

Goal-Referenced Selection of Verbal Action: Modeling Attentional Control in the Stroop Task

Ardi Roelofs

Max Planck Institute for Psycholinguistics

This article presents a new account of the color-word Stroop phenomenon (J. R. Stroop, 1935) based on an implemented model of word production, WEAVER++ (W. J. M. Levelt, A. Roelofs, & A. S. Meyer, 1999b; A. Roelofs, 1992, 1997c). Stroop effects are claimed to arise from processing interactions within the language-production architecture and explicit goal-referenced control. WEAVER++ successfully simulates 16 classic data sets, mostly taken from the review by C. M. MacLeod (1991), including incongruency, congruency, reverse-Stroop, response-set, semantic–gradient, time-course, stimulus, spatial, multiple-task, manual, bilingual, training, age, and pathological effects. Three new experiments tested the account against alternative explanations. It is shown that WEAVER++ offers a more satisfactory account of the data than other models.

This article addresses the issue of the attentional control of verbal action. The basic modes of language use, namely speaking, listening, reading, and writing, seem to use different configurations of basic processing components (e.g., Caplan, 1992; Levelt, 1989; Shallice, 1988). These configurations are selectively controlled. Hearing or reading a word does not unavoidably lead to its production; this is under the control of a speaker. Similarly, seeing an object does not necessarily lead to the naming of it. Furthermore, words do not occur in isolation but are typically part of a spoken discourse, a text on a page, or appear on objects in the real world, for example, as labels and signs. Without apparent problems, a speaker can read aloud one of the words or name the object and ignore the words. How does a speaker exert task-relevant control over the language processes underlying object naming and oral reading?

On the brink of the cognitive revolution in psychology, a clear answer to the question of how actions are controlled was given by Miller, Galanter, and Pribram (1960) in their book *Plans and the Structure of Behavior*: Human voluntary behavior is goal directed and actions are selected with explicit reference to these goals. Goals keep attention focused and help with decisions between potential actions. During subsequent years, this answer was further developed and refined to become the received view in the field (e.g., Newell & Simon, 1972). The proposal was still defended some 25 years later, for example, by Anderson (1983) in his book *The Architecture of Cognition*, and has been advocated even more recently (e.g., Anderson & Lebiere, 1998). However, the theoretical landscape has been extended. Along with the rise of connectionism in the 1980s came a new functional conception of the attentional control of action, which holds that attention does its

work by biasing activation levels of potential responses, as proposed, for example, by Norman and Shallice (1986). Goals are sources of activation that predispose for certain actions, but the selection of actions does not make explicit reference to these goals.

Two prominent, computer-implemented models developed in this new direction were published in 1990, namely those of Cohen, Dunbar, and McClelland (1990) and Phaf, Van der Heijden, and Hudson (1990), each representing an important subclass of approach. Both were models of the same task, the Stroop color-word task (Stroop, 1935). The Stroop task is one of the most widely used tasks in cognitive psychology, as observed in surveys of the field (e.g., Gazzaniga, Ivry, & Mangun, 1998). Stroop's (1935) article is among the most cited publications (if not the most cited one) in the history of experimental psychology (MacLeod, 1992). The control of verbal action has, in its most elementary form, perhaps been most exhaustively studied using Stroop stimuli (although from some perspectives, even the Stroop task is complicated). In a prototypical Stroop experiment, a person is presented with color words written in colored ink. The task is to name the color of the ink and to try to ignore the word. For example, speakers have to say "red" to the red ink of the word GREEN, the *incongruent* condition, to say "red" to the red ink of the word RED, the *congruent* condition, or to say "red" to the red ink of a series of Xs, the *control* condition. In another version of the task, speakers are asked to read aloud the word and ignore the color. The Stroop models of Cohen et al. (1990) and Phaf et al. (1990) have had a great impact (e.g., Styles, 1997). They are very explicit and open to empirical testing (e.g., Kanne, Balota, Spieler, & Faust, 1998; Mewhort, Braun, & Heathcote, 1992; Sugg & McDonald, 1994).

Cohen et al. (1990) assumed evidence accumulators for verbal actions, whereby an action is selected after sufficient evidence—expressed as activation—has been collected for it, with attention providing just one of the sources of activation in the accumulation process. Henceforth, I refer to this approach as *selection through accumulation*. Phaf et al. (1990) assumed an interactive activation network with basins of attraction for verbal actions, whereby an action is selected by moving to a stable state of activation for it, with attention providing just one of the sources of activation

I am grateful to Pim Levelt, Antje Meyer, Markus Damian, and Colin M. MacLeod (who helped me with the article outside of the review process), for helpful discussions and comments. I thank Monique van de Ven and Otto Kokke for their help in preparing and running the experiments.

Correspondence concerning this article should be addressed to Ardi Roelofs, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, the Netherlands. E-mail: ardi@mpi.nl

biasing the settling process. I refer to this approach as *selection through attraction*. Although the models of Cohen et al. (1990) and Phaf et al. addressed several data sets on attention (e.g., Cohen et al., 1990, set out to explain training data on Stroop task performance from MacLeod & Dunbar, 1988), both models have attempted to account by computer simulation for the findings on the time course of the Stroop phenomenon obtained by M. O. Glaser and Glaser (1982). Since the seminal chronometric work of Donders (1868/1969) and Cattell (1886) in the early days of experimental psychology (Cattell was a student of Wundt), many researchers have convincingly argued that time-course data provide among the most detailed “fingerprints” of a cognitive process (e.g., Luce, 1986; McGill, 1963; Posner, 1978; Townsend & Ashby, 1983; Vorberg, 1985). Donders was the first who measured naming times and reported that “distraction during the appearance of the stimulus is always punished with prolongation of the process” (1868/1969, p. 428).

In this article, I begin by arguing that the accumulator model of Cohen et al. (1990) and the attractor model of Phaf et al. (1990), for different specific reasons and a common general one, fail to account for critical aspects of the time-course findings. I then describe goal-referenced selection of action, which I refer to as *selection through verification*. I present a further development of the mechanism for attentional selection by verification in the WEAVER++ model of word production (e.g., Levelt, Roelofs, & Meyer, 1999b; Roelofs, 1992, 1997c) and apply the model to the Stroop task.

Unlike the models of Cohen et al. (1990) and Phaf et al. (1990), WEAVER++ has been designed to account for spoken word production and has been applied to semantic and phonological effects from the picture–word interference task and to key findings obtained with several other word-production tasks (e.g., Levelt et al., 1999a, 1999b; Roelofs, 1992, 1993, 1994, 1996a, 1996b, 1996c, 1997a, 1997b, 1997c, 1998b, 1999; Roelofs & Meyer, 1998; Roelofs, Meyer, & Levelt, 1996, 1998), but the model has never been applied to distractor effects from the color-word Stroop task. In a picture–word interference experiment, participants name pictures while trying to ignore spoken or written distractor words. A semantic effect concerns, for example, the difference in effect of the distractors DOG and HOUSE in naming a pictured cat, where both distractor words are incongruent with the target. A phonological effect concerns, for example, the difference in effect between CAP and HOUSE, where *cap* shares part of its form with the target *cat*, but again, both distractor words are incongruent with the target. Thus, the Stroop phenomenon constitutes a new test bed for the model: WEAVER++ accounts for semantic and phonological effects, but it could, in principle, be far off concerning incongruency and congruency effects relative to a neutral control condition.

In WEAVER++, selection takes place with explicit reference to goals for verbal actions. I show that WEAVER++ does not suffer from the particular problems that the models of Cohen et al. (1990) and Phaf et al. (1990) have with the time course of the Stroop phenomenon but that WEAVER++ in its published form does not give a full account either. To deal with this shortcoming, I adopt the distinction between expectancy-induced and automatic verbal priming proposed by Posner and Snyder (1975), Neely (1977, 1991), and M. O. Glaser and Glaser (1982), among many others. Expectancy-based priming occurs when the time between the presentation of a distractor and target is sufficiently long for

verbal response anticipation to occur. WEAVER++ embodies automatic but not expectancy-induced priming. I show that by including verbal response anticipation in the model, the time course of the classic Stroop phenomenon and many variants of it can be captured.

Basic Characteristics of the Stroop Phenomenon

The Stroop task (Stroop, 1935) requires naming the ink color of a color word (Stroop’s Experiment 2) or, alternatively, reading the word aloud (Stroop’s Experiment 1). The color word may correspond to the ink color, or it may be the name of a different color. Performance in these conditions is compared with that in a control condition. For color naming, the control condition generally consists of a patch of color or a row of colored x’s. For reading, the control condition consists of an uncolored word. Different from many modern versions of the task, Stroop’s task did not use a congruent condition, which was introduced by Dalrymple-Alford and Budayr (1966). Stroop’s (1935) original experiments measured the time taken to complete a list of stimuli, which is still typical of psychometric applications of the task (e.g., Abramczyk, Jordan, & Hegel, 1983; Wapner & Krus, 1960; Wysocki & Sweet, 1985). In this original form, participants are presented with four cards. One card has color words on it printed in black ink, and the words have to be read aloud. Another card contains colored patches, and the task is to name the color. The two remaining cards have words on them each printed in a conflicting ink color, with the words to be read (one card) or the colors to be named (the other card). The critical measure is the time it takes to complete each card. Using cards with multiple stimuli generally gives the same pattern of results as those obtained from measuring the time to name or read individual Stroop stimuli, which is the standard in modern experimental settings.

In a monumental article, MacLeod (1991) reviewed the literature on the Stroop task covering over 50 years and some 400 studies. He made a list of robust, key, empirical results that must be explained by any successful account of the Stroop phenomenon, with the following findings being absolutely central (see also Dyer, 1973; W. R. Glaser & Glaser, 1989; W. R. Glaser, 1992). Table 1 lists the findings that are addressed in the current article.

The most critical finding is that color naming is much slower in the incongruent condition (e.g., saying “red” because of the red ink in which the word GREEN is written) than in a control condition where there is no color word but, typically, a string of colored x’s. This effect persists even after thousands of trials (see, e.g., MacLeod, 1998). The delay in color naming is often referred to as *Stroop interference*. The amount of interference differs somewhat from study to study, but is often 100 ms or more (e.g., Dunbar & MacLeod, 1984; M. O. Glaser & Glaser, 1982; Kahneman & Chajczyk, 1983). There is less interference when the color word does not correspond to one of the colors used in the experiment. For example, when the colors in an experiment are red and green, the word BLUE in red ink produces less interference than the word GREEN in red ink (e.g., Klein, 1964; Proctor, 1978); this is called the *response set* effect. The amount of interference decreases with increasing semantic distance: When printed in red, the color-neutral distractor word THEME produces less interference than the color-related distractor word SKY, and SKY yields less interference than BLUE (Dalrymple-Alford, 1972; Klein, 1964), which is often

Table 1
Sixteen Basic, Robust Findings on Stroop Performance From Over Half a Century of Research That Are Addressed in the Current Article

No.	Finding
1 ^a	Interference exceeds facilitation in color naming.
2 ^a	Response set membership increases interference for color naming as well as for oral reading in the word–word task.
3 ^a	There is no reverse Stroop effect (i.e., no effect of colors on oral reading), except when there is a very high proportion of congruent trials.
4 ^a	Interference in color naming increases with decreasing distractor preexposure (distractor-first SOAs) and decreasing distractor postexposure (distractor-second SOAs); maximal impact is observed around SOA = 0 ms.
5 ^a	Facilitation in color naming is constant at distractor-first SOAs.
6 ^a	The picture-word task yields results equivalent to those obtained with the color-word task.
7	Reading with word–word stimuli yields interference but no semantic gradient.
8	Task uncertainty yields SOA curves of interference equivalent to those obtained with task certainty.
9	There is less interference with color–color than with color–word stimuli.
10 ^a	Interference reduces with spatial separation of the color and word; with spatial certainty, there is no interference around SOA = 0 ms with color–color and word–word stimuli.
11 ^a	Interference in color naming decreases with increasing semantic distance.
12 ^a	Manual responding yields results that are equivalent to those obtained with spoken responding.
13 ^a	There is less interference between than within languages of bilingual individuals.
14 ^a	With extensive training on calling arbitrary shapes by color names, interference from colors in shape naming changes into interference from shapes in color naming; with moderate practice, there is interference in both tasks.
15 ^a	Interference across the life span is U-shaped.
16	There is increased interference with frontal lobe lesions, Alzheimer dementia, and schizophrenia.

Note. Compared with MacLeod's (1991) list, this table adds a few findings, such as those on word–word performance, which in my view are theoretically very important, and omits a few findings, such as the absence of gender differences in Stroop performance, which are theoretically not so revealing. SOA = stimulus onset asynchrony.

^a Findings from MacLeod (1991).

referred to as the *semantic gradient*. Furthermore, color naming is often faster in the congruent condition (the word RED in red ink) compared with the control condition, but it is important to note that this facilitation is always less than the interference from the incongruent condition (e.g., Dunbar & MacLeod, 1984; Dyer, 1971b; MacLeod, 1998). Facilitation is not always observed (e.g., Mewhort et al., 1992), but usually it ranges from about 20 ms (Regan, 1978) to 50 ms (Kahneman & Chajczyk, 1983).

When the task is to read the word aloud rather than to name the ink color, the response is usually unaffected by congruent and incongruent colors (e.g., the reading aloud of the word RED is virtually unaffected by red or green ink). That is, there is no “reverse” Stroop effect, unless a very high proportion of congruent trials biases the use of the color to initiate responding (e.g., M. O. Glaser & Glaser, 1982; Logan & Zbrodoff, 1979). However, when word–word stimuli are used, and one of the words has to be read

(e.g., the first of two words presented above each other with a certain stimulus onset asynchrony [SOA], or the word that is designated by a bar marker), one obtains a Stroop conflict and response set effect but no semantic gradient (M. O. Glaser & Glaser, 1982; W. R. Glaser & Glaser, 1989; Van der Heijden, 1981).

The Stroop phenomenon is not only observed with naming colors and reading color words but occurs in many other verbal domains, such as with position naming and position word reading (e.g., Virzi & Egeth, 1985), numerosity naming and numeral reading (e.g., Flowers, Warner, & Polansky, 1979), and picture naming and word reading (e.g., W. R. Glaser & Dünghoff, 1984). Pressing a colored button to refer to the color of a color-word Stroop stimulus is interfered with by incongruent words, but pressing a word-labeled button to refer to the word is not influenced by incongruent colors (e.g., Sugg & McDonald, 1994). Stroop interference is also observed with typing the word (Logan & Zbrodoff, 1998).

Usually, a group of participants is asked to name the color and to ignore the word; sometimes (e.g., W. R. Glaser & Dünghoff, 1984; M. O. Glaser & Glaser, 1982), another group of participants is asked to read the word aloud and to ignore the color. Stroop (1935) also asked one group of participants to name colors and another group to read words. However, when words and color patches are presented with a certain SOA, and a single group of participants is asked to refer to the first (or second) stimulus by reading it when it is a word and naming it when it is a color, the standard Stroop patterns are replicated (e.g., W. R. Glaser & Glaser, 1989).

The magnitude of Stroop interference varies with age, with young children and older adults exhibiting the largest interference scores (e.g., Comalli, Wapner, & Werner, 1962; Schiller, 1966). Persons diagnosed with schizophrenia, frontal lobe lesions, or Alzheimer dementia show Stroop interference scores that are disproportionately high (e.g., Abramczyk et al., 1983; Janer & Pardo, 1991; Spieler, Balota, & Faust, 1996). Whereas monolinguals are not influenced by irrelevant words from a foreign language, Stroop interference does occur between the languages of bilingual individuals (e.g., Dyer, 1971a). The amount of interference is usually less between than within languages. Extensive training on color naming does not yield interference of colors on word reading (e.g., MacLeod, 1998; but see Stroop, 1935, Experiment 3). However, interference may come and go as a function of training on some tasks, such as calling novel shapes by color names (MacLeod & Dunbar, 1988).

