongruent	BLUE (blue)	BLUE (green)
18	CC-WW-inc-inc	Incongruent	Incongruent	YELLOW (blue)	YELLOW (blue)
19	CC-WW-neu-neu	Neutral	Neutral	ACTION (red)	ACTION (red)
20	REP-cgr-cgr	Congruent	Congruent	GREEN (green)	GREEN (green)

Note. The repetition ID is an acronym in which W = Word, C = Color, inc = Incongruent, cgr = Congruent, neu = Neutral, and REP = Repeat.

- For each between-trial effect identified at the previous step, all combinations between the within-trial conditions of the previous trial and the current trial were considered. For example, the WC effect occurs when the previous trial is incongruent and the current trial is either incongruent or neutral; the WC&CC effect occurs when the previous trial is congruent and the current trial is either incongruent or neutral; and so on (see Table 6 for a list of all the possible combinations).
- The baseline for a particular effect was identified as the subset of trials without any repetition and with the same combination of conditions in the previous and the current trial. For example, the baseline for the CW&CC-incongruent-congruent effect is the subset of congruent trials preceded by incongruent trials without any kind of repetition.

Each of the 20 effects identified above was submitted to a mixed effects analysis with reaction time as dependent variable, the particular effect under consideration as a fixed factor and participant (subject) as a grouping factor. Table 7 shows the results of these analyses. The table shows the magnitudes of these effects and the tests of their statistical significance. However, our objective is not to test whether these effects are significantly different than zero. We are interested to see to what extent these data match the predictions of our theory; sometimes the theory predicts a zero effect.

Table 7
Results of the LME analysis for each between-trial effect and interaction.

	Estimate	Std. error	DF	*t*-Value	*p*-Value
WC-inc-inc	0.000	0.032	250	−0.009	0.993
WC-inc-neu	0.025	0.021	685	1.178	0.239
CW-inc-inc	−0.046	0.031	234	−1.452	0.148
CW-neu-inc	−0.053	0.022	636	−2.384	0.017
CC-inc-inc	0.019	0.035	229	0.544	0.587
CC-inc-neu	0.018	0.022	662	0.842	0.400
CC-neu-inc	−0.002	0.024	614	−0.086	0.932
CC-neu-neu	0.002	0.019	813	0.108	0.914
WW-inc-inc	−0.072	0.032	259	−2.252	0.025
WW-neu-neu	0.006	0.027	668	0.219	0.827
WC-CW-inc-inc	−0.034	0.045	161	−0.763	0.447
WC-CC-cgr-inc	0.094	0.027	553	3.437	0.001
WC-CC-cgr-neu	0.048	0.018	879	2.625	0.009
WC-WW-inc-cgr	0.040	0.018	693	2.208	0.028
CW-CC-inc-cgr	0.035	0.021	621	1.687	0.092
CW-CC-neu-cgr	0.007	0.018	882	0.396	0.692
CW-WW-cgr-inc	−0.055	0.023	649	−2.430	0.015
CC-WW-inc-inc	−0.045	0.042	171	−1.065	0.288
CC-WW-neu-neu	0.021	0.049	608	0.428	0.669
REP-cgr-cgr	−0.027	0.016	950	−1.663	0.097

Although we do not test the null hypothesis for each particular effect, the question remains as to how reliable each measurement is. For each of the 20 conditions, there were on average approximately five observations per participant, and there were 38 participants. Given that each LME analysis has one fixed factor and one grouping factor (participant), there are enough observations for a reliable estimation of the magnitude of each effect. As shown by Maas and Hox (2005), the number of groups (participants in our case) is more important for parameter estimation than the number of observations in each group.

In order to test the predictions of our suppression theory, the cognitive model described in the previous section was first fit to the aggregated data for the three within-trial conditions (incongruent, congruent, and neutral). The model was run with 50 simulated subjects and 150 trials per subject. Fig. 2 shows the fit of the suppression model to the reaction time for each condition (correlation = 0.999, mean deviation = 0.005 s).

Next, the model was run with 50 simulated subjects and 500 trials in order to gather enough data to generate predictions for the between-trial effects. The same coding procedure and analysis were applied to the simulated data as to the real data. Fig. 3 shows the model predictions plotted against the actual data for all the between-trial effects.

One way to qualify the fit between the model predictions and the empirical data (correlation = 0.69, mean deviation = 0.026 s) is to compare our suppression theory with the competing theory, that is, the no-suppression theory. Our computational model can easily be modified to generate predictions for the no-suppression theory by removing the repetition-suppression mechanism. The no-suppression theory would postulate that representations of the previous stimuli maintain traces of activation and, thus, are more readily available in case of stimulus repetitions. Fig. 4 shows the predictions of the no-suppression model plotted against the actual data. The fit of this model to the data is significantly worse than the fit of the suppression model (correlation = −0.55; mean deviation = 0.043 s).

