The Picture-Word Interference Effect is a Stroop Effect After All

Leendert van Maanen (leendert@ai.rug.nl)
Hedderik van Rijn (D.H.van.Rijn@rug.nl)
Artificial Intelligence, University of Groningen
P.O. Box 407
9700 AK Groningen, The Netherlands

Abstract

The idea that picture-word interference (PWI) and the Stroop effect are two manifestations of the same process has been widely accepted. However, recently Dell’Acqua and colleagues (2007) have questioned the identical nature of the Stroop effect and picture-word interference, based on the observation that the loci of both effects in the mental processing stream differ. In this paper, we will present a dynamic computational model of semantic interference for both the Stroop effect and PWI. The model is based on competition between possible responses in both Stroop and PWI tasks. The differences between both tasks are solely explained by differences in processing speed of the stimuli. This way, our model does justice to the different loci of effects observed by Dell’Acqua et al, while at the same time supporting the long-held view that both Stroop and PWI are indeed similar in the underlying processes.

Keywords: Stroop effect; picture-word interference; cognitive model; ACT-R; RACE.

Introduction

Through the years, the idea that the picture-word interference (PWI) effect and the Stroop effect are two manifestations of the same process has been widely accepted. For instance, MacLeod’s influential review on the Stroop effect (MacLeod, 1991) also discusses picture-word interference, and even lists the picture-word task in his list of “eighteen major empirical results that must be explained by any successful account of the Stroop effect” (MacLeod, 1991, Appendix B). In previous work, we have also taken for granted that picture-word interference is just an instance of the Stroop effect. Van Maanen and Van Rijn (2007b) described a computational model of picture-word interference, and when discussing related models, we actually focused on models of the Stroop task.

The wide acceptance that PWI and the Stroop effect are one and the same effect probably stems from the (behavioral) similarities between the two. First, the prototypical paradigm for both tasks is that participants are requested to respond to one stimulus, while being presented with a second stimulus. In the case of the Stroop task, the to-respond-to stimulus (henceforth referred to as the target) is the color a word is printed in, and the to-be-ignored stimulus (the distractor) is a written word. Similarly, in picture-word interference tasks, the target is a picture, and the distractor is a word. Over the years, a wide range of variations of this paradigm have been presented (for an overview, see MacLeod, 1991).

Second, both tasks display similar latency distributions. The typical Stroop effect is that it is harder to name the ink color if it spells a different color word than if it does not spell a color word (e.g., M. O. Glaser & Glaser, 1982). That is, participants are slower in naming the color of a written word if that word is another color (e.g., saying “red” to a stimulus that is the word “green” written in red), as compared to naming a colored spot alone (that is, without the distractor “green”). This effect is usually referred to as Stroop interference. In picture-word interference tasks, a similar interference pattern can be observed (e.g., W. R. Glaser & Düngehoff, 1984). Picture-naming latencies are increased for pictures that are accompanied by incongruent words relative to latencies of these pictures in isolation. For example, naming a picture of a cat together with the word “dog” takes longer than naming that same picture without the word.

However, when target and distractor relate to the same concept, participants typically respond faster than to the target alone, an effect called facilitation. For example in the Stroop task, the word “red” written in red ink elicits a faster color naming response than a red spot. The PWI analog is that a picture of a cat accompanied by the word “cat” is responded to faster than a picture of a cat alone.

Besides the interference and facilitation effects and their prototypical design, PWI and Stroop share many other characteristics. For instance, both interference and facilitation typically disappear if the task is changed to naming the written word. This effect is sometimes referred to as Stroop asynchrony (W. R. Glaser & Düngehoff, 1984). Given all these similarities1, it has regularly been assumed that both tasks tap a similar set of underlying processes.

Difference between Stroop and PWI

Recently it has been suggested that the Stroop effect and picture-word interference may not be caused by the same process (Dell’Acqua et al., 2007). In particular, it seems that the loci of the two effects in the mental processing stream differ. Many researchers now agree that the locus of the Stroop effect is on the level of response selection (e.g., Fagot & Pashler, 1992; Kuipers, La Heij, & Costa; MacLeod, 1991; Roelofs, 2003). That is, the Stroop interference comes about because an incorrect response possibility that is triggered by the distractor, interferes with the correct response that is triggered by the target stimulus.