Over the past half century, several explanations for the color-word and related Stroop phenomena have been advanced and have been extensively reviewed by Dyer (1973), W. R. Glaser (1992), and MacLeod (1991). Most accounts construe performance in the Stroop task as a problem in the attentional control of action as opposed to, for example, a problem of selective attention in visual processing. On the latter account (e.g., Hock & Egeth, 1970), the visual processing of the word interferes with the processing of the color by diverting attention from it, which would explain the delay of color naming caused by the color word compared with the xs in the control condition. However, this view fails to explain why congruent words (which are also words) often help in naming the color (e.g., Kahneman & Chajczyk, 1983; Regan, 1978). Furthermore, Duncan-Johnson and Kopell (1980, 1981) observed the

normal Stroop patterns in color naming times but no effect on the P300, an evoked brain potential that is usually taken to be an index of task-relevant visual encoding and evaluation. The P300 latency increased when hues were used that were more difficult to discriminate (e.g., reddish purple and bluish purple for the words *red* and *blue*). Thus, the P300 reflected visual processing but not the Stroop conflict. This does not, of course, exclude the aspect of visual selection in the Stroop phenomenon. In fact, good evidence exists that such selection plays some role. Although weakened, Stroop interference is obtained when the ink color and word are physically separated by presenting the word below or above a color patch (e.g., Kahneman & Henik, 1981; MacLeod, 1998). However, with spatially separated color-color and word-word stimuli, Stroop interference is only observed when the location of the target (above or below the distractor) at each trial is uncertain. Under spatial certainty (e.g., the target color patch always appears below the distractor patch), no interference around SOA = 0 ms is obtained (M. O. Glaser & Glaser, 1982), which suggests that under this condition, the distractor may be filtered out during early, spatial processing of the visual stimuli.

The most influential account of the Stroop phenomenon has been one of “late” selection. In a classic view, the phenomenon is attributed to a response-related bottleneck, like a response buffer that can hold a single word only (e.g., Morton, 1969; Posner & Snyder, 1975). Reading is typically faster than color naming (about 100–200 ms; see Cattell, 1886; Dyer, 1973; Fraisse, 1969; M. O. Glaser & Glaser, 1982), and therefore the response word that is derived from the perceived word occupies the buffer before the word that is derived to name the color. Stroop interference reflects the time it takes to clear the buffer so that the word for the color can be produced. In this view, reading is not affected by the color because the word for the color is simply derived too slowly to occupy the buffer before the perceived word.

Two decades ago, however, this relative-speed-of-processing explanation was shown to be flawed. First, Dunbar and MacLeod (1984) showed that when reading is made much slower than color naming by presenting the distractor words upside down or backward, Stroop interference in color naming persists virtually unaltered. Second, the relative-speed account predicts that a reverse Stroop effect (i.e., interference on reading aloud by an incongruent color) should be obtained when the color is presented sufficiently in advance of the word to be read. However, M. O. Glaser and Glaser (1982) showed that when an incongruent color patch is presented before the word (e.g., 100, 200, 300, or even 400 ms), still no reverse Stroop interference is obtained. Because an SOA of 300 or 400 ms fully compensates for the 100–200 ms difference in processing time between color and word, this suggests that the relative speed of processing is not the crucial factor (for a mathematical proof, see Vorberg, 1985). What, then, accounts for the phenomenon?

The remainder of this article is organized as follows. First, I describe the key findings on the time course of the Stroop phenomenon in greater detail. Next, I indicate why the accumulator model of Cohen et al. (1990) fails to account for the findings. Cohen et al. (1990; Cohen & Huston, 1994) suggested that their model fails because the time course findings are contaminated by “habituation.” I report an experiment that shows, however, that there is no evidence for such contamination. Next, I indicate why the attractor model of Phaf et al. (1990) also has difficulty ac-

counting for the time course findings. Then I explain the alternative to attentional control through activation biasing provided by “production rule system” models, in which actions are selected with explicit reference to goals. The WEAVER++ model falls into this general class of model. In subsequent sections, I show that WEAVER++ accounts for the Stroop phenomenon and its time course, and report two new empirical tests of the model. Furthermore, the model is shown to account for performance in variants of the task that manipulate response set membership, semantic distance, stimulus types, task and location uncertainty, response mode and type, basis of responding, and amount of training. Also, findings from the picture-word task and different age groups, bilingual people, and clinical groups are explained. The article ends by discussing some wider implications.

Time Course of the Stroop Phenomenon

The first study that manipulated the SOA between the word and color was by Dyer (1971b). He used SOAs ranging from 0 to 500 ms preexposure of the word (i.e., 0, 20, 40, 60, 80, 100, 140, 200, 300, and 500 ms) in exploring the effects of incongruent and congruent word-color combinations on color naming. On each trial, a word was presented in black with a certain SOA before the same word appeared in a color that had to be named. Dyer observed that interference decreased when the SOA increased. The interference was about 90 ms at the SOA of 0 ms, increased to 118 ms at an SOA of 40 ms, but dropped to 68 ms at SOA = 200 ms and then further diminished to 32 ms at an SOA of 500 ms. Facilitation increased a bit from SOA = 0 ms to the SOA of 200 ms.

Using Dyer’s (1971b) presentation technique, Thomas (1977) varied the SOA from the word in black preceding the same word in color by 300 ms (henceforth called *distractor-first* SOAs, indicated by a minus sign) to the word in color leading the same word in black by 300 ms (called *distractor-second* SOAs). Note that Dyer used distractor-first SOAs only. Consistent with Dyer’s findings, the interference on color naming in Thomas’s study was greatest around SOA = 0 ms. Facilitation occurred when the word preceded or even when it closely followed the corresponding ink color (up to 50 ms).

Similarly, M. O. Glaser and Glaser (1982) examined the time course of Stroop interference and facilitation by presenting words and colors with a wide range of distractor-first and distractor-second SOAs. Moreover, they instructed one group of participants to name the colors and another group of participants to read the words aloud. In his review of the Stroop literature, MacLeod (1991) called this the “definitive” study (p. 180). Unlike in Dyer’s (1971b) procedure, the word was presented in white on a dark background, and the color was presented on this background as a colored rectangle. In color naming, the onset of the presentation of the word could be 400, 300, 200, or 100 ms before the onset of the presentation of the color patch (the distractor-first SOAs), the onset of word and color could coincide (the classic situation of SOA = 0 ms), or the word could follow the color patch with a lag of 100, 200, 300 or 400 ms (the distractor-second SOAs). Identical SOAs were used for word reading. Trials were blocked by SOA.

Consistent with Dyer (1971b) and Thomas (1977), M. O. Glaser and Glaser (1982) observed that the interference with color naming in the incongruent condition increased considerably as the preexposure time of the word (i.e., the distractor-first SOA) became

shorter. In particular, the interference at SOA = -400 ms (i.e., 25 ms) was about half of that at SOA = -100 ms (45 ms) and about a third of that at SOA = 0 ms (72 ms). The interference also decreased as the postexposure time of the word (i.e., the distractor-second SOA) became longer. In particular, the interference at SOA = 100 ms (i.e., 73 ms) was about three times as big as the interference at SOA = 200 ms (i.e., 24 ms). At later SOAs, no interference was left. For the congruent condition, a rather flat pattern of facilitation was observed from a 31-ms effect at the distractor-first SOA of -400 ms to a 30-ms effect at the distractor-first SOA of -100 ms. Figure 1 shows how the Stroop interference and facilitation varied with SOA for color naming.

These SOA patterns of interference and facilitation have been replicated with the Stroop task, for example, by Neumann (1980), Goolkasian (1981), and Long and Lyman (1987), and by W. R. Glaser and Dungelhoff (1984) using the picture-word variant. For reading aloud, M. O. Glaser and Glaser (1982) observed basically no facilitation or interference for any SOA. This corresponds to what Lassen (1975) and W. R. Glaser and Dungelhoff observed for the picture-word task.

In interpreting the SOA patterns of interference and facilitation from color words on color naming, M. O. Glaser and Glaser (1982) adopted the distinction between "automatic" primary effects and

"controlled" secondary effects proposed by Taylor (1977). According to M. O. Glaser and Glaser, the huge amount of interference around the SOA of 0 ms is a fast, primary effect reflecting the basic Stroop conflict. According to them, the conflict has no fast, primary facilitatory complement. The latter claim seems to be incorrect because other researchers have obtained facilitation at SOA = 0 ms (e.g., Kahneman & Chajczyk, 1983; Long & Lyman, 1987; MacLeod, 1998; Regan, 1978). The effects at more negative SOAs were interpreted as being slow, secondary effects due to anticipation of the right or wrong response. The idea is that facilitation and interference effects occur here because the target is anticipated correctly on congruent trials but incorrectly on incongruent trials (e.g., reading RED induces the expectation of "red" as target). A similar distinction between controlled and automatic priming has been proposed by Posner and Snyder (1975) and Neely (1977). The distinction is independently supported by the finding that interference and facilitation with long distractor-first SOAs persist under conditions that yield no effect around SOA = 0 ms (e.g., M. O. Glaser & Glaser, 1982). A distractor preexposure time of less than 250 ms has been shown to eliminate effects that are considered to be due to anticipation (for a review, see Neely, 1991). In summary, analyses suggest that the Stroop phenomenon reflects automatic priming with short distractor preexposure times and expectancy-induced priming at longer preexposure times. What selection mechanism accounts for the SOA findings?

Selection Through Accumulation

The prototype accumulator model in the word perception and production literature is the logogen model of Morton (1969). Each word is represented by a *logogen*, which is a counter that collects perceptual or conceptual evidence for a word. When the tally exceeds threshold, the logogen fires and the articulatory program for the word is made available. The model accounts for the Stroop effect by assuming a limited-capacity buffer for holding the program. As MacLeod (1991) pointed out, a problem with the accumulator account in its simplest form is that it predicts symmetrical Stroop interference and facilitation, but facilitation is typically smaller than interference. Another problem is that the account relies on the relative speed of processing of the color and word in explaining the lack of a reverse Stroop effect. It thereby runs into difficulties explaining the findings of Dunbar and MacLeod (1984), M. O. Glaser and Glaser (1982), and MacLeod and Dunbar (1988). The model of Cohen et al. (1990) remedies these problems by assuming that evidence accumulates through nonlinear pathways of different strengths.

The model of Cohen et al. (1990; Cohen & Huston, 1994; Cohen & Servan-Schreiber, 1992; Cohen, Servan-Schreiber, & McClelland, 1992; Cohen, Usher, & McClelland, 1998) assumes a feed-forward activation network consisting of input, hidden, and output nodes, which are connected to evidence accumulators (Figure 2). Processing occurs through activation spreading from input nodes via hidden nodes to output nodes, whereby nodes change their activation with time in a continuous, nonlinear manner. The strengths of the connections between input, hidden, and output nodes were set using the error back-propagation algorithm. This resulted in excitatory links between nodes representing compatible information and inhibitory links between nodes standing for in-

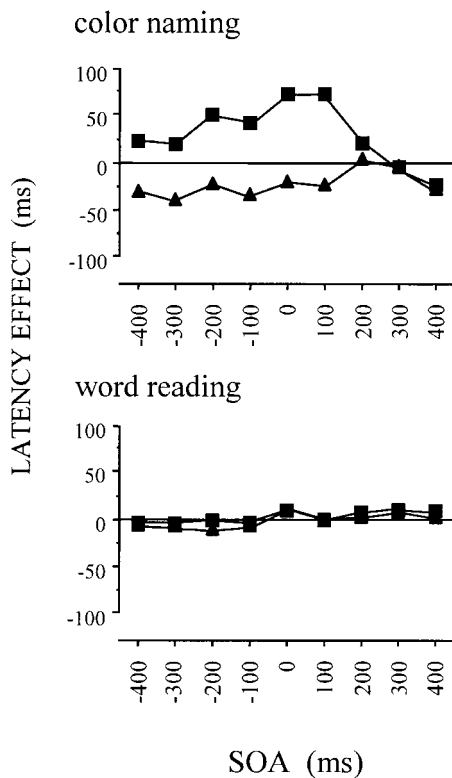


Figure 1. Time course of the color-word Stroop phenomenon as empirically measured by M. O. Glaser and Glaser (1982, Experiment 1) for color naming and word reading. For each SOA, the mean latency effects are shown. Preexposure SOAs are indicated by a minus sign. Positive effects indicate interference and negative effects indicate facilitation relative to the control condition. ■ = incongruent; ▲ = congruent; SOA = stimulus onset asynchrony.

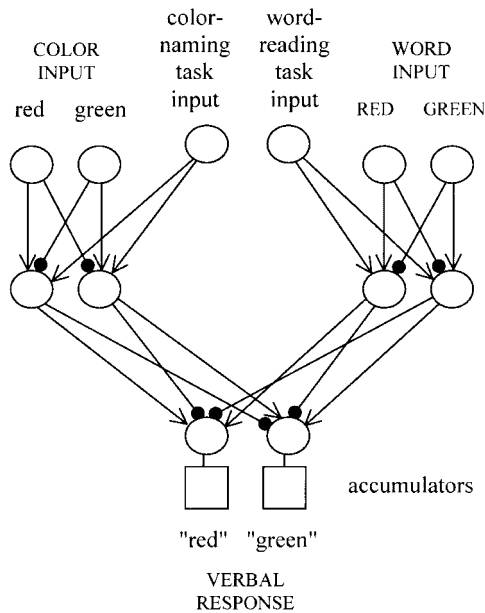


Figure 2. Network architecture of the model of Cohen, Dunbar, and McClelland (1990). The lines with black circles denote unidirectional inhibitory connections and the lines with arrows indicate unidirectional excitatory connections.

compatible information. To reflect the assumption that reading is a more highly practiced skill than color naming, reading was given more back-propagation learning trials than color naming, which resulted in stronger connections in the reading than in the color pathway.

In the Cohen et al. (1990) model, selective attention is achieved by task nodes for color naming and reading. These nodes provide extra input to one of the pathways, depending on the task. At the beginning of a simulated Stroop trial, all accumulators are set to zero. A task node is activated and the model is run until the activation levels of all nodes stabilize. This allows the system to settle into a “ready state” for the task. Next, the components of a Stroop stimulus are presented with the appropriate SOA. In each time step, the evidence accumulators add a small amount of activation to their accumulated total. A response is selected when one of the accumulators exceeds a fixed threshold.

The stronger pathway for reading than for color naming explains why there is Stroop interference in color naming but not in reading. However, the model does not adequately account for the time course of the Stroop phenomenon as obtained by M. O. Glaser and Glaser (1982) and many others. Furthermore, the model wrongly predicts a lack of interference for processes of comparable strength (e.g., MacLeod & Dunbar, 1988).

The Time Course of the Effects

In the Cohen et al. (1990) model, the amount of evidence accumulated for the written word in color naming is a positive function of the preexposure time of the word. Thus, interference peaks at the longest distractor-first SOA and decreases when the preexposure time becomes shorter. Similarly, facilitation peaks at the longest distractor-first SOA and decreases when the SOA

becomes shorter. Figure 3 shows the performance of the model for color naming in simulations run by Cohen et al. (1990). The effects differ from the empirical results, as shown in Figure 1. Empirically, maximal impact of incongruent words is observed when the words appear within 100 ms of the colors, and facilitation from preexposed congruent words is constant. As Figure 3 shows, the model also predicts a small reverse Stroop effect at distractor-first SOAs in reading, contrary to the real data. Clearly, the model misses an important aspect of the Stroop data, as Cohen et al. (1990) admitted.

Cohen et al. (1990; Cohen & Huston, 1994) considered the possibility that a process of habituation affects preexposed distractor words, such that it would reduce the amount of interference and facilitation at distractor-first SOAs but not at SOA = 0 ms.¹ Next, I report a new experiment that specifically tested for an effect of stimulus condition repetition in a replication of M. O. Glaser and Glaser (1982). Congruent, incongruent, and control Stroop stimuli were tested at two SOAs: -400 and 0 ms. If habituation is effective so that Stroop interference decreases with increasing stimulus condition repetition at distractor-first SOAs but not at SOA = 0 ms (according to Cohen et al., 1990, the explanation for the increase of interference from SOA = -400 ms to SOA = 0 ms), an interaction should be obtained between distractor, SOA, and repetition. Such an interaction would be consistent with the two-process account of Cohen et al. (1990). However, if habituation plays no role, and therefore the distractor effects do not disappear with increasing stimulus condition repetition, no interaction between distractor, SOA, and repetition should be obtained.

Experiment 1: A Test of Habituation

Method

Participants. The experiment was carried out with 24 paid participants from the pool of the Max Planck Institute. All participants were young adult native speakers of Dutch.