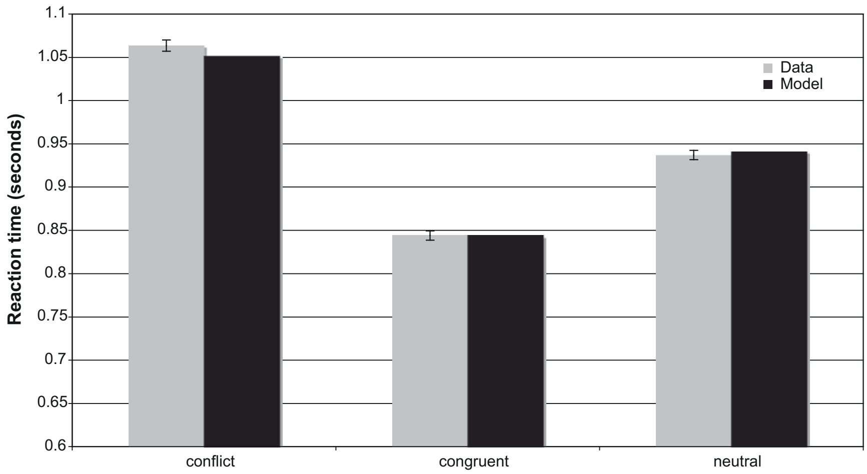


Fig. 2. Within-trial effects as shown in the empirical data and in the model.

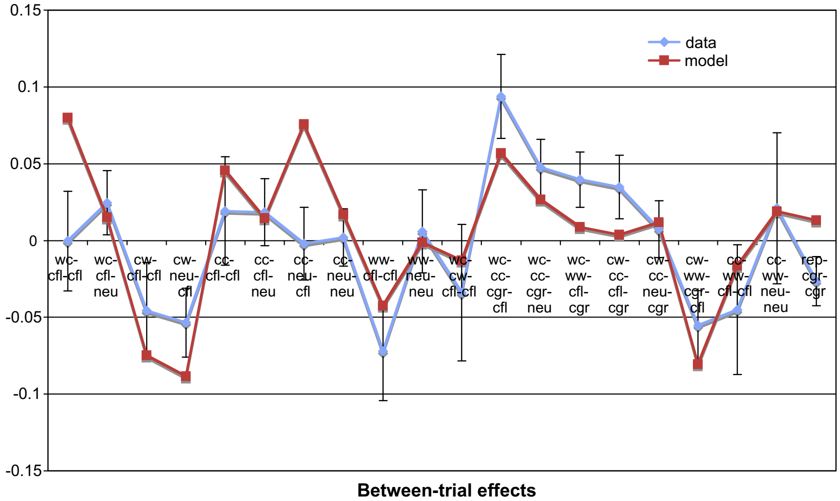


Fig. 3. The pattern of between-trial effects as estimated from empirical data and simulated by the “suppression” model. On the horizontal axis the between-trial effects and their interactions are deployed. The vertical axis shows the magnitudes of these effects: positive values indicate an increase and negative values indicate a decrease in reaction time as compared to baseline trials. The error bars indicate standard errors of the means.

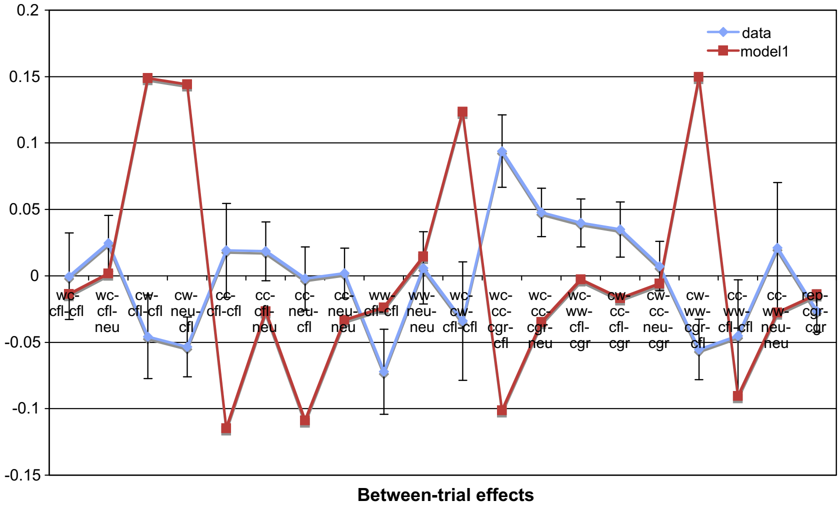


Fig. 4. The pattern of between-trial effects as estimated from empirical data and simulated by the “no-suppression” model. On the horizontal axis the between-trial effects and their interactions are deployed. The vertical axis shows the magnitudes of these effects: positive values indicate an increase and negative values indicate a decrease in reaction time as compared to the baseline trials. The error bars indicate standard errors of the means.