Fagot and Pashler (1992, Experiment 7) studied whether the Stroop effect would persist in a psychological refractory

---

1 For a more extensive comparison between PWI and Stroop, see (Dell’Acqua, Job, Peressotti, & Pascali, 2007; MacLeod, 1991).
period (PRP) design. They hypothesized that if the locus of the Stroop effect would be early in the processing stream, the Stroop-critical processing would be parallel to the secondary task, and the Stroop effect would disappear. In their experiment, participants were instructed to respond to an auditory stimulus first, and then, after a SOA (Stimulus Onset Asynchrony) interval the Stroop stimulus was presented. Fagot and Pashler did not find an interaction between congruency condition and SOA, which can be interpreted as evidence that the locus of the Stroop effect (operationalized as the latency difference between the congruent and the incongruent Stroop stimulus) could not be early in the processing stream. This finding, they argued, is in line with a response selection account (that is, late in the stream of mental processing) of the locus of the Stroop effect.

In a similar experiment, Dell’Acqua and colleagues (2007) tried to answer the same question for the PWI effect. They performed the same experiment as Fagot and Pashler (1992 Experiment 7), with the only difference that the primary task was not Stroop naming, but instead picture-naming. Interestingly and surprisingly, they did find an interaction between congruency condition and SOA, indicating that the locus of picture-word interference is early, possibly during perceptual encoding of the stimulus. Based on this, Dell’Acqua and colleagues state that their findings “are obviously incompatible with the often reiterated principle that the PWI effect comes about for limitations of the cognitive system that are analogous to those causing the Stroop effect” (Dell’Acqua et al., 2007, p. 720). They conclude that their analysis “favor[s] an interpretation of the present findings that points to the functional dissociation of the sources of Stroop and PWI effects” (Dell’Acqua et al., 2007, p. 722).

In the remainder of this paper, we will argue that this conclusion might be premature. Using computational cognitive models, we will demonstrate that an apparent difference in locus can be achieved using the same underlying process. To achieve this, we will present a cognitive model that simulates both the Stroop task and PWI. The only difference between the two simulations of the Stroop task and the PWI task is a single parameter value, and we will argue that this parameter is sufficient to account for the perceived shift in locus. Since the model relies on the ACT-R cognitive architecture (Anderson, 2007) as well as a recent extension that models the memory retrieval process (RACE, Van Maanen & Van Rijn, 2007b), we will first discuss the critical components of both, before explaining the implementation of the combined Stroop/PWI model.

**ACT-R**

ACT-R (Anderson, 2007) can be perceived as a theory of cognitive information processing. That is, the theory predicts how humans handle pieces of information in various (cognitive) tasks. The theory is characterized by a central production system that communicates with different modules through a set of buffers. The central production system is a set of condition-action pairs (called production rules) that specify which actions to perform, given the content of the buffers. Actions are instructions to the modules and are performed by manipulating buffer contents. Whereas different modules can run in parallel, the central production system constitutes a serial bottleneck.

An example of a module is declarative memory. Declarative memory is represented by a module that can be queried via a retrieval buffer. Declarative facts, that are stored in declarative memory, are called chunks. Given that there is just one retrieval buffer that can contain only a single chunk, only one chunk can be retrieved from memory at a time. Each chunk is associated with a quantity called activation that indicates the likelihood that a particular chunk is needed at this moment in the current task. In ACT-R, the level of activation of the chunks is determined by a set of equations that take the history of usage of a chunk into account (Anderson & Schooler, 1991). However, as this account is not detailed enough for current purposes, we will use a more elaborate model of memory retrievals (Van Maanen & Van Rijn, 2007b) to compute the activation values. This model, termed Retrieval by ACcumulating Evidence (RACE) will be explained in the next section.

**RACE**

RACE stems from an effort to better understand the process of memory retrievals within a cognitive architecture (Van Maanen & Van Rijn, 2005, 2006, 2007a, 2007b). By studying cognitive phenomena that were difficult to explain with the default ACT-R retrieval mechanism, we developed a model that predicts what happens during the actual retrieval process.

RACE is driven by two key assumptions:

1. The activation of one chunk is determined (in part) by the activation of other chunks.
2. The activation of one chunk relative to the activation of other chunks determines the likelihood that it will be needed.