Materials and design. The stimuli were the three Dutch color words *rood* (“red”), *groen* (“green”), and *blauw* (“blue”) and their colors. The control condition consisted of a row of five x’s (xxxxx). The colors were presented as colored rectangles 3 cm high × 9 cm wide. The words were on average 1.5 cm high × 6 cm wide. There were three independent variables, which were varied within participants. The first independent variable was SOA with two levels: -400 and 0 ms. SOA was varied within participants but between trial blocks. Twelve participants were first tested on SOA = 0 ms and then on SOA = -400 ms. For the remaining 12 participants, the order of testing the SOAs was reversed. The second independent variable was distractor. In the stimulus set, three congruent, three incongruent (*ROOD*-blue, *GROEN*-red, *BLAUW*-green), and three con-

¹ Cohen and colleagues (Cohen et al., 1990; Cohen & Huston, 1994) suggested that habituation plays a critical role, but they did not specify how it works. Nevertheless, their suggestion is important because it carries the burden of explaining the discrepancy between their model and the real data. Habituation may be to a distractor condition (i.e., stimulus type) or a particular stimulus (i.e., stimulus token). Relevant for the model of Cohen et al. (1990) is that habituation should lead to reduced interference at the level of distractor conditions, which may result from repetition of either a stimulus type or token. At the level of distractor conditions, an increase in interference from SOA = -400 ms to SOA = 0 ms should go together with a decrease of interference with repetition at SOA = -400 ms but not at SOA = 0 ms.

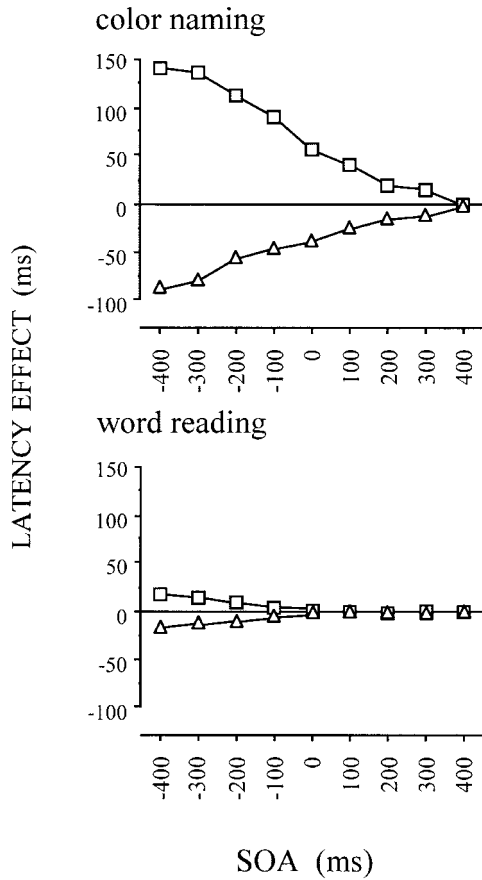


Figure 3. Time course of the color-word Stroop phenomenon as simulated by Cohen, Dunbar, and McClelland (1990) for color naming and word reading. □ = incongruent; △ = congruent; SOA = stimulus onset asynchrony.

control pairings were created, making up the congruent, incongruent, and control condition. Each of the congruent, incongruent, and control pairings occurred 10 times within an SOA block. Thus, for each SOA, the participants received a congruent stimulus 30 times, an incongruent stimulus 30 times, and a control stimulus 30 times. The stimuli were presented in random order. The repetition of the congruent, incongruent, and control condition was the third independent variable, stimulus condition repetition.

Procedure and apparatus. The participants were tested individually. They were seated in front of a Multisync computer monitor (NEC Computers, Wijnchen, the Netherlands) and an electronic Sennheiser microphone (Wedemark, Germany). The distance between participant and screen was approximately 50 cm. The participants were asked to name the color patches and to ignore the color words. After the participant had read the instructions, a block of 12 practice trials with SOA = 200 ms was administered, which was followed by the two experimental SOA blocks. The structure of a trial was as follows. First, the participant saw a warning signal (an asterisk) for 500 ms. Next, the screen was cleared for 500 ms, followed by the display of the components of a Stroop stimulus with the appropriate SOA. The stimuli were presented on a black background. The asterisk and words were presented in white. The written words and colored rectangles were shown in the middle of the screen; the words always appeared in the center of the colored rectangles. Before the start of the next trial there was a blank interval of 500 ms. The total duration of a trial was 3 s. A Herman computer (Scherpenzeel, the Netherlands) controlled the stimulus presentation and data collection.

Analysis. After each trial, the experimenter coded the response for errors. Five types of incorrect responses were distinguished: wrong response word, wrong pronunciation of the word, a disfluency, triggering of the voice key by a nonspeech sound, and failure to respond within 1,500 ms after target presentation. Incorrect responses were excluded from the statistical analyses of the production latencies. The latencies and errors were submitted to analyses of variance (ANOVAs) with the crossed variables distractor, SOA, and stimulus condition repetition. All variables were tested within participants.

Results and Discussion

Table 2 gives the means and standard deviations of the production latencies and the mean error percentages (all excluded responses) for the distractor by SOA cells. At SOA = -400 ms, the incongruent condition yielded an interference effect of 27 ms and the congruent condition produced 32 ms facilitation compared with the control condition. For SOA = 0 ms, 91 ms interference and 29 ms facilitation was obtained. These findings replicate M. O. Glaser and Glaser (1982), who obtained 25 ms interference and 31 ms facilitation for SOA = -400 ms, and 72 ms interference and 16 ms facilitation for SOA = 0 ms. Figure 4 shows for each SOA how the latency effects varied with stimulus condition repetition. At both SOAs, the interference is present at all repetitions; the amount of interference is roughly constant. For SOA = -400 ms, it is certainly not the case that the interference is high with few repetitions (the Cohen et al., 1990, model predicts about 150 ms interference) and then reduces with more repetitions. The same holds for the facilitation.

Production latencies. The statistical analyses of the production times confirmed these observations. There were main effects of distractor, $F(2, 46) = 122.11, MSE = 23,475, p < .001$; SOA, $F(1, 23) = 5.88, MSE = 126,942, p < .05$; and stimulus condition repetition, $F(29, 667) = 3.23, MSE = 11,180, p < .001$. SOA interacted with distractor, $F(2, 46) = 27.25, MSE = 16,243, p < .001$, and with repetition, $F(29, 667) = 1.65, MSE = 9,766, p < .05$, but there was no interaction between distractor and repetition,

Table 2
Means (RTs, in Milliseconds) and Standard Deviations (SDs) of the Production Latencies and Mean Error Percentages (E%) as a Function of Distractor and Stimulus Onset Asynchrony (SOA) for Experiment 1

Distractor	SOA (ms)	
	-400	0
Congruent		
RT	561	569
SD	113	142
E%	0.6	0.6
Incongruent		
RT	620	689
SD	127	154
E%	0.8	3.8
Control		
RT	593	598
SD	119	124
E%	0.3	0.3

Note. RT = response time.

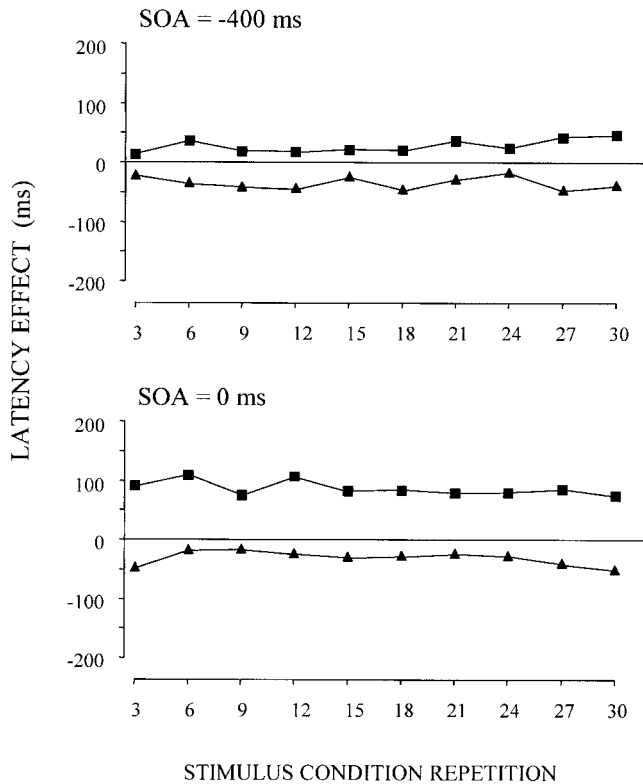


Figure 4. Latency effect of congruent and incongruent stimuli compared with control stimuli as a function of stimulus condition repetition at the SOAs of -400 and 0 ms in Experiment 1. ■ = incongruent; ▲ = congruent; SOA = stimulus onset asynchrony.

$F(58, 1334) = 0.62$, $MSE = 109,333$, $p > .50$. Importantly, there was no interaction between distractor, SOA, and repetition, $F(58, 1334) = 0.98$, $MSE = 10,090$, $p > .50$ —the power to detect an interaction, $p(\text{reject } H_0 | H_1 \text{ true})$, was very high, namely .985 at the .05 level. The incongruent condition was slower than the control condition at both the SOA of -400 ms and that of 0 ms, respectively, $F(1, 23) = 13.71$, $MSE = 18,515$, $p < .001$, and $F(1, 23) = 105.20$, $MSE = 26,993$, $p < .001$. The interference at SOA = -400 was smaller than that at SOA = 0 ms, $F(1, 23) = 39.84$, $MSE = 17,448$, $p < .001$. The congruent condition was faster than the control condition at both the SOA of -400 ms and that of 0 ms, respectively, $F(1, 23) = 29.64$, $MSE = 12,164$, $p < .001$, and $F(1, 23) = 22.48$, $MSE = 13,220$, $p < .001$, but there was no interaction with SOA, $F(1, 23) = 0.13$, $MSE = 11,531$, $p > .50$. None of the effects interacted with stimulus condition repetition; all F s were around 1.0.

Error rates. Table 2 shows that most errors were made in the incongruent condition at SOA = 0 ms and that the other conditions did not differ much. The statistical analyses of the error rates yielded main effects of distractor, $F(2, 46) = 14.17$, $MSE = 0.01$, $p < .001$; SOA, $F(1, 23) = 9.47$, $MSE = 0.01$, $p < .005$; and stimulus condition repetition, $F(29, 667) = 1.88$, $MSE = 0.01$, $p < .005$. SOA interacted with distractor, $F(2, 46) = 12.95$, $MSE = 0.01$, $p < .001$, confirming that most errors were made in the incongruent condition at SOA = 0 ms. There was no interaction between SOA and repetition, $F(29, 667) = 1.09$, $MSE = 0.01$,

$p > .10$, nor between distractor and repetition, $F(58, 1334) = 1.22$, $MSE = 0.01$, $p > .10$, or between distractor, SOA, and repetition, $F(58, 1334) = 1.01$, $MSE = 0.01$, $p > .50$.

To summarize, the current experiment replicates the findings of M. O. Glaser and Glaser (1982): Interference increases when the preexposure time of the word becomes shorter, but facilitation remains constant. Importantly, the experiment shows that repetition does not differentially influence the distractor effect between SOA = -400 ms and SOA = 0 ms, contrary to the habituation hypothesis. The magnitude of interference at SOA = -400 ms in the model of Cohen et al. (1990) is about 150 ms, but in the real data interference does not exceed 50 ms at any repetition. It is certainly not the case that interference in the real data starts at 150 ms and then diminishes during the experiment to yield, on average, 27 ms interference. To conclude, the increase of interference from SOA = -400 ms to SOA = 0 ms cannot be explained as being due to habituation at distractor-first SOAs.

No Interference for Processes of Comparable Strength?

The Cohen et al. (1990) model achieves attentional control by task nodes that selectively enhance the activation in the relevant pathway. However, if, for example, the second of two color patches has to be named and the first ignored, the two stimuli are processed by a single pathway of a certain strength. Consequently, the task cannot be performed by selectively enhancing the activation in the relevant pathway. Yet, with color-color and word-word stimuli, patterns of interference and facilitation are obtained that are empirically similar to those obtained with color-word stimuli, where the pathways differ (e.g., M. O. Glaser & Glaser, 1982; W. R. Glaser & Glaser, 1989). One might propose to split up the color and word pathways in the model so that there are separate color and word pathways for the first stimulus and second stimulus presented (e.g., one pathway processing the first color and another pathway processing the second color). However, there is no reason to assume that such pathways differ in strength (e.g., for first color vs. second color), which is a function of practice in the model. When pathways are of approximately equal strength, the model produces no interference (see Cohen et al., 1990, Figure 12). Of course, in the course of an experiment, participants may become more practiced in responding to the first (or second) color. However, this would increase the strength of the target pathway, which should make performance even less susceptible to interference, according to the model. Thus, the presence of specific color and word pathways for the first stimulus and the second stimulus predicts a lack of interference in the color-color and word-word variants of the Stroop task, contrary to the empirical findings of M. O. Glaser and Glaser (1982) and W. R. Glaser and Glaser (1989).

Furthermore, the model of Cohen et al. (1990) holds that interference patterns can be reversed if one pathway is made stronger than the other pathway by extensive training. Training color naming to the extent that it becomes faster than oral reading appears to be impossible empirically (e.g., Brown, 1915). However, MacLeod and Dunbar (1988) observed that with extensive training on calling arbitrary shapes by color names, interference from colors in shape naming changed into interference from shapes in color naming; with intermediate practice, there was interference in both tasks. Cohen et al. (1990) were successful in simulating the rever-

sal of interference as a function of the amount of training. However, when the pathways for color naming and shape naming were of approximately equal strength, the model produced virtually no interference from colors on shape naming or from shapes on color naming, contrary to the data of MacLeod and Dunbar.

In summary, the accumulator model proposed by Cohen et al. (1990) has difficulty accounting for the time course of the Stroop effects. The model predicts that the largest interference and facilitation scores in color naming should be obtained at the longest distractor-first SOAs, where most evidence for the right or wrong response is collected. Instead, empirically the interference increases when distractor-first SOAs become shorter and there is a flat pattern of facilitation (e.g., M. O. Glaser & Glaser, 1982). There is good evidence that habituation is not present to reduce the effects at the distractor-first SOAs, thereby producing the decreasing patterns predicted by Cohen et al. (1990): A new experiment showed that the amount of interference does not vary with repetition between SOA = -400 ms and 0 ms, contrary to what the habituation account predicts. Finally, the Cohen et al. (1990) model predicts no interference when the pathways are of comparable strength. However, in the color-color and word-word variants of the Stroop task, where there are two processes that are, by definition, of equal strength, the normal Stroop patterns are obtained (M. O. Glaser & Glaser, 1982; W. R. Glaser & Glaser, 1989). Also, the model has difficulty with the training data of MacLeod and Dunbar (1988).

Strength or an Architectural Difference Between Color Naming and Word Reading?

Training color naming to the extent that it becomes faster than reading aloud appears to be impossible (e.g., Brown, 1915; Cattell, 1886; Ligon, 1932), which suggests a difference in the architecture for reading and naming rather than a difference in pathway strength. Brown (1915) found that even after 10 days of practice, colors were still named 131 ms slower than their names were read. Of course, this could mean that reading is so much more practiced from the start that 10 days of training does not make up for the difference. However, this does not seem to be the case. Ligon (1932) tested hundreds of schoolchildren whose ages varied from 6 to 16 years (Grades 1-9). Color naming and word reading times decreased with age, but the difference between them remained constant at 286 ms. Thus, the difference is there already from the start (i.e., when children learn to read) and stays.

Color naming is conceptually driven, whereas reading aloud typically is not (e.g., Caplan, 1992; Shallice, 1988). Smith and Magee (1980) argued that in the picture-word task, pictures have priority access to meaning before pronunciation, whereas the reverse holds for words. The same appears to be true for colors and words. Thus, color naming seems to require an extra processing step. Furthermore, written words in alphabetical systems are intrinsically tied to their sounds, whereas colors are not. These structural differences not only explain why reading is faster than naming but also why this cannot be reversed by training. In contrast, when an architectural difference is absent, interference patterns may be reversed: I refer to the training study of MacLeod and Dunbar (1988) for an artificial version of this. Although the attractor and verification models that I discuss below differ in their

attentional mechanism, both assume an architectural difference between reading aloud and color naming.