4.3. Discussion of the second study

The second study provides an empirical test of a substantive theory. A computational model is used to transform the qualitative statements of our substantive theory in numerical predictions. We do not test whether the difference between a particular condition and its baseline is significantly different than zero. This null hypothesis refutation would not help much in corroborating a particular theory, because there are potentially many theories that could explain a particular finding. Instead, we test how well our suppression theory predicts the whole set of conditions. The data are compared with the numerical predictions of our model. A theory increases its verisimilitude when it is able to make accurate and risky predictions (Meehl, 1990).

With regard to accuracy, we can characterize the overall pattern of predictions by looking at the correlation between the 20 data points with their corresponding predictions and the mean deviation of the data points from their corresponding predictions. The correlation is significantly positive ($r = 0.69$, $t = 4.05$, $df = 18$, $p = 0.0007$) but it is clearly not ideal, as also indicated by the mean deviation 0.026 s. We can also look at how accurate each point prediction was. Each point prediction can be characterized as a hit, a near-miss or a far-miss depending on whether it falls within a standard error from the mean of the data (hit), it has the same sign as the data (near-miss), or it has an opposite sign as compared to the data (far-miss). There are 11 hits, 8 near-misses, and 1 far-miss. Of 20 predictions, one was totally off. This is the case in which a congruent stimulus repeats as such, for example, when the word "red" colored in *red* re-occurs in the next trial. Our suppression theory predicts an increase in reaction time in this case, whereas the data show a decrease in reaction time (an effect known in literature as *repetition priming*). In agreement with Klein (2004), we would argue that this case is unique, in the sense that a different type of selection strategy is involved, that is, a heuristic of the kind "if no stimulus change, then repeat the previous response".

With regard to how risky our prediction is, we can ask, absent the theory, what is the probability of getting this particular combination of values for our data points? It is hard to numerically characterize this probability because we do not know the *a priori* range of values for each data point. Fortunately, we can obtain a numerical characterization of the state in which the theory is absent by removing the suppression mechanism from our model. Now the altered model has only a positive priming mechanism to influence the between-trial interference. Representations from previous trials maintain their activations for a while, and, thus, they are more readily available in case of repetitions. This model instantiates the no-suppression theory. The predictions of the no-suppression theory are much worse than the predictions of the suppression theory. The correlation is significantly negative ($r = -0.55$, $t = -2.78$, $df = 18$, $p = 0.01$) and the mean deviation is 0.043 s. Of the 20 point predictions, 5 are hits, 8 are near-misses, and 7 are far-misses. These results seem to corroborate more the suppression theory than the no-suppression theory. An important aspect to be noticed is that the no-suppression theory is able to get a considerable number of hits and near-misses. This illustrates the danger of studying only a limited set of effects.

5. General discussion and conclusion

Criticism has recently been expressed with reference to psychological theories that postulate suppression (cognitive inhibition) as an explanatory mechanism for the observed behavioral effects (e.g., MacLeod et al., 2003). Since we agreed that many of these points of criticism were justified, we 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", and we did not make strong assumptions with regard to the exact implementation of cognitive suppression in the brain. 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 interference independently of 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 first study and estimated their magnitudes while controlling for within-trial effects, as suggested by MacLeod (1991). To our knowledge, an integrated account for these effects has not been proposed previously, although a similar account has been proposed in the task-switching paradigm (Hubner & Druey, 2006). We have shown that a repetition-suppression account can explain all these effects and we have implemented this account in a computational cognitive model.

The second study makes a stronger point in favor of the repetition-suppression account by analyzing all the possible combinations among between-trial and within-trial effects. Christie and Klein (2008) argued for the necessity to analyze the full set of effects (in proportion to their possibility of occurrence) in order to select the theoretical account that best explains the whole set and rule out a multitude of possible accounts. By considering the full set of conditions and the proper control (baseline) for each condition, we ensured that our theory complies with the *ceteris paribus* clause, that is, it controls for all the factors that are relevant to the theorized phenomenon (Meehl, 1990).

Although not all the predictions were confirmed by the data, our suppression theory predicts a rich dataset much better than competing accounts. According to Meehl (1990), a theory deserves to be defended and amended when it is able to make successful or near-miss predictions of low prior probability, that is, accurate and risky predictions. A risky prediction is about a reality that would be highly improbable, if the theory were not true. When we changed our model by removing the repetition-suppression mechanism, predictions were negatively correlated with the data. Thus, the empirical data presented here would be improbable in the absence of an inhibitory control mechanism dedicated to between-trial interference.