The first assumption represents the fact that the relevance of information is context-dependent. This is for instance reflected in priming studies, in which a related prime decreases the response latency on a target stimulus (e.g., Neely, 1976, 1977). We operationalized this by adopting a spreading activation strategy (Collins & Loftus, 1975), in which increased activation of one chunk increased the activation of related chunks. This way, context, as represented by increased activation of associated chunks, has an effect on the likelihood that a chunk will be needed.

\[
C_i(t) = \alpha C_i(t-1) + \beta \sum_{j \in k} C_j(t-1)S_{ij}
\]

Equation 1

Equation 1 incorporates the first assumption. The equation reflects how the activation of a chunk \(C_i(t)\) accumulates during a retrieval process. The activation dynamics depend on the previous activation of a chunk (at \(t-1\), as well as on spreading activation from other chunks.
\( S_j \). \( \alpha \) and \( \beta \) are scaling parameters that determine the relative contributions of both components to the new activation value. Because \( \alpha \) is set to a value in the range \((0,1)\), it can be regarded as temporal decay of activation. The accumulated activation thus decays away after a retrieval has been attempted.

The second assumption reflects the insight that if multiple memory representations are relevant, responding becomes more difficult (Luce, 1986). Following ACT-R, the activation of a chunk determines the likelihood that it will be needed in the near future. However, given that these likelihoods are not independent, the actual need likelihood of a chunk is relative to the activation of other chunks. In RACE, this is accounted for by taking the ratio of activation of one chunk (chunk \( i \) in Equation 2) with other relevant chunks (chunks \( j \) in Equation 2). In the current model, these chunks (\( j \)) are the other chunks that match the retrieval criteria specified in a retrieval request.

\[
\sum_j e^{A_j} \geq \theta
\]

Equation 2

If the ratio specified in Equation 2 crosses a threshold (\( \theta \), the retrieval ratio), the relative activation of the chunk in the denominator (chunk \( i \)) is high enough to produce a memory retrieval of that chunk. The duration of the retrieval process constitutes the interval between the onset of the retrieval process and the moment at which the activation of one of the chunks crosses the retrieval ratio.

Using Equations 1 and 2, we have already provided quantitative fits for variants of both picture-word interference and the Stroop task (Van Maanen & Van Rijn, 2007a, 2007b). The previous model of PWI had a focus on the effect of SOA differences (Van Maanen & Van Rijn, 2007b); the Stroop model fitted a data set in which the distractors were presented subliminally (Van Maanen & Van Rijn, 2007a).

A cognitive model of both Stroop and PWI

The model presented here, in line with the proposal of Dell’Acqua et al (2007), assumes that both the Stroop task and PWI consist of three main stages, the perceptual encoding stage, the response selection stage, and the response execution stage. The interference effects observed in the Stroop task and PWI are accounted for by competition effects in the RACE declarative memory retrieval mechanism: If multiple chunks are eligible for retrieval from declarative memory, they compete for retrieval as outlined above in the section describing RACE. This causes interference if multiple chunks contribute to the likelihoods of different chunks. Besides retrieval competition, interference may also be caused by another factor. In the incongruent conditions, the model sometimes selects an incorrect response (that is, retrieves a chunk reflecting the picture, when the word has to be named), and has to retry to select the correct response (as is illustrated by Figure 1). This also leads to increased latencies.

Facilitation occurs if multiple chunks contribute to the likelihood that one chunk is the requested one, and speed up the retrieval time of that chunk. In this case, the activation increase of that chunk due to spreading activation from other chunks exceeds the competitive effect from the activation of the other chunks themselves.

For ease of explanation, we will focus in this section on the Stroop task only. However, given our hypothesis that the Stroop and PWI tasks are the same process, one can substitute all references to the color dimension of the stimulus with references to the picture to get a description of how the model performs a typical picture-word interference trial.

The Stroop task examined here has six different conditions. Two task conditions: color naming and word reading, and three different congruency conditions. The first congruency condition is one in which both the word and the ink color refer to the same color concept (e.g., the word “red” written in red ink, congruent condition). In the second condition, the word and the ink color refer to different color
concepts (“red” in green ink, incongruent condition). The third congruency condition is a neutral condition (actually, two neutral conditions, which can be considered identical for current purposes: one for color naming in which a series of x-s is written in colored ink, and one for word reading in which a color word is written in black). This follows a prototypical experimental setup for the Stroop task (M. O. Glaser & Glaser, 1982), as well as for picture-word interference (W. R. Glaser & Düngelhoff, 1984).