Selection Through Attraction

Attractor models have been proposed for many cognitive phenomena. One of the most influential models of this type is the reading model of McClelland and Rumelhart (1981). Hinton and Shallice (1991) proposed an attractor model for reading disorders (i.e., acquired dyslexia). In the area of speech production, Stemberger (1985) was the first to propose an attractor model for spoken word production, which was modeled after McClelland and Rumelhart's reading model. Phaf et al. (1990) developed an attractor model for selective attention, called SLAM (for SeLective Attention Model), within the framework of Van der Heijden's (1992) general theory of attention. It follows the basic design characteristics of the reading model of McClelland and Rumelhart, adapted to the domain of action selection.

Phaf et al.'s (1990) model assumes an interactive-activation network consisting of input, hidden, and output nodes (see Figure 5). Input nodes represent words and colors in a particular spatial position (e.g., the word RED at a position below the color patch). Processing occurs through activation spreading from color input nodes via hidden nodes to output nodes, and directly from word input to output nodes, whereby nodes change their activation with time in a continuous, nonlinear manner. There are excitatory links between nodes representing compatible information, and there are inhibitory links between nodes standing for incompatible information. Selective attention to the color naming and reading tasks is achieved by adding extra external activation to all hidden color nodes for color naming and to all output nodes for word reading. The task activation is given from trial onset onward. On each simulated trial, word and color input is given to the network and activation cycles around from one unstable pattern to another until a stable pattern of activation is reached, depending on the inputs (e.g., color and color-naming task input). A response is chosen by

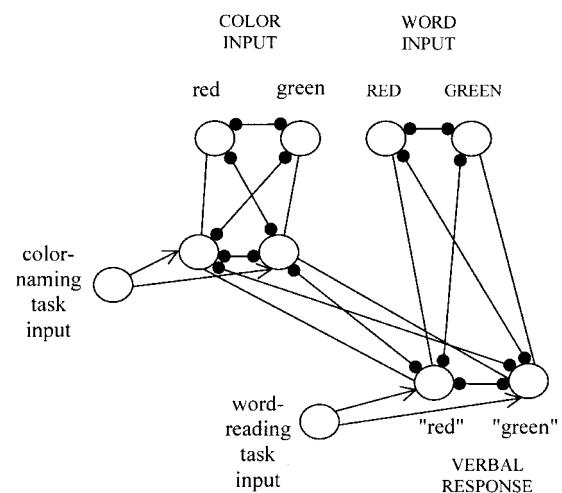


Figure 5. Network architecture of the model of Phaf et al. (1990). The lines with black circles denote bidirectional inhibitory connections, the plain lines denote bidirectional excitatory connections, and the lines with arrows indicate unidirectional excitatory connections.

feeding the activation of the response layer into a sampling and recovery procedure that stochastically favors the most highly activated response node.

In color naming, a preexposed distractor word puts the system in an attractor basin for the word rather than the color. The Stroop interference is explained as the extra time it takes to move the system from the basin for the word into the basin for the color name. The model predicts a constant SOA effect in color naming. After perceiving the word, the system quickly settles into a stable state of activation for the response corresponding to the word. By definition, the stable state does not vary with time; hence, making the distractor-first SOA longer or shorter has no effect until an SOA is used that is too short for the distractor to reach an attractor basin. Consequently, the amount of time it takes for the color name to overcome the interference from the word is constant. Figure 6 shows the model's predictions for the time course of the Stroop phenomenon from simulations by Phaf et al. (1990). Note that the model does not adequately account for the time course of the Stroop conflict in color naming that has been observed by M. O. Glaser and Glaser (1982) and others. Also, the model incorrectly predicts a small reverse Stroop effect. At all distractor-first SOAs, interference on reading in the model is over 25 ms, whereas the real data show no effect.

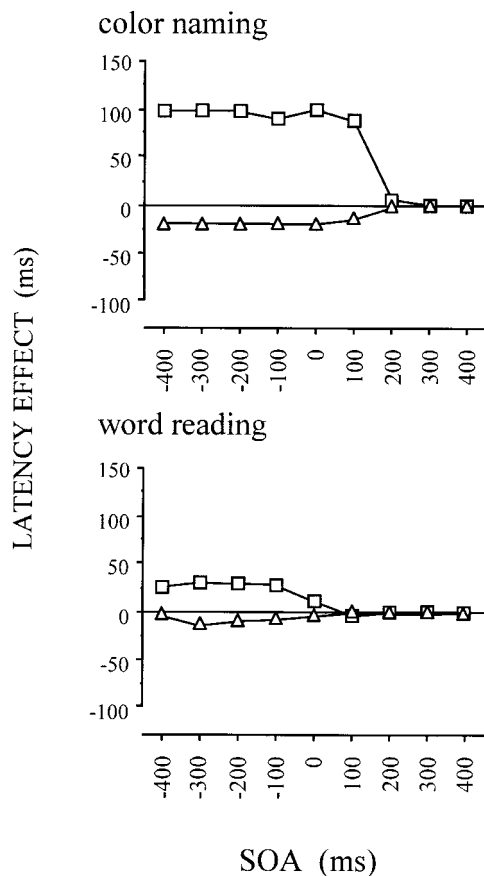


Figure 6. Time course of the color-word Stroop phenomenon as simulated by Phaf et al. (1990) for color naming and word reading. \square = incongruent; \triangle = congruent; SOA = stimulus onset asynchrony.

In line with Phaf et al. (1990), Cohen and Huston (1994) discussed an attractor version of the model proposed by Cohen et al. (1990). The behavior of this model is similar to that of Phaf et al.: The amount of interference at distractor-first SOAs is constant (see Cohen and Huston, 1994, Figure 18.11), contrary to the real data.

In the next few sections, I describe a third theoretical alternative, namely "classic" goal-referenced selection of verbal action, as implemented in WEAVER++. In subsequent sections, I show that selection by verification in WEAVER++ does account for the time course of Stroop interference and for many other findings.

Selection Through Verification

WEAVER++ is a model of word production that falls into the general class of "production rule system" models of cognition, which includes ACT (Anderson, 1983; Anderson & Lebiere, 1998), CAPS/READER (Just & Carpenter, 1992), EPIC (D. E. Meyer & Kieras, 1997), and Soar (Newell, 1990), among others. WEAVER++ is a so-called "hybrid" model in that it combines a symbolic declarative network and production rule system with activation spreading and activation-based production rule triggering and firing (cf. ACT). Many of the properties of the WEAVER++ model reflect that it was originally developed as an attempt to situate conceptually driven word production firmly in the context of sentence generation (Levelt et al., 1999a, 1999b; Roelofs, 1992, 1993). The model was designed to account simultaneously for a wide variety of findings (for a recent review, see Levelt et al., 1999b). Unfortunately, the trend in the word-production literature has moved away from trying to simultaneously account for a wide range of data in favor of a focus on a few findings (e.g., Caramazza, 1997; Cutting & Ferreira, 1999; Starreveld & La Heij, 1996).

WEAVER++ distinguishes between conceptual preparation, lemma retrieval, and word-form encoding, with the encoding of forms further divided into morphological, phonological, and phonetic encoding (Roelofs, 1997c). During conceptual preparation, concepts are selected and "flagged" as goal concepts. In lemma retrieval, a goal concept is used to activate and select a lemma, which is a memory representation of the syntactic properties of a word. For example, the lemma of *red* says that it can be used as an adjective. Lemma retrieval makes these properties available for syntactic encoding (the latter process can, e.g., make the lemma *red* the head of an adjective phrase). In word-form encoding, the selected goal lemma is used to activate and select the morphophonological properties of the word from memory to construct ("weave") an appropriate word form. For example, for *red*, the morpheme <red> and the segments /r/, /e/, and /d/ are activated and selected, and a motor program for [red] is generated. Finally, articulation processes execute the motor program, which yields overt speech. Lemma retrieval and word-form encoding are discrete processes in that only the word form of a selected lemma becomes activated and encoded (Levelt et al., 1991).

Naming a perceptual entity such as a color proceeds through the processing stages shown in Figure 7. First, there is the conceptual identification of the color based on perceptual input. The concept is flagged as goal concept (i.e., RED(X)). Second, the lemma of the corresponding word is activated and selected (i.e., *red*), in the Stroop literature often referred to as *response selection* (except that

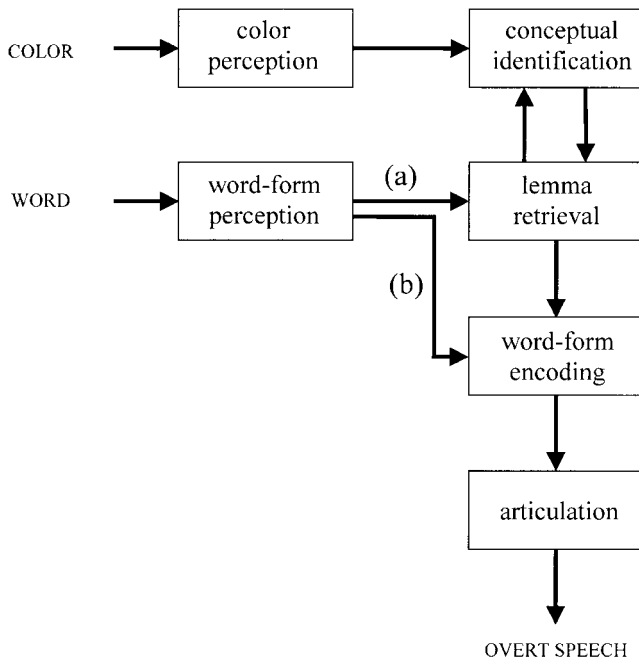


Figure 7. Levels of processing in WEAVER++. Word reading may involve lemma selection (Route a) or may not (Route b). Adapted from *Cognition*, 42, A. Roelofs, "A Spreading-Activation Theory of Lemma Retrieval in Speaking," p. 114, Figure 1, Copyright 1992, with permission from Elsevier Science.

here it involves lemmas, which is new). Third, the form properties of the word are activated and selected (i.e., [red]), called *response programming*. And finally, the name is articulated, called *response execution*. A perceived written word activates its lemma and its output form in parallel. Oral reading, which I discuss later, may be achieved by a shallow form-to-form route (e.g., from the orthographic form RED to the output form [red], Route b in Figure 7) or it may involve an extra step of lemma retrieval (i.e., from orthographic RED via the lemma *red* to [red], Route a in Figure 7), roughly corresponding to what is traditionally called the *semantic* route (e.g., Caplan, 1992; Shallice, 1988).

Although this architecture has been developed as a functional architecture on the basis of behavioral evidence (e.g., Levelt et al., 1999b; Roelofs, 1992), it is receiving increasing support from neuroimaging and electrophysiological studies (Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998; Van Turennout, Hagoort, & Brown, 1997, 1998). Indefrey and Levelt (2000) performed a meta-analysis of 58 brain imaging studies on word production in the literature, which provided a rough anatomical localization of this functional architecture in the brain. The studies included picture naming, verb generation (generating a use for a noun, e.g., saying "hit" to HAMMER), word reading, and pseudoword reading. As would be expected from the classic neurology literature and most later studies, the system is basically located in the left hemisphere. Visual and conceptual processing involves the occipital, ventro-temporal, and anterior frontal regions of the brain; the middle part of the left middle temporal gyrus seems to be involved with lemma retrieval (activity in these areas occurs within the first 275 ms after an object is presented). Next, activation spreads to

Wernicke's area, where the phonological code of the word seems to be retrieved; activation is then transmitted to the left middle superior temporal lobe and Broca's area for postlexical phonological processing such as syllabification (taking some 125 ms). Finally, phonetic encoding takes place (for the next 200 ms), with a contribution of the supplementary motor area and the cerebellum, while the sensory-motor areas control articulation.

The WEAVER++ model assumes that the mental lexicon is a network that stores declarative pieces of information about words. The network comprises three major strata, corresponding to the stages of conceptual preparation, lemma retrieval, and form encoding: a conceptual stratum, a syntactic stratum, and a word-form stratum, shown in Figure 8. The conceptual stratum represents concepts as nodes and labeled links in a semantic network, following Quillian's (1969) seminal work, Collins and Loftus (1975), and many others. For example, the concepts red and color are represented by the node RED(X) and the superordinate node COLOR(X), respectively. The labels on the links, such as IS-A, are omitted from Figure 8, because they are not important in the current article. The syntactic stratum contains lemma nodes, such as *red* and *color*, which are connected to nodes for their syntactic class (e.g., *red* can be used as an adjective) and to their concept nodes in the network. And finally, the word form stratum contains nodes representing morphemes, segments, and syllable programs. The form of monosyllables such as *red* presents the simplest case with one morpheme node <red>, segments such as /r/, /e/, and /d/, and one syllable program node [red]. Polysyllables such as *color* have their segments connected to more than one syllable program node; for *color*, these program nodes are [kʌ] and [lɔr]. And polymorphemic words such as *blackboard* have one lemma connected to more than one morpheme node; for *blackboard* these morphemes are <black> and <board>. For extensive discussions of the theoretical and empirical motivation of these assumptions, see Levelt (1989, 1992, 1999), Levelt et al. (1999a, 1999b), Roelofs (1992, 1993, 1994, 1996a, 1996b, 1996c, 1997a, 1997b, 1997c, 1998b, 1999), Roelofs et al. (1996, 1998), and Roelofs and Meyer (1998).

Information is retrieved from the network by spreading of activation. For example, a perceived color (e.g., red) activates the corresponding concept node (i.e., RED(X)) in the network. Activation then spreads through the network following a linear activation rule with a decay factor (see Appendix A). Each node sends a proportion of its activation to the nodes it is connected to. For example, RED(X) sends activation to other concepts such as GREEN(X) and also to its lemma node *red*. Selection of nodes is accomplished by production rules (e.g., Anderson, 1983; Anderson & Lebiere, 1998; Newell, 1990), which are associated with nodes in the network. The *left-hand side* of a production rule specifies a condition to be satisfied and the *right-hand side* specifies an action to be taken when the condition is met. A rule is triggered when "its" nodes become active. Application of a production rule to select a node is called *verification*. A lemma retrieval production rule selects a lemma if the connected concept is flagged as the goal concept. For example, *red* is selected for RED(X) when it is the goal concept, and *red* has reached a critical difference in activation compared with other lemmas. The actual moment in time of selection (i.e., the "firing" of the production rule) is determined by the ratio of activation of the lemma node and the sum of all the

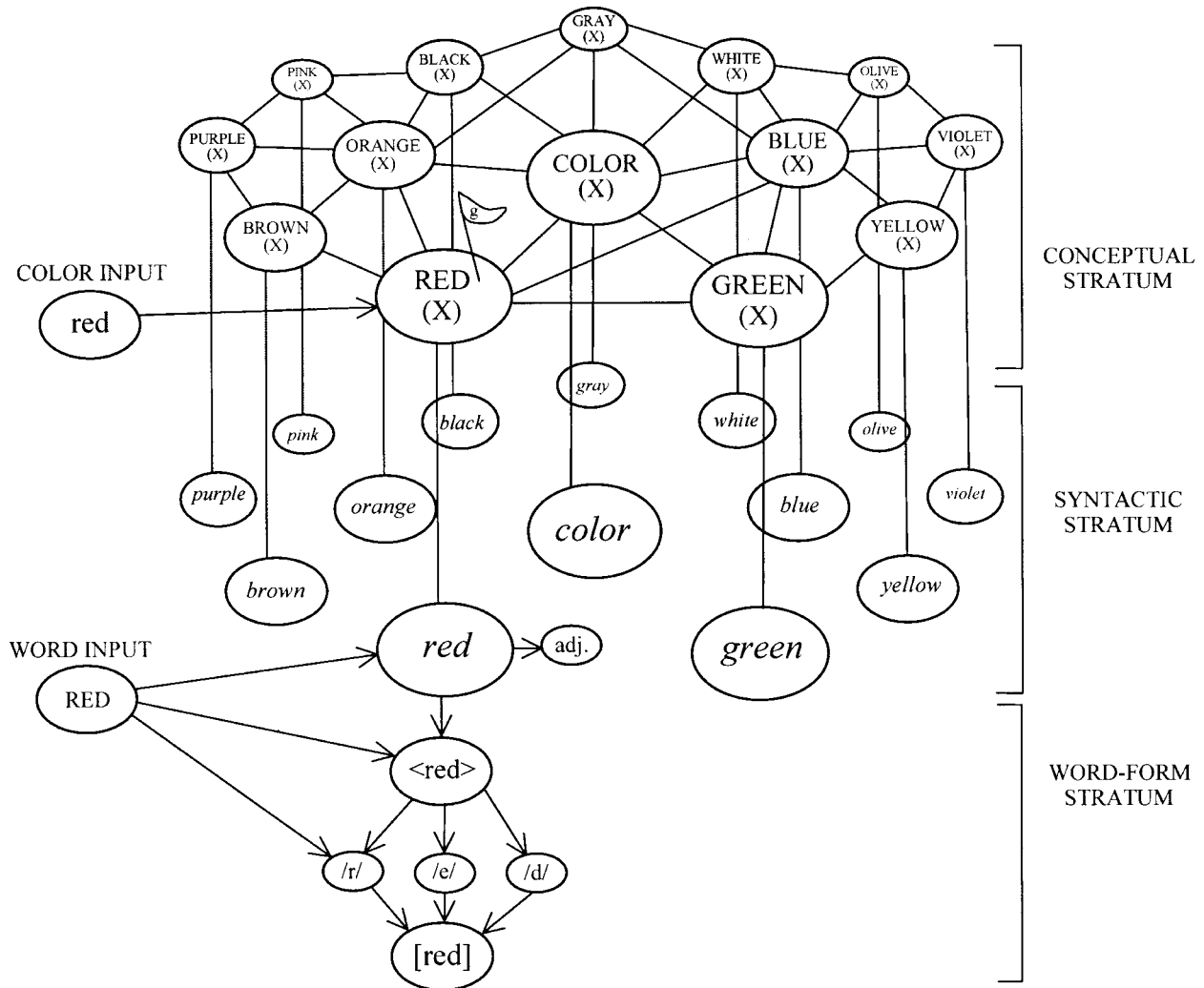


Figure 8. Network architecture of WEAVER++. The lines with arrows indicate unidirectional excitatory connections and the lines without arrows denote bidirectional excitatory connections. RED(X) has been flagged as goal concept. g = goal.

other lemma nodes. Thus, how fast a node is selected depends on how active the other nodes are (see Appendix A).