Although the repetition-suppression account has been shown to explain the presented data better than the no-suppression account, we do not have sufficient grounds to generalize the outcome of this comparative analysis beyond the task and the dataset presented here. We do not imply that the theoretical accounts that are shown to fail here are fundamentally invalid. Neither do we imply that the repetition-suppression account presented here should explain all the data in the negative priming and inhibition-of-return literatures. We do acknowledge that the sequence effects in the modified Stroop paradigm presented here and the hypothesized inhibitory control mechanisms might not be reproducible in different tasks under different circumstances (but, see Druey and Hubner (2008), for a very similar account in the task-switching field).

It is characteristic of inhibitory control mechanisms to be employed only in specific circumstances (Lowe, 1985; Neill & Westberry, 1987; 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. We also admit that there might be other theoretical accounts that would be able to explain the data equally well. For all these reasons, we make the task software, the data, and the model publicly available (http://www.andrew.cmu.edu/user/ijuvina/Publications.htm) and invite researchers to replicate our findings, ideally in different settings, and eventually challenge or corroborate our theoretical account.

The computational mechanism that we have used to model cognitive 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, which allows FINSTed memories to be retrieved, only delaying their retrieval. The classical FINST in ACT-R was a tagging mechanism similar in principle with the tagging postulated by the episodic retrieval account (Neill, 1997); instead of *do-not-respond* tags attached to the recently ignored distractors, we would have *do-not-retrieve* tags attached to recently retrieved memories. By adding a sub-symbolic quantity (FINST activation), we made this mechanism more flexible and more in line with the intuitive concept of cognitive suppression (MacLeod, 2007b). This mechanism needs more research in order to be fully validated.

In conclusion, we have attempted to demonstrate that a repetition-suppression mechanism is a viable theoretical account for explaining a large range of between-trial effects in an integrated and parsimonious way. A research program focused on cognitive inhibition as a means of interference control seems worth pursuing.

Acknowledgements

This research was supported by Office of Naval Research grant N00014-08-10541 and by Air Force Office of Scientific Research grant FA95500710359. The authors thank Daniel Dickison for his contribution to the research reported here.

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, NY: Oxford University Press.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *Neuroscientist, 13*(3), 214–228.

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.

Christie, J., & Klein, R. M. (2001). Negative priming for spatial location? *Canadian Journal of Experimental Psychology, 55*(1), 24–38.

Christie, J. J., & Klein, R. M. (2008). On finding negative priming from distractors. *Psychonomic Bulletin and Review, 15*(4), 866–873.

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.

Druey, M. D., & Hubner, R. (2008). Response inhibition under task switching: Its strength depends on the amount of task irrelevant response activation. *Psychological Research, 72*, 515–527.

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.

Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychologica, 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.

Garson, G. D. (n.d.). *Linear mixed models. In Statnotes: Topics in multivariate analysis*. http://www2.chass.ncsu.edu/garson/pa765/statnote.htm Retrieved 05.21.2008.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, A. Towse, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York, NY: Oxford University Press.

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. *Journal of Cognitive Neuroscience, 18*(1), 22–32.

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.

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.

Hubner, R., & Druey, M. D. (2006). Response execution, selection, or activation: What is sufficient for response-related repetition effects under task shifting? *Psychological Research, 70*, 245–261.

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* (pp. 545–559). Cambridge, MA: MIT Press.

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research, 56*(3), 130–135.

Law, M. B., Pratt, J., & Abrams, R. A. (1995). Color-based inhibition of return. *Perception and Psychophysics, 57*(3), 402–408.

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* (pp. 189–239). San Diego, CA: 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.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86–92.

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, NY: 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, Dodd, M. D, Sheard, E. D, Wilson, D. E, & Bibi, U. (2003). In opposition to inhibition. In H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 43, pp. 163–214). Elsevier Science.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*(2), 108–141.

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.

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 and performance* (Vol. 10, 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 and 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.

Rothermund, K., Wentura, D., & De Houwer, J. (2005). Retrieval of incidental stimulus–response associations as a source of negative priming. *Journal of Experimental Psychology: Learning Memory and Cognition, 31*, 482–495.

Shallice, T. (2004). The fractionation of supervisory control. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences III* (pp. 943–956). Cambridge, MA: MIT Press.

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., Jerreat, L. M., & Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 478–499.

Van Maanen, L., & Van Rijn, H. (2007). An accumulator model of semantic interference. *Cognitive Systems Research, 8*(3), 174–181.

Vivas, A. B., & Fuentes, L. J. (2001). Stroop interference is affected in inhibition of return. *Psychonomic Bulletin and Review, 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.