The model performs a typical word reading trial as follows: In the perceptual encoding stage, the stimulus features are transferred to the perceptual buffer and at the same time, the relevant chunks are being retrieved from memory. This reflects the observation that perceptual cues can be processed before they enter awareness (Marcel, 1983; Merikle, Smilek, & Eastwood, 2001). Because the model’s task is to read the word (and ignore the color), the model retrieves a chunk that encodes the syntactic properties of the stimulus (that is, the lemma chunk, Levelt, Roelofs, & Meyer, 1999). For word reading it is not necessary to retrieve a concept chunk, since words can be pronounced without active access to the word’s meaning.

Although perceptual cues can influence processing before they enter awareness (e.g., Marcel, 1983), it seems that processing of the complete visual stimulus cannot commence until one is aware of those cues (Treisman & Gelade, 1980). Therefore, the second stage is only initiated if the model is aware of the stimulus and if it has retrieved a chunk that is associated with the desired stimulus dimension. In the case of word reading, this means that the model should have retrieved a lemma chunk. However, since a lemma is also the desired chunk type in the response selection stage, the model directly continues with the response execution stage. At this time the model retrieves an associated motor program (that is, a lexeme, Kempen & Huijbers, 1983) storing information on how the lemma is articulated and the response is uttered.

If the task is not word reading but instead color naming, the process is a bit more complex (Figure 1). Again, during the perceptual encoding stage, the stimulus features are transferred to the perceptual buffer. Then, a chunk is requested from declarative memory that represents conceptual information on the color of the stimulus. During the response selection stage, the model tries to retrieve a lemma that is associated with the just retrieved concept. Because multiple chunks may be retrieved and accumulate activation, lemma retrieval could result in a retrieval of an incorrect lemma. That is, the model did not retrieve the lemma that represents the syntactic properties of the stimulus color, but instead retrieved the lemma describing the word. To be able to deal with this, a perceptual check is performed (c.f., Van Rijn & Anderson, 2004). If the correct lemma was retrieved, the model continues with the response execution. If not, the model tries to retrieve another lemma.

Latency differences between various conditions come from two sources. The first source is the perceptual check that ensures that a lemma describing the relevant stimulus feature is retrieved. The second one is the competition caused by the RACE mechanism. All retrieval times (for concept chunks, lemma chunks, and lexeme chunks) depend on the activation of other chunks. If multiple chunks spread activation to different chunks, as in the incongruent condition, retrieval times are increased because of the competition between those chunks. If multiple chunks spread activation to the same chunk, as in the congruent condition, retrieval times are decreased because of facilitation. The model thus assumes that interference in declarative memory is not localized, but is distributed over the various processing stages (McClelland, 1979; Wheeldon, 1999).

Model fits of experimental data
We ran the model for 36 simulation runs per condition, similar to the number of trials in typical Stroop or PWI experiments (e.g., M. O. Glaser & Glaser, 1982 and W. R. Glaser & Düngelhoff, 1984 used 36 trials per condition for...

When observing the typical temporal activity traces in Discussion 16 and 18 participants, respectively). Figure 2a presents the fit of the Stroop model for both the color naming task and the word reading task. The model captures both Stroop interference in the incongruent color naming condition, and facilitation in the congruent color naming condition (RMSE=32.6, \( R^2=.93 \)). In addition, the Stroop asynchrony between color naming and word reading can be observed. Given our hypothesis that the Stroop effect and picture-word interference are manifestations of the same process, the challenge is to demonstrate that both effects can be fitted with the same model.

The only difference in the two simulations that yield the data presented in Figure 2 is a single parameter that controls the speed of processing of the stimulus and that is adjusted to reflect the differences in the two tasks: The processing speed of the combined picture-word stimulus was set slower than the processing speed of the combined color-word stimulus. This reflects the idea that line drawings are more complex visual stimuli than colors, which can be considered as stimuli consisting of only one feature (e.g., Fleetwood & Byrne, 2006; Rossion & Pourtois, 2004).