A selected lemma is flagged as the goal lemma. A morphological production rule selects the morpheme nodes that are connected to the selected, goal lemma (<red> is selected for *red*). Phonological production rules select the segments that are connected to the selected, goal morphemes (/r/, /e/, and /d/ for <red>) and syllabify the segments (e.g., /r/ is made syllable onset: onset(/r/)) to create a phonological word representation. Finally, phonetic production rules select syllable-based articulatory programs that are appropriately connected to the syllabified segments (i.e., [red] is selected for onset(/r/), nucleus(/e/) and coda(/d/)) that are part of the goal phonological-word representation (e.g., Levelt 1989; Levelt & Wheeldon, 1994). The moment of selection of syllable program nodes is determined by the ratio of activation of the target syllable program node and the sum of all the other syllable program nodes. How fast selection occurs depends on how active other nodes are. Once a concept has been designated as the goal

concept, planning the word proceeds fully automatically. Verification guarantees that the planning of a word is insulated from interfering cross-talk, for example, from the concurrent planning of other words in connected speech or from hearing or seeing words (for an extensive discussion, see Roelofs, 1997c). In summary, similar to ACT (Anderson, 1983; Anderson & Lebiere, 1998), WEAVER++ combines a symbolic network and a goal-driven production rule system with activation spreading and activation-based production rule triggering and firing.

Basic Assumptions About Stroop Performance

Most Stroop studies use from two to six colors (MacLeod, 1991). WEAVER++'s Stroop performance has been assessed by computer simulations using three colors (red, green, and blue). To make sure that the results do not depend on the size and content of the network, I also ran simulations that embedded these three colors and words in a larger network. The larger network consisted

of the 13 basic English color terms from the system developed by the National Bureau of Standards plus their hyperonym *color*. The terms are *red, orange, yellow, green, blue, violet, purple, pink, brown, and olive* for the chromatic colors, and *black, gray, and white* for the achromatic colors. These 14 terms, plus some adjectives (e.g., *dark*) and a rule involving the use of *-ish*, make up the basic terminology of the English color system (Miller & Johnson-Laird, 1976). Secondary color terms (e.g., *greenish blue*) are defined in terms of these. It appeared that both small and larger networks give equivalent simulation results, as I show below. Also, equivalent results were obtained when the simulations included lemma retrieval only versus when they included lemma retrieval and word-form encoding. Thus, in the model, the crucial Stroop conflict in color naming occurs during accessing the lemma. After showing that the model scales up by using a larger network and by including word-form encoding, I use the small network with lemma retrieval only in all further simulations of color naming.

In WEAVER++, words are planned by spreading activation and goal-factored production rule application. Planning is controlled by task-independent “slave” production rules (accomplishing the appropriate selections in lemma retrieval and word-form encoding) and task-dependent “executive” production rules. The executive rules set the overall task goal (e.g., naming the color), and they achieve task-relevant selections in word planning (e.g., selecting the color concept rather than its superordinate) as well as perceptual input control (e.g., ignoring the word). I discuss and motivate these aspects of the model in turn.

First, a distinction is made in the model between executive and slave aspects of performing a task (cf. Baddeley, 1986). There are task-dependent production rules like P1 that refer to the overall goal and define general aspects of the task (paraphrased in an informal form), and task-independent rules like P2 that accomplish the actual lexical selections, such as lemma selection.² The latter rules are stored with each lemma in memory and may operate in parallel.

- P1. IF the goal is to say the name of the color,
and the concept is the color of the stimulus
THEN select the concept,
and flag the concept as goal concept,
and enhance its activation.
- P2. IF RED(X) is flagged as goal concept,
and the activation of *red* exceeds threshold
THEN select *red*.

This distinction between executive and slave aspects of performing a task is in line with Norman and Shallice’s (1986) distinction between the operations of what they call a “supervisory attentional system” and “contention scheduling” (pp. 5–6). Contention scheduling concerns the routine selection of actions, which is decentralized. Lemma selection is such a highly practiced, routine operation. By contrast, a supervisory attentional system like that proposed by Norman and Shallice sets nonroutine tasks and achieves nonroutine selections. Performance of a Stroop task requires such a system, one that ensures that task-specific instructions are followed. As with the task nodes of Cohen et al. (1990) and Phaf et al. (1990), Norman and Shallice’s supervisory system

operates by biasing the activation levels of the units involved in contention scheduling (the output layer of Cohen et al., 1990, and Phaf et al., 1990). In contrast, I propose verification as the mechanism. Norman and Shallice associate the supervisory system with the functioning of the frontal lobes of the brain and assume that frontal lobe lesions yield impairments of the supervisory system. The deficits of patients with frontal lobe lesions are quite diverse (including repetitive, disintegrated, and pointless behavior), but many deficits share in common a difficulty with the formulation and maintenance of goals in the face of environmental distractors (e.g., Duncan, 1986; Jahanshahi & Frith, 1998; Shallice, 1988; Shallice & Burgess, 1991).

Second, in addition to goal setting and task-relevant selection, executive production rules achieve perceptual input control. Whereas Norman and Shallice (1986) concentrate on the role of an executive system in the selection of actions, Posner and colleagues (e.g., Posner & Raichle, 1994) point to a crucial role of such a system in controlling perception. In the control of perception, they distinguish between an executive system centered on the anterior cingulate cortex (part of the frontal lobes), called the *anterior attention system*, and a *posterior attention system* that is responsible for input selection (e.g., Rafal & Henik, 1994).

The idea is that the executive attentional system of the anterior cingulate tries to achieve selective perceptual processing in other brain regions given the current goals and task demands. Interactions of the anterior system with posterior regions of the brain seem to selectively increase the gain of one spatial location or one type of perceptual input over another. In the color-word Stroop task, the system controls the gain of color information relative to visual word-form information in the extrastriate visual cortex (e.g., S. E. Peterson, Fox, Snyder, & Raichle, 1990). Measurements of evoked brain potentials have shown that cortical responses to attended and ignored stimulus locations may diverge already at about 70–90 ms after stimulus onset; attention to nonspatial stimulus dimensions such as color is indexed by longer latency event-related brain potential components in the 150–350 ms range. The activation of the attended information increases, whereas that of the ignored information decreases (e.g., Hillyard, Mangun, Woldorff, & Luck, 1995). Evidence for the involvement of the anterior cingulate in resolving the Stroop conflict comes from neuroimaging studies revealing blood flow increase in this cortical area during incongruent and congruent trials compared with neutral trials (for reviews, see Bush, Luu, & Posner, 2000; MacLeod & MacDonald, 2000; Posner & DiGirolamo, 2000). However, the executive system involved cannot uniquely be associated with the cingulate cortex. Janer and Pardo (1991) examined Stroop performance by a patient who had the anterior cingulate bilaterally removed by operation. When tested 2 weeks after her operation, performance appeared to be deteriorated, but the deficits were absent 8 months later. Supposedly, other brain areas can take over some of the executive functions.

² In Roelofs (1992, 1993), I proposed to achieve lemma selection in Stroop-like tasks by so-called intersection search (e.g., Charniak & McDermott, 1985). Levelt et al. (1999b) assumed intersection search plus verification. In hindsight, this was overkill. I therefore propose here to simplify the earlier proposals by assuming verification only.

Thus, whereas the anterior system seems to be involved with managing goals for action, the posterior system, in conjunction with the anterior system, is responsible for input selection. Clearly, input selection cannot involve selecting a location and increasing the gain of all information coming from that location because the word and color are spatially integrated in a classic Stroop stimulus. So-called object-based selection (e.g., Kanwisher & Driver, 1992) does not work here either because evidence for it suggests facilitation of all information within the spatial boundaries of the object, which does not separate the word from the color. As Carr (1999) suggested, input selection is presumably based on whether the dimension is the color or the word, and this is also assumed for the model (e.g., Roelofs, 1992; Levelt et al., 1999b). Colors and visual word forms are processed in different areas of the extrastriate cortex of the brain (e.g., Gazzaniga et al., 1998; Posner & Raichle, 1994). Input control may favor the color channel over the visual word-form channel. In WEAVER++, input control is achieved by a production rule (P3) that lets the network receive perceptual input much longer for the relevant than for the irrelevant stimulus dimension. The rule operates with a latency that I have referred to as *distractor duration* (cf. Roelofs, 1992, p. 120):

- P3. IF the goal is to say the name of the color,
and input is received from a word
THEN block out the word input.

To summarize, color naming is achieved through color perception, conceptual identification, lemma retrieval, word-form encoding, and articulatory processing, whereas word reading minimally involves word-form perception, word-form encoding, and articulatory processing. This word planning system is controlled by an executive system, centered on the anterior cingulate, which sets and maintains the overall task goal. Furthermore, the executive system achieves input control. Figure 9 illustrates how the word planning and executive control processes relate to areas of the human brain, following Indefrey and Levelt (2000) and Posner and colleagues (e.g., Bush et al., 2000; Posner & DiGirolamo, 2000; Posner & Raichle, 1994).

The Computer Simulations

In what follows, WEAVER++'s Stroop performance is evaluated by means of computer simulation. For the simulations of lemma retrieval, which are central, there was a basic set of seven parameters: (a) the spreading rate r_C within the conceptual stratum, (b) the spreading rate r_L within the lemma stratum, (c) the decay rate d , (d) the size $extin$ of the external input, (e) the duration du of the distractor input, (f) the response threshold $critdiff$, and (g) the cycle time Δt , the real-time equivalent of a time step (which relates model time to real time in a linear fashion). All parameter values were identical to those in the simulations of lemma retrieval reported in earlier articles on the model (Levelt et al., 1999b; Roelofs, 1992, 1993, 1996a, 1997a; Roelofs et al., 1996). Two parameter values were changed slightly for Stroop performance to fine-tune the fit of the model to the current data: The distractor duration du was set to 100 ms and the response threshold $critdiff$ to 1.6. Parameter du determines the gain of the distractor input relative to the target input and is thus a central attentional parameter in the model. It reflects input selection. In addition to these

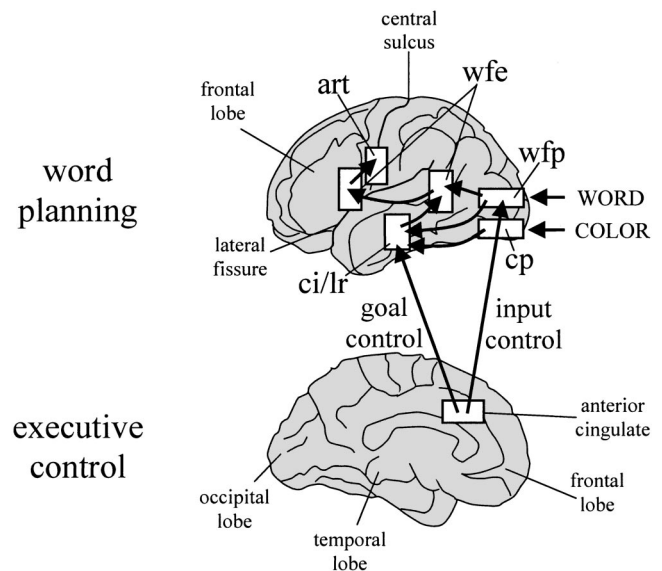


Figure 9. Word planning and executive control in the Stroop task. Lateral view (top panel) and medial view (bottom panel) of the left hemisphere of the human brain. The word-planning system achieves color naming through color perception (cp), conceptual identification (ci), lemma retrieval (lr), word-form encoding (wfe), and articulatory processing (art); word-form perception (wfp) activates lemmas and word forms in parallel. Word reading minimally involves wfp, wfe, and art. The executive system centered on the anterior cingulate achieves goal and input control.

seven “old” parameters, a new one, *bias*, was included. This parameter is a constant by which the activation of a node in the selection ratio is multiplied to simulate target anticipation.

The parameters were kept fixed in all simulations reported below, except when an experimental or group manipulation was explicitly linked to a parameter variation. For example, because spatial information becomes available faster than nonspatial information, locational certainty of the components of a Stroop stimulus may decrease parameter du ; in contrast, frontal cortex impairment may lead to problems in controlling the input and hence increase du . The tasks, participants, materials, design, and details of experimental procedures differed between studies in the literature, so sometimes identical conditions yielded numerically different results between studies. In particular, such differences occurred when the task was changed from color naming to picture naming, manual responding, or oral reading. To accommodate such differences, one or another parameter was fine-tuned in some of the simulations. This was done in 4 of the 16 simulations reported. In particular, $critdiff$ and $bias$ were slightly adjusted for picture naming and manual responding, and $critdiff$ and du for oral reading (see Appendix A for details). Later, I show that the critical behavior of the model is robust under parameter variation. In total, the target data sets in the present article contain 250 data points from 16 different, classic studies in the literature, which are simultaneously accounted for using three free parameters only (i.e., $critdiff$, du , and $bias$) accommodating task changes. Thus, the fit of the model to the data is nontrivial; there are many more data points than free parameters. Furthermore, it should be mentioned that I have made no attempt to fit the data quantitatively. Appendix

A gives the equations for the spreading of activation, the selection ratio, the expected latencies, and the parameter values.

The model is applied to findings from both the individual stimulus version of the Stroop task (where individual stimuli are presented and timed) and the original version of the task with stimulus cards. Whereas the average time to name individual Stroop stimuli is typically in the range of 500–700 ms and the conflict is often 100 ms, the production times are typically some 1.5 times longer for the card version of the Stroop task, and the conflict is often 400 ms or more. It appeared that the model accommodates these differences in effect size by scaling one parameter in the model by a constant (i.e., 4.0), namely the duration du . This suggests that it is more difficult to selectively modulate processing of task-irrelevant information in the card version than in the individual version of the Stroop task. A theoretically neutral alternative to scaling this parameter would be to scale the simulated Stroop scores themselves by this constant, which appeared to give equivalent results.

To illustrate the simulations, I go through a simulated trial. Assume that the color must be named of a red color patch on which the word GREEN is superimposed, with the word presented 100 ms before the color patch (the simulated SOA is -100 ms). The simulation starts with the lemma node of *green* receiving external activation (for du milliseconds, the latency of production rule P3). Activation spreads through the network, with the node *green* sending a proportion of its activation to GREEN(X), and this node in its turn to the other concept nodes. After the number of time steps that is the equivalent of 100 ms (the SOA), the concept node RED(X) receives external input from the color. Next, production rule P1 fires, RED(X) becomes flagged as goal concept, and its activation level is selectively enhanced. The input to the target concept is clamped until the word form has been planned. Empirical support for this comes from studies by Meyer and colleagues using an eye tracker to measure gaze duration during the naming of objects (e.g., A. S. Meyer, Sleiderink, & Levelt, 1998). These studies show that a speaker keeps fixating a perceived object until its phonological form has been prepared. After the response threshold of *red* is exceeded, production rule P2 fires, and *red* is selected as response. Although the threshold has been reached for *green* earlier, its production rule did not fire because one critical condition was not satisfied, namely GREEN(X) was not flagged as the goal concept.

Accounting for the Basic Stroop Findings (Findings 1–3)³

Interference Exceeds Facilitation (Finding 1)

Incongruent words interfere with color naming, whereas congruent words facilitate it. Figure 10 shows that WEAVERTT accounts for these findings. The real data shown are from Dyer (1971b). Many other studies have reported the same effects. In the incongruent condition (the word GREEN in red ink), it takes longer for the lemma of *red* to reach the response threshold than in the control condition because of the activation of the lemma of *green*. By contrast, in the congruent condition (RED in red ink), activation from the color and word converges on the target lemma *red*, with the consequence that this node reaches threshold earlier. The finding that Stroop interference exceeds facilitation is explained as a floor effect in speeding up responses. When exceeding the

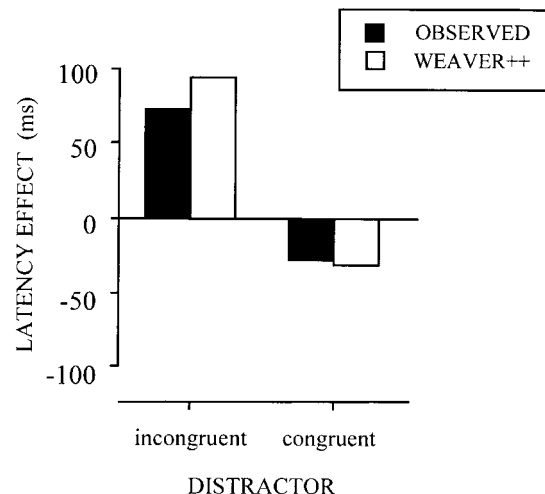


Figure 10. Basic latency effects in the congruent and incongruent conditions compared with the control condition: Observed effects from Dyer (1971b) and WEAVERTT simulations.

response threshold takes a minimal amount of time because of priming by the congruent word, the maximal amount of facilitation is reached. However, there is, in principle, no upper limit to delay exceeding the threshold by an incongruent word.

Response Set Effects (Finding 2)

There is less Stroop interference in color naming when the color words are not names of the target colors. For example, when the colors are red and green, the word BLUE in red ink produces less interference than the word GREEN in red ink (e.g., W. R. Glaser & Glaser, 1989; Klein, 1964; Proctor, 1978).

WEAVERTT assumes that potential responses are flagged as being part of the response set and that the response criterion (critdiff) may hold for response set members only. Making the assumption that attention in response selection can be selectively allocated to a restricted set of words appears to be very fruitful in explaining semantic effects in naming versus categorization tasks (Roelofs, 1992, 1993). For example, in saying “cat” to a pictured cat (picture naming), superimposing the semantically related distractor word DOG on the picture delays responding compared with a semantically unrelated word like HOUSE (e.g., W. R. Glaser & Dtingelhoff, 1984). However, when the picture of a cat must be categorized by saying the hyperonym *animal*, then the distractor DOG speeds up naming relative to the distractor HOUSE. The same is observed when the word must be categorized: Compared with a pictured house, a pictured dog speeds up saying “animal” to the word CAT (W. R. Glaser & Dtingelhoff, 1984; W. R. Glaser & Glaser, 1989). Elsewhere (in Roelofs, 1992, 1993), I have shown that the semantic interference and facilitation effects and their exact time course can be explained if one assumes that the response criterion in conceptually driven naming may hold for potential responses only (e.g., for hyperonyms in categorizing). By

³ The finding numbers (e.g., Finding 1) refer to the numbers of the Stroop findings in Table 1.

contrast, models that assume some kind of fixed competition (e.g., Cutting & Ferreira, 1999; Starreveld & La Heij, 1996) have difficulty accounting for these findings (see Levelt et al., 1999b; Roelofs, 1993).

For the color-word Stroop task, response set flagging explains why interference is attenuated for words that are not part of the response set. However, these distractors do yield interference. If a word is not a response, it nevertheless activates responses that are part of the response set because their concept nodes are connected at the conceptual stratum. For example, if “red” and “green” (but not “blue”) are potential responses, and *red* is the target, BLUE produces interference on color naming because it activates *green*, which is a competitor to *red*. Figure 11 shows that the model accounts for the response set effect. The real data, which are typical, are from Proctor (1978).

No Reverse Stroop Effect (Finding 3)

The model yields no interference and facilitation from colors on word reading. In the model, word reading can be accomplished by a fast, “shallow” mapping of the orthographic code of the word onto the corresponding articulatory program (Route b in Figure 7) with the following task production rule:

- P4. IF the goal is to say the name of the word,
and the morpheme is the name of the stimulus
THEN select the morpheme,
and flag the morpheme as goal morpheme.

This is followed by the normal construction of the word form. Because a lemma is not selected, Route a in Figure 7 does not control the reading response. The irrelevant color activates its lemma, but this does not yield activation of the corresponding word form because this depends on selection of the lemma in the model (Levelt et al., 1991, 1999a, 1999b), which is task dependent. Because the form of the name of the color is not active, planning the target form is unaffected by the color. Thus, by assuming (a) a shallow form-to-form mapping for reading aloud and (b) word-

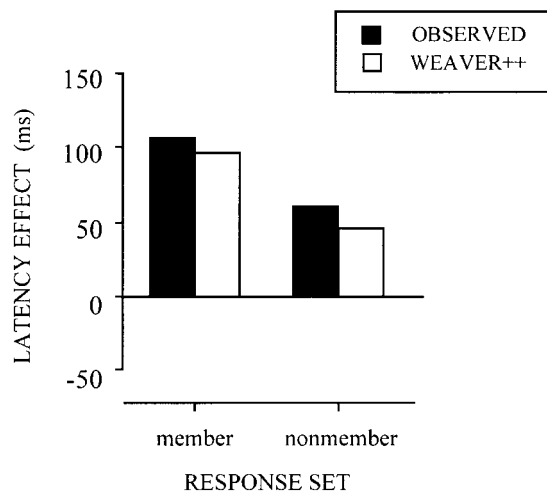


Figure 11. Basic latency effect of response set: Observed effects from Proctor (1978) and WEAVERTT simulations.

form activation by selected lemmas only (resulting in task-dependent activation spreading), WEAVERTT explains why there is no reverse Stroop effect.

How critical is the assumption that only selected lemmas activate their forms? Examination of this issue by simulation revealed that it is crucial for WEAVERTT's account of the absence of a reverse Stroop effect that activation spreading from the lemma level to the word-form level is functionally discrete: Only the word form of a selected lemma should become significantly active. Functional discreteness may happen under certain conditions without WEAVERTT's current architectural discreteness.⁴ WEAVERTT yields no reverse Stroop effect if activation spreads from nonselected lemmas to word forms provided that the spreading rate between lemmas and forms is lower than 40% of the spreading rate at the form level. Thus, functional discreteness may occur with weak cascading of activation from nonselected lemmas. Moreover, the reverse Stroop effect is absent if, in addition, backward links are included in the network provided that the backward spreading rate is lower than 45% of the spreading rate at the form level. To conclude, functional discreteness may arise in WEAVERTT when only selected lemmas activate their word forms (i.e., the current discrete architecture) or with weak backward spreading or forward cascading of activation from nonselected items (i.e., a “quasi” discrete architecture). The fact that the lemma and word-form levels are functionally discrete provides sufficient explanation for the absence of a reverse Stroop effect (for further discussion, see Levelt et al., 1999a, 1999b).

Empirical support for the assumption that only selected lemmas significantly activate their forms comes from double-task experiments. Levelt et al. (1991) asked participants to name pictured objects. On one third of the trials (the critical ones), a spoken probe was presented at a certain SOA after the picture, and participants had to perform a lexical decision on this probe. R. R. Peterson and Savoy (1998) also asked participants to name pictures, but on the critical trials in their study written words were presented that had to be read aloud. The lexical decision and reading latencies showed that in naming a perceptually given entity, there is no detectable activation of the word forms of nonsynonymous semantic relatives (i.e., fellow category members) of the target. For example, in naming a pictured cat, the lexical decision latencies and reading times for CAP (related in form to *cat*) and DOG (semantic alternative of *cat*) were affected, but those for DOLL (related in form to the semantic alternative *dog*) were not influenced at all. This suggests that in naming a cat, there is lemma activation for *cat* and *dog* and word-form activation for *cat*, but the word form of *dog* is not significantly activated. Can this be generalized to colors (which is crucial for the account of the absence of a reverse color-word Stroop effect)?

Below, I report an experiment that replicated Levelt et al. (1991) and R. R. Peterson and Savoy (1998) using the color-word Stroop task. Experiment 2A tested whether in naming a color—for example, red—distractor word RENT (a form relative of the congruent word RED) yields facilitation, and distractor word GREED (a form relative of the incongruent word GREEN) yields interference relative to an unrelated word, replicating the normal Stroop conflict. The

⁴ I am indebted to David Plaut (personal communication, November 21, 2000) for the terminology.

SOAs were 0, +100, and +200 ms. Experiment 2B tested the critical prediction that when colors have to be named and on one third of the trials a word has to be read aloud (presented 100 or 200 ms after the color), a red patch facilitates reading RENT but has no effect on reading GREED. Presumably, facilitation will be obtained at a late SOA only (+200 ms), when planning the color name has reached the form level, and the reading of RENT may benefit from the activation of the segments that are shared with the color name (i.e., /r/ and /e/). Such a replication of Levelt et al. (1991) and R. R. Peterson and Savoy (1998) would suggest that in naming a red color patch, the lemmas of *red* and *green* become active as well as the word form of *red*, but not the word form of *green*, supporting functional discreteness.

Experiment 2A: Color Naming With Form Relatives as Distractors

Method

Participants. Experiment 2A was carried out with 12 paid participants from the pool of the Max Planck Institute. All participants were young adults who were native speakers of Dutch. None of them had participated in Experiment 1.

Materials and design. The target stimuli were the colors red, green, and blue and the corresponding Dutch color words *rood*, *groen*, and *blauw*. The distractors were form relatives of the target words, namely ROOF (“robbery”), GROEP (“group”), and BLAAM (“blame”). They occurred in target–distractor pairings where they were directly related to the form of the target (e.g., distractor ROOF–target *rood*), the *direct* condition, or in pairings where they were related to the form of a response alternative (e.g., GROEP–*rood* and BLAAM–*rood*), the *mediated* condition. The control condition consisted of the color-neutral, form-unrelated words HUUR (“rent”), KNIP (“snip”), and SPUIT (“squirt”). These unrelated words rather than a series of Xs were used as controls because the same materials served as targets for reading aloud in Experiment 2B (Xs cannot be read aloud). The presentation sizes of the color patches and written words were the same as in Experiment 1.

There were two independent variables, which were varied within participants. The first independent variable was SOA with three levels: 0, +100, and +200 ms. SOA was varied within participants but between trial blocks. The order of SOA blocks was counterbalanced across participants. The second independent variable was distractor. In the stimulus set, three direct, six mediated, and nine control pairings were possible. Each of the direct, mediated, and control pairings occurred respectively four times, two times, and once within an SOA block. Thus, for each SOA, the participants received a direct stimulus 12 times, a mediated stimulus 12 times, and a control stimulus 9 times. The stimuli were presented in random order.

Procedure, apparatus, and analysis. This was the same as in Experiment 1. Participants were asked to name the colors and to ignore the written words. After the participant had read the instructions, a block of nine practice trials with SOA = 100 ms was administered, which was followed by the three experimental SOA blocks. The latencies and errors were submitted to ANOVAs with the crossed within-participant variables distractor and SOA.

Results and Discussion

Table 3 gives the means and standard deviations of the production latencies and the mean error percentages for the distractor by SOA cells. For all SOAs, the latencies were faster in the direct than in the control condition, and slower in the mediated than in the control condition. Figure 12 shows the latency effects per SOA.

Table 3
Means (RTs, in Milliseconds) and Standard Deviations (SDs) of the Production Latencies and Mean Error Percentages (E%) as a Function of Task, Distractor/Probe, and Stimulus Onset Asynchrony (SOA) for Experiment 2

Task	SOA (ms)		
	0	100	200
Color naming: Distractor (Experiment 2A)			
Direct			
RT	553	584	571
SD	118	124	124
E%	0.7	4.2	1.4
Mediated			
RT	637	657	607
SD	156	167	170
E%	7.6	6.9	1.4
Control			
RT	608	634	586
SD	134	149	137
E%	2.8	3.7	0.9
Word reading: Probe (Experiment 2B)			
Direct			
RT		534	513
SD		72	82
E%		0.5	2.3
Mediated			
RT		542	536
SD		69	83
E%		2.3	2.3
Control			
RT		544	539
SD		75	97
E%		0.0	3.1
Color naming (Experiment 2B) ^a			
RT		597	623
SD		119	135
E%		2.4	2.7

Note. SOA of 0 ms was not used in Experiment 2B. RT = response time.
^a The SOA for color naming in Experiment 2B indicates the SOA of the reading trials.

Production latencies. The statistical analyses of the production times confirmed these observations. There was an overall effect of distractor, $F(2, 22) = 34.59$, $MSE = 1,072$, $p < .001$, but not of SOA, $F(2, 22) = 2.31$, $MSE = 6,805$, $p > .10$. Distractor did not interact with SOA, $F(4, 44) = 1.93$, $MSE = 1,049$, $p > .10$. The direct condition was faster than the control condition, $F(1, 11) = 48.75$, $MSE = 673$, $p < .001$. Pairwise comparisons showed that this held for the SOAs of 0 ms and +100 ms ($ps < .05$) but not for +200 ms. The mediated condition was slower than the control condition, $F(1, 11) = 5.91$, $MSE = 1,238$, $p < .05$. The effect of mediated distractors did not vary with SOA, $F(2, 22) = 0.02$, $MSE = 1,207$, $p > .50$.

Error rates. The statistical analyses of the error rates yielded no significant results. The statistics for distractor, SOA, and the interaction were, respectively, $F(2, 22) = 1.49$, $MSE = 0.01$, $p > .10$; $F(2, 22) = 3.18$, $MSE = 0.004$, $p > .05$; and $F(4, 44) = 1.16$, $MSE = 0.003$, $p > .10$.

To summarize, the experiment shows that relative to unrelated control words, form relatives of targets in color naming yield Stroop interference in the mediated condition and facilitation in the

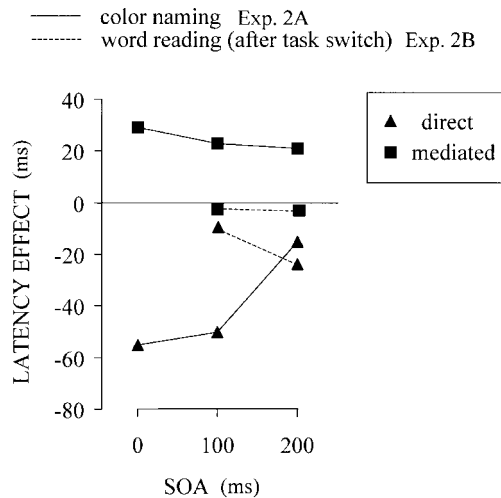


Figure 12. Latency effect of direct and mediated stimuli compared with control stimuli as a function of task and stimulus onset asynchrony (SOA) in Experiment 2. Experiment 2A used color naming; Experiment 2B used word reading. Exp. = experiment.

direct condition. This shows that the stimuli are effective in producing a Stroop conflict. In the experiment, interference was less than facilitation, which is presumably because unrelated words constitute a slower control condition than a series of xs (cf. Klein, 1964). Experiment 2B tested whether the colors activate these form relatives in a Stroop replication of Levelt et al. (1991) and R. R. Peterson and Savoy (1998).