Figure 2b presents the model fit on picture-word interference reaction time data. Similar to the fit of the Stroop task, all the behavioral patterns (interference, facilitation, and asynchrony) are captured by the model (RMSE=56.8, \( R^2=.84 \)).

It could be argued that this model has a significant number of degrees of freedom, since a number of other strategies could also have been implemented. However, we believe that this model is a convincing theory of the process of both tasks. The model is constrained by integrating it in a cognitive architecture that has proven its merits (ACT-R, Anderson, 2007), and implemented the tasks in a relatively straightforward way (c.f., Newell and Simon’s “listen to the architecture” argument, Newell & Simon, 1972). However, we deviated from the central ACT-R assumptions with respect to memory retrieval processes, and extended the standard system with RACE. As we have argued earlier (Van Maanen & Van Rijn, 2007b), default ACT-R cannot account for interference patterns found in PWI and Stroop tasks. In addition, the model parameters were chosen to fit the Stroop data, and afterwards only the perceptual encoding speed was adapted to fit the PWI data.

Discussion

When observing the typical temporal activity traces in Figure 1 of the model for both tasks, it becomes clear that the time the model has spent on retrievals from declarative memory is temporally shifted in the incongruent conditions. The picture-word interference trace is characterized by a long retrieval during the perceptual encoding stage, followed by three relatively quick retrievals during the response selection and execution stages. In contrast, the Stroop trace does not show a long retrieval during perceptual encoding, but it encounters relatively more interference during the response selection and execution phases. This pattern in the sub-processes of both tasks seems similar to the one theorized by Dell’Acqua et al. (2007), in which the locus of PWI is in the perceptual encoding phase, and the locus of the Stroop-effect is in the response selection phase. However, contrary to Dell’Acqua et al’s argument that this is an indication for different underlying mechanisms, we show here that a single mechanism accounts for both phenomena.

We further analyzed the contributions of the various retrieval steps to the final reaction times. Since the suggestion that the Stroop task and PWI are different processes is primarily based on the incongruent conditions, and only on the naming tasks (Dell'Acqua et al., 2007), we focused on the retrieval times in these conditions only. As the temporal activity traces in Figure 1 and the barplot in Figure 3 suggest, the retrieval time for the concept chunk differed between simulations of the Stroop task and PWI (t-test, \( t = 11.13, df = 36.43, p < 0.001 \)). Also, an ANOVA with as factors Task and Chunk Type showed a significant interaction effect between task and chunk type (F(1, 140)=41.78, p<0.001). Post-hoc paired t-tests demonstrate that the retrieval times for concepts and lexemes differ within each task (\( t = 5.71, df = 35, p < 0.001 \) for PWI, and \( t = 7.58, df = 35, p < 0.001 \) for Stroop). That is, in the picture-word interference task, the retrieval times of the concept retrieval are significantly higher than those of the

Figure 3: Relative contributions to the latency in the incongruent condition, for Stroop and PWI in our models. The bars show the percentage of the latency dedicated to the retrieval of the concept chunk, the lemma chunk, and the lexeme chunk respectively. Error bars denote standard errors. A clear interaction between the task (Stroop or PWI) and the relative contribution of concept retrievals and lexeme retrievals can be observed.

2 The model lisp code as well as the RACE ACT-R module can be downloaded from http://www.ai.rug.nl/cogmod.

3 Note that these analyses were conducted on the same model runs as reported earlier. The number of runs was kept equal to the reported empirical data.
lexeme retrieval, whereas in the Stroop task retrieval times for the concepts are significantly lower than those of the lexemes.

Conclusion

These results suggest that for picture-word interference, the contribution of the perceptual stage to the total interference effect exceeds the contributions of the response selection and execution stages. On the other hand, the Stroop interference effect is primarily caused in the later stages of response selection and execution. This is exactly the conclusion that Dell’Acqua et al (2007) draw based on theirs and Fagot and Pashler’s (1992) dual-task studies. However, in contrast to the conclusion Dell’Acqua et al present, we demonstrate here that this difference between PWI and Stroop can be explained with a single cognitive process underlying both tasks.

Acknowledgments

This research is supported by the Netherlands Organization for Scientific Research (ToKeNi/FR project grant no. 634.000.002).

References