Experiment 2B: Word Reading in the Context of Color Name Preparation

Method

Participants. Experiment 2B was carried out with 36 paid participants from the pool of the Max Planck Institute. None of them had participated in the earlier experiments.

Materials and design. The materials were the same as in Experiment 2A. There were now two types of trials, color naming and reading trials, which occurred in random order. As in the studies of Levelt et al. (1991) and R. R. Peterson and Savoy (1998), one third of the trials involved oral reading, and the remaining trials involved color naming. On a color naming trial, one of the colors red, green, or blue was presented but no probe word, and the color had to be named. On a reading trial, one of the colors was presented but also one of the distractor words of Experiment 2A, which had to be read aloud. A word was presented after the color so that participants were already preparing the color name when the probe word appeared (cf. Levelt et al., 1991; R. R. Peterson & Savoy, 1998). The probes were ROOF, GROEP, and BLAAM (form relatives of the color words *rood*, *groen*, and *blauw*, respectively) and HUUR, KNIP, and SPUIT (unrelated control words).

There were two independent variables. The first variable was SOA with two levels: +100 and +200 ms. SOA was varied between participants. For each SOA, 18 participants were tested. The second variable was probe. For the reading trials, three direct, six mediated, and nine control color-word pairings were possible, making up the direct, mediated, and control condition. Each of the direct, mediated, and control pairings occurred respectively four times, two times, and once per SOA. Thus, participants received a direct stimulus 12 times, a mediated stimulus 12 times, and a control

stimulus nine times. On the 66 other trials, only a color was presented, which had to be named. The 99 stimuli were presented in random order.

Procedure, apparatus, and analysis. For the most part, this was the same as in Experiments 1 and 2A. Participants were asked to name the color patches. They were told that on some of the trials, a word would appear shortly after the color. When a word appeared, they had to read it aloud instead of naming the color patch. After the participant had read the instructions, a block of 27 practice trials was administered. On nine of these trials, a probe word was presented with SOA = +100 ms or +200 ms, depending on the SOA group. The practice trials were followed by the experimental trials. The latencies and errors were submitted to ANOVAs with the crossed variables probe and SOA, with SOA tested between participants.

Results and Discussion

Table 3 gives the means and standard deviations of the production latencies and the mean error percentages for the probe by SOA cells. For SOA = +100 ms, the latencies in the direct, mediated, and control conditions were almost the same. For SOA = +200 ms, the latencies were faster in the direct than in the control condition, but the mediated and control condition did not differ. Figure 12 shows the latency effects per SOA.

Production latencies. The statistical analyses of the production latencies confirmed these observations. There was a main effect of probe, $F(2, 68) = 6.63$, $MSE = 567$, $p < .005$, but not of SOA, $F(1, 34) = 0.55$, $MSE = 5,223$, $p > .10$. For SOA = +100 ms, there was no effect of probe, $F(2, 34) = 1.32$, $MSE = 427$, $p > .10$, but SOA = +200 ms showed a clear probe effect, $F(2,34) = 5.62$, $MSE = 706$, $p < .01$. For this latter SOA, the direct condition was faster than the control condition, $F(1, 17) = 7.40$, $MSE = 954$, $p < .05$, but the mediated condition did not differ from the control condition, $F(1, 17) = 0.26$, $MSE = 1,024$, $p > .50$.

Error rates. The statistical analyses of the error rates yielded no significant results. The statistics for distractor, SOA, and the interaction were, respectively, $F(2, 68) = 0.64$, $MSE = 0.001$, $p > .50$; $F(1, 34) = 3.37$, $MSE = 0.002$, $p > .05$; and $F(2, 68) = 1.66$, $MSE = 0.001$, $p > .10$.

To conclude, the experiment shows that in a task-switch situation, colors yield facilitation in the direct condition but no effect at all in the mediated condition. The experiment exactly replicates Levelt et al. (1991) and R. R. Peterson and Savoy (1998) using the color-word Stroop task. Form relatives of the target become activated (e.g., *roof* in preparing to say “rood”), but form relatives of semantic alternatives remain inactive (e.g., *groep*, related in form to the color word *groen*, in preparing to say “rood”). This suggests that there are two stages of lexical access and that these stages are functionally discrete.

Accounting for the Time Course of the Stroop Effect (Findings 4–6)

Expectancy-Induced Versus Automatic Priming

Maximal impact of incongruent words in color naming is empirically observed when the words appear within 100 ms of the colors (e.g., M. O. Glaser & Glaser, 1982). Simulations showed that interference in WEAVER++ also peaks around SOA = 0 ms but that there is no effect at long distractor-first SOAs. This was to be expected if a distinction holds between automatic and

expectancy-induced verbal priming as argued by M. O. Glaser and Glaser (1982), Neely (1977), Posner and Snyder (1975), and many others. WEAVER++ includes automatic spreading of activation but has no expectancies. Expectancies can be captured, though, by introducing a response bias in the model. A preexposed distractor momentarily increases the weight of its lemma node in the selection ratio (this is the bias parameter, set to 2.0); if the distractor and response coincide, the response threshold is lowered (set to 0.0). This manipulation is applied in all simulations, assuming that it takes a preexposure time of 100 ms or more for expectancies to develop. The literature often assumes longer preexposure times for verbal anticipation (cf. Neely, 1991), namely 250 ms. However, this larger value concerns the anticipation of a different, associated word (e.g., the anticipation of the word *cat* on seeing the prime word *DOG*), whereas in the Stroop task the distractor word is expected to become the target (e.g., seeing the word *RED* leads to the anticipation of *red* as target). Taking a more conservative estimate for the anticipation time in the model gives equivalent simulation results.

Thus, the proposal here is that the bias factor of each lemma node has a default value of 1.0 and that the value for a preexposed word momentarily increases. This speeds up selection in the congruent condition but delays selection in the incongruent condition. In addition, several studies have demonstrated that the proportion of congruent and incongruent distractors per se also affects the magnitude of the effects in an experiment. In particular, increasing the proportion of incongruent trials reduces the Stroop interference, whereas increasing the congruent trials leads to more interference on the incongruent trials (e.g., M. O. Glaser & Glaser, 1982; Logan & Zbrodoff, 1979). This can be captured in the model by adjusting the default value of the bias parameter or the response threshold (cf. Roelofs, 1992, 1993): Decreasing the values reduces the magnitude of the effects.

Entity-Based Color Naming With Word-Color Stimuli (Findings 4 and 5)

In naming the color of a color-word Stroop stimulus, the interference in the incongruent condition increases when word preexposure and postexposure times become shorter, and interference peaks around $SOA = 0$ ms. In the congruent condition, a flat facilitation curve is observed for preexposed words. Figure 13 shows the classic SOA curves measured by M. O. Glaser and Glaser (1982) and those generated by WEAVER++. Like the real data, interference increases in the model when the distractor-first and distractor-second SOAs become shorter, with interference peaking at the SOA of 0 ms. The facilitation does not vary much with the preexposure time of the words. Unlike the real data of M. O. Glaser and Glaser, facilitation in the simulation slightly “peaks” around the SOA of 0 ms. However, in a replication of this study, Long and Lyman (1987) and W. R. Glaser and Glaser (1989) did observe most facilitation around $SOA = 0$ ms. As discussed, the amount of facilitation at short SOAs varies between studies (MacLeod, 1991). Figure 13 also shows the results for reading. As in the real data, there is no effect in the model of incongruent and congruent colors on reading color words because of the functional discreteness of stages and the form-to-form mapping in oral reading.

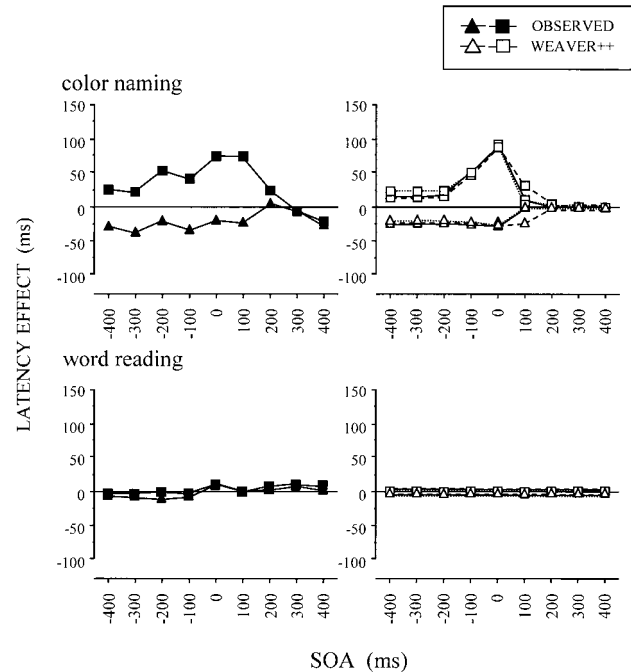


Figure 13. Time course of the color-word Stroop phenomenon for color naming and word reading: entity-based responding. The left-hand panels show the empirical data of M. O. Glaser and Glaser (1982, Experiment 1). The right-hand panels show WEAVER++ simulation results. The solid lines show the small-lexicon simulations with lemma retrieval, the dotted lines show the large-lexicon simulations, and the dashed lines show the simulations with lemma retrieval and word-form encoding. □, ■ = incongruent; △, ▲ = congruent; SOA = stimulus onset asynchrony.

Furthermore, Figure 13 shows the effect of varying the size of the lexicon and the effect of including word-form encoding. The solid lines indicate the results for the minimal lexicon with lemma retrieval only; the dotted lines give the results for the larger lexicon, and the dashed lines show the results of running the simulations with lemma retrieval and word-form encoding. The figure shows that including all basic color words of the English language in the model does not basically affect the simulation outcomes. Also, including form encoding does not change the outcomes. Thus, in the model, the Stroop conflict occurs during lemma retrieval, when the system has to select one verbal action over another.

The values of the parameters of WEAVER++ have been kept fixed across all applications of the model since its beginning (Roelofs, 1992) except for the parameters *du* and *critdiff*, which sometimes have been allowed to vary between studies to accommodate differences between tasks, participants, and materials. Thus, *du* and *critdiff* are the only real free parameters in the model. In addition, the current article introduces *bias* as a parameter. Importantly, maximal interference around $SOA = 0$ ms is a robust property of WEAVER++ and does not depend on the values of *du*, *critdiff*, or *bias*. Figure 14 shows how Stroop interference and facilitation vary as a function of the values of these parameters in the model.

The top panel of Figure 14 shows that maximal interference around $SOA = 0$ ms is still obtained when the value of *du*

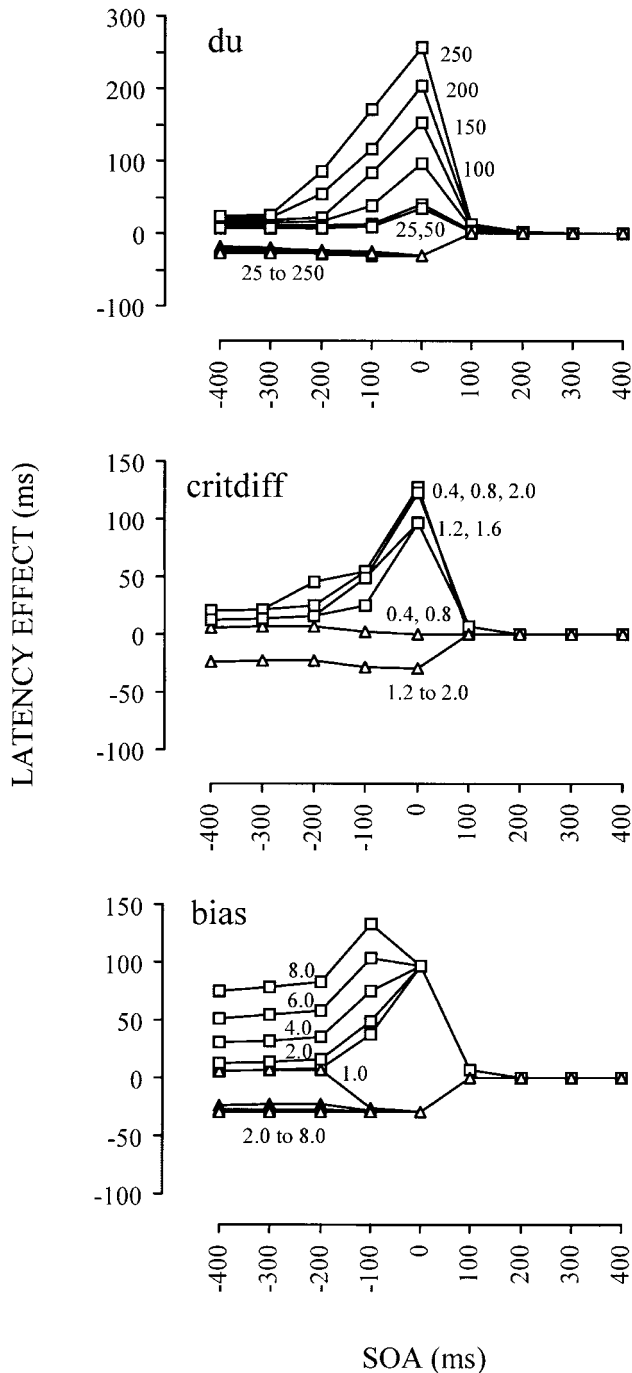


Figure 14. The effect of manipulating the free parameters in WEAV-ER++ on the time course of the Stroop phenomenon in the model: the duration (du) of the distractor input, the response threshold (critdiff), and bias (a constant by which the activation of a node in the selection ratio is multiplied to simulate target anticipation). \square = incongruent; \triangle = congruent; SOA = stimulus onset asynchrony.

increases or decreases (it also holds for values larger than 250 ms, which are not shown). Increasing du increases the Stroop interference while leaving facilitation virtually unaffected. Thus, factors that influence input control (du) in the model may have a differ-

ential effect on interference and facilitation. This corresponds to what was empirically observed by MacLeod (1998), who found that amount of practice and spatial separation affected interference but did not affect facilitation.

The middle panel of Figure 14 shows that maximal interference around SOA = 0 ms is still obtained when the value of the response threshold critdiff increases or decreases (it also holds for values larger than 2.0, which are not shown). The size of the threshold affects performance in the congruent and incongruent conditions, but also in the baseline control condition. With low values of critdiff (e.g., 0.4, 0.8), responding in the control condition is fast, and therefore interference is observed but no facilitation. With higher threshold values, both interference and facilitation are observed. Note that with critdiff set at 2.0 there is facilitation, but with it set at 0.8 there is not, although both values yield equivalent interference. Thus, according to the model, interference and facilitation do not necessarily go together. This corresponds to what is empirically observed: Studies report interference and facilitation or, sometimes, interference but no facilitation (MacLeod & MacDonald, 2000).

MacLeod (1998; MacLeod & MacDonald, 2000) argued that the independence of interference and facilitation in the Stroop task suggests that the effects may not have a common base. He proposed that facilitation is due to inadvertent reading of the distractor word on some trials (yielding apparent facilitation, because reading is faster than color naming). However, we saw that WEAV-ER++ produces the independence even though interference and facilitation have a common base in the model. Furthermore, there are a number of findings that challenge the inadvertent reading view of Stroop facilitation. First, facilitation is also observed in the color-color variant of the Stroop task (e.g., M. O. Glaser & Glaser, 1982), where no word can be inadvertently read aloud (although here matching is possible as a strategy). Second, in Experiment 2A above, RENT facilitated naming a red color patch. Because RENT facilitates the response "red," it is likely that RED also facilitates "red." Nonetheless, it seems important to examine to what extent facilitation is due to inadvertent reading or to convergence of activation. Convergence predicts that facilitation should persist under masked priming conditions and also with rhyming distractors (e.g., HEAD and the color name *red*), whereas the inadvertent reading hypothesis predicts that facilitation should disappear under these conditions, which may be tested in future research.

The bottom panel of Figure 14 shows that maximal interference around SOA = 0 ms is still obtained when the value of bias increases or decreases (it also holds for values larger than 8.0, which are not shown). With values of 6.0 or higher, the peak of interference shifts from SOA = 0 ms to SOA = -100 ms. Importantly, however, interference still increases with decreasing preexposure time.

The effects of manipulating the other parameters of WEAV-ER++ are similar to those from manipulating du, critdiff, and bias. It is important that maximal interference around SOA = 0 ms persists when the values of these other parameters increase or decrease. Increasing the size of the external input to the network (extin) or decreasing the decay rate (d) has the same effect as increasing critdiff: More external input from a distractor yields more interference, and the same holds when the distractor lemma activation persists longer because of reduced decay. Manipulating

the spreading rates at the conceptual stratum (r_c) and the syntactic stratum (r_1) also has the same effect as manipulating critdiff. Decreasing the spreading rate between concepts and lemmas leads to an increase of interference, because lemmas get less activation from their concepts. Consequently, the influence of distractor lemma activation is greater.

Why is there maximal interference in the model around SOA = 0 ms? Interference occurs in the model when the activation patterns of the target and distractor overlap, which happens when target and distractor are presented close together in time. Figure 15 shows how the activation of lemma nodes varies over time for color naming in WEAVER++. The trial is the word GREEN superimposed on a red color patch. The figure shows the activation levels for the lemma nodes of *red* and *green*. The activation curves are plotted for the SOAs of -100 and 0 ms. The activation of *red* is not much affected by the SOA at which GREEN is presented, but the activation of *green* itself depends very much on the SOA. With preexposure of the word GREEN by 100 ms, the activation of *green* is already decaying when the color is presented (at T = 0 ms). Because the activation of *green* is low, the response threshold for *red* is reached quickly. However, when color and word are presented simultaneously, *green* and *red* build up their activation in parallel, with *green* having a small head start. This is due to the direct access that words have to the syntactic stratum (the access by colors to this stratum is indirect, via the conceptual stratum). Therefore, the selection threshold for *red* is reached later, and a delay in responding is observed. Thus, in the model, whether interference is obtained critically depends on the timing of color and word. If the word is presented too early, there is no effect of the word on selecting the name of the color; the same holds when the word is presented too late. Note that verification in WEAV-ER++ guarantees correct performance. Although the activation of *green* exceeds that of *red* for the first 125 ms after color presentation (with SOA = 0 ms), the model correctly selects *red* and not *green*. This is because the concept RED(X) has been flagged as the goal concept (by production rule P1), and selection of a lemma makes explicit reference to this.

In the control condition, lemma retrieval in the model takes on average 105 ms, and word-form encoding (excluding syllabary access) takes 147 ms. Visual and phonetic-articulatory processing are not included in the model. On the basis of a survey of the

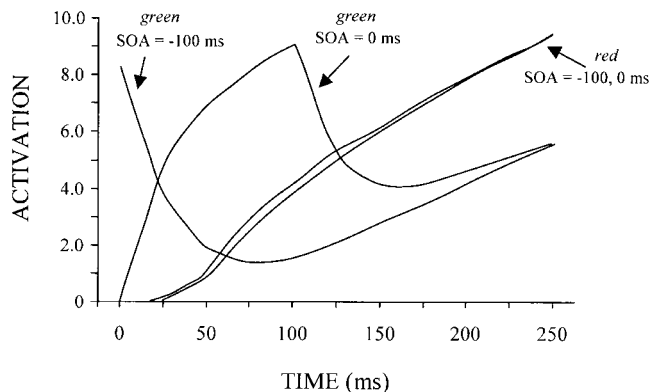


Figure 15. Activation curves for the lemma nodes of *red* and *green* in WEAV-ER++. SOA = stimulus onset asynchrony.

literature, Levelt et al. (1998) estimated that visual processing and accessing the corresponding lexical concept takes about 100–150 ms and phonetic-articulatory processing (including syllabary access) about 200 ms. This yields a lower estimate of 552 ms for color naming in the model. This corresponds well to the 540 ms for color naming observed by M. O. Glaser and Glaser (1982). Reading aloud in the Stroop task involves visual processing (but not concept access and lemma retrieval), word-form encoding, and phonetic-articulatory processing, which yields an estimate for the reading time in the model of about 447 ms. This corresponds to the 425 ms for word reading observed by M. O. Glaser and Glaser.

Although the error rates in a Stroop experiment are low (typically below 5%) and often do not reach statistical significance, the error tendencies should be briefly discussed. Empirically, most errors are made in the incongruent condition, and virtually no errors are made in the congruent and control conditions. This was also the case in Experiment 1 reported above (i.e., my replication of M. O. Glaser & Glaser, 1982). Furthermore, most errors are made at the shortest SOAs, that is, around SOA = 0 ms (e.g., Long & Lyman, 1987). Thus, the errors and the production latencies go in the same direction: There is no evidence for a speed-accuracy tradeoff. In WEAV-ER++, speech errors occur when verification fails or is skipped to gain speed, so that selection is based on the level of activation only (Levelt et al., 1999b; Roelofs, 1997c). Thus, most errors are made when the impact of incongruent words is largest, which is when the words appear within 100 ms of the colors, in agreement with the empirical findings. As Figure 15 illustrates, the activation of the lemma of *green* exceeds that of the target lemma *red* much longer for SOA = 0 ms than for SOA = -100 ms. Thus, more errors are predicted for the SOA of 0 than for -100 ms, in agreement with the empirical findings (cf. Long & Lyman, 1987, Experiment 2).

Picture-Word Task (Finding 6)

“A critical question is how directly picture-word and color-word findings relate to each other” (MacLeod, 1991, p. 168). WEAV-ER++ gives a unified account of performance in both tasks. Figure 16 shows the patterns of interference and facilitation of distractor words on picture naming. For example, participants had to name a pictured cat and ignore the incongruent word DOG or congruent word CAT. The empirical results are from W. R. Glaser and Dungelhoff (1984). Other studies have obtained the same findings (e.g., Lassen, 1975). The figure shows that WEAV-ER++ captures the empirical patterns.⁵ Figure 16 also shows the finding of W. R. Glaser and Dungelhoff that, like colors, pictures do not interfere with oral reading. This also holds for the model. These findings suggest that the conflict in the color-word and picture-word tasks is the same.

Order-Based Word Reading With Word-Word Stimuli (Finding 7)

Whereas words interfere with color naming, distracting colors do not yield a Stroop conflict in reading aloud. However, distract-

⁵ The model yields interference at SOA = -400 ms, whereas the real data show no effect. However, the model captures the overall empirical pattern: Interference increases with decreasing distractor preexposure, in agreement with the real data.

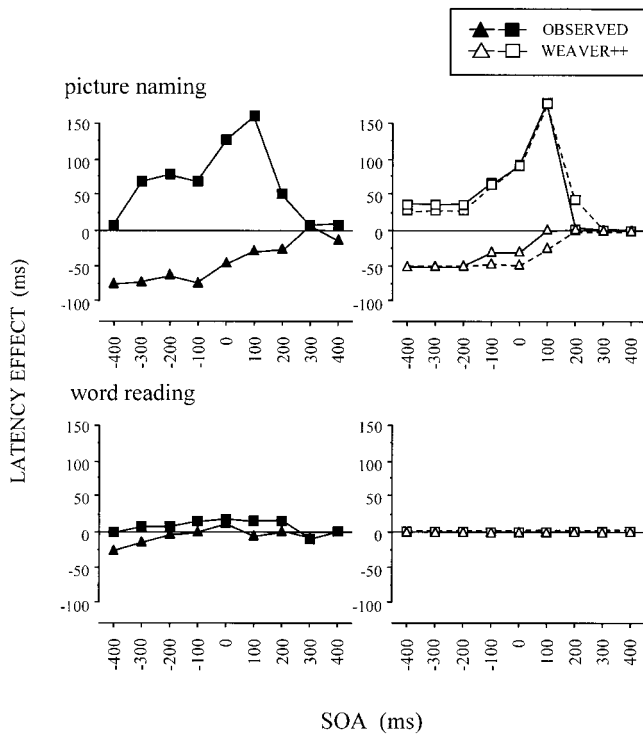


Figure 16. Time course of the congruency and incongruency effects from the picture-word task. The left-hand panels show the empirical data of W. R. Glaser and Dtingelhoff (1984, Experiment 1) and the right-hand panels show the WEAVERTT simulation results. The solid lines show the small-lexicon simulations with lemma retrieval, and the dashed lines show the simulations with lemma retrieval and word-form encoding. \square , \blacksquare = incongruent; \triangle , \blacktriangle = congruent; SOA = stimulus onset asynchrony.

ing words do produce the conflict in word reading with the word-word variant of the Stroop task (M. O. Glaser & Glaser, 1982; Van der Heijden, 1981). Van der Heijden (1981) presented the word-word stimuli tachistoscopically and indicated the relevant word by a bar marker. M. O. Glaser and Glaser (1982) and W. R. Glaser and Glaser (1989) presented the two color words with a certain SOA and instructed one group of participants to read aloud the second word (the condition with distractor-first SOAs) and another group of participants to read aloud the first word (the condition with distractor-second SOAs). Van der Heijden (1981), M. O. Glaser and Glaser (1982), and W. R. Glaser and Glaser (1989) observed the normal Stroop conflict with the word-word task variant. This also held for the time course of the effects. For reading aloud, patterns of interference and facilitation were obtained similar to those typically observed for color naming. That is, performance exhibited a peak of interference around SOA = 0 ms in the incongruent condition and a flat pattern of facilitation at distractor-first SOAs for the congruent condition. However, with this word-word task variant, no semantic gradient was obtained. When all words are in the response set, the color-neutral distractor word *THEME* produces the same amount of interference on reading *RED* as the color-related distractor word *SKY* and the color word *BLUE* (W. R. Glaser & Glaser, 1989, Experiment 3). Figure 17 shows the Stroop conflict in word reading empirically observed by W. R. Glaser and Glaser (1989) and the WEAVERTT simula-

tions. The semantic manipulation had no effect on the results and is not shown.

Why is the normal Stroop conflict obtained with word-word stimuli? According to WEAVERTT, when lemmas are not selected to accomplish the task, response selection takes place at the form level, where one morpheme is chosen over another (by contrast, in conceptually driven naming, a morpheme is selected for the selected lemma). Consequently, in reading aloud the target word *RED*, the morpheme <green> activated by the distractor word *GREEN* competes with the target <red> during the selection of morphemes (and, as with color naming, during the selection of motor programs in phonetic encoding), whereas the distractor *RED* helps in selecting the morpheme <red>. Although perceived distractor words activate their lemmas and concepts in the network, this activation does not affect the reading response because the activation does not significantly activate the form level (this happens only for a selected lemma). Thus, semantic influences do not reach form selection. Hence, although a Stroop conflict is obtained with word-word stimuli, a semantic gradient does not occur.

An alternative account of the lack of a semantic effect holds that because reading aloud is a fast process and semantic effects take time to develop, semantic influences are simply too slow to influence the reading response. However, this alternative view cannot explain all aspects of the data. The semantic gradient is absent even when the distractor word is presented well in advance of the target word (i.e., with long distractor-first SOAs). So, there is no gradient even when semantic effects do have sufficient time to develop. The findings are not restricted to color words. W. R. Glaser and Glaser (1989, Experiment 6) observed the Stroop conflict and the absence of semantic effects with reading words from a variety of other semantic domains (clothing, animals, furniture, etc.).

The model's account of a lack of semantic effects predicts that the effects should occur when the lemma level is involved in generating a response to words, such as in word categorizing (e.g., saying "animal" to the word *CAT*). This is exactly what is empir-

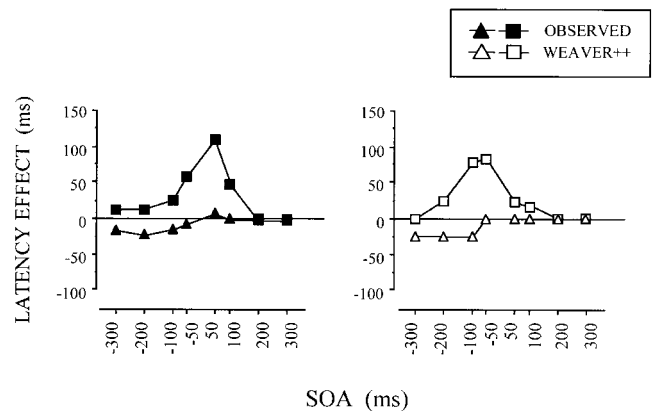


Figure 17. Time course of the Stroop phenomenon for order-based oral reading with word-word stimuli. The left-hand panel shows the empirical data of W. R. Glaser and Glaser (1989, Experiment 2), and the right-hand panel shows the WEAVERTT simulation results. \square , \blacksquare = incongruent; \triangle , \blacktriangle = congruent; SOA = stimulus onset asynchrony.

ically observed, both with picture (W. R. Glaser & Dünghoff, 1984) and word distractors (W. R. Glaser & Glaser, 1989). As with color naming, response set effects are obtained in the word–word task (e.g., W. R. Glaser & Glaser, 1989). This supports the assumption of the model that two lexical levels are involved in speaking words, a lemma and a morpheme level, and that the form-to-form route in oral reading involves a lexical rather than just a grapheme-to-phoneme mapping, where, by definition, the identities of the words involved play no role.

In summary, distracting colors do not yield a Stroop conflict in reading aloud, but distracting words produce the conflict and a response set effect in word reading. However, unlike distractor words in color naming, distractor words in reading do not yield a semantic gradient. This supports the assumption (implemented in WEAVER++) that two lexical levels are involved in word production, a lemma and a word-form level, and that word planning progresses from one level to the other in a functionally discrete fashion. Models that assume a single level (e.g., Caramazza, 1997; Starreveld & La Heij, 1996) have difficulty accounting for the full set of data. For example, Starreveld and La Heij proposed a model for picture-word interference in which concepts are directly mapped onto word forms without mediating lemmas. Activation automatically spreads from concepts to word forms and back again. Simulations by Starreveld and La Heij (p. 906) showed that their model captures the absence of a semantic gradient in reading aloud. However, this is because the model does not yield a Stroop conflict with word–word stimuli at all, contrary to the empirical data.

Thus, semantic effects occur, according to the WEAVER++ model, if the lemma level is involved in responding to a word, such as in word categorizing (e.g., saying “animal” to the word CAT). Furthermore, as shown by Roelofs (1992), categorizing yields a “reverse” semantic gradient in the model, that is, faster latencies in the semantic than in the unrelated condition, exactly as empirically observed. It may be argued that a semantic effect of distractor pictures in word categorizing occurs because the conceptual rather than the lemma level is involved. According to the model, however, accessing the conceptual stratum is not critical for the semantic effect, but the reverse semantic gradient should also be obtained when only lemma-level information, such as a word’s grammatical gender, needs to be accessed. Experiment 3, reported next, used the picture–word variant of the Stroop task to test this critical prediction. The experiment exploited the linguistic fact that nouns take gender-marked articles in Dutch definite descriptions: the article *het* with neuter gender and the article *de* with non-neuter gender. So if a noun is presented and participants have to read the noun aloud preceded by its gender-marked article (not presented), the grammatical gender of the noun needs to be retrieved to determine the right article, *de* or *het* (color nouns in Dutch all take the article *het*, so the classic color-word Stroop task cannot be used here). For example, if participants have to respond to the word PISTOOL (“pistol”) by saying “het pistol,” the gender of the noun *pistol* needs to be accessed to determine the correct determiner, *het*. Consequently, a semantic effect from distractor pictures should be obtained.

Experiment 3 tested for semantic effects in word reading, word categorizing, and word reading with a determiner. Distractor pictures should not affect word reading, but they should yield a reverse semantic gradient in word categorizing, which would rep-

licate findings in the literature (i.e., W. R. Glaser & Dünghoff, 1984). Crucially, distractor pictures should also yield such a reverse semantic gradient in word reading with a determiner, because generating the response requires lemma access. Why is a reverse semantic gradient predicted? As explained in Roelofs (1992), word distractors in picture naming yield slower latencies in the semantic than in the unrelated condition in WEAVER++, because the target picture activates the lemma of a semantically related distractor word more than the distractor word activates the target lemma because of distances in the lexical network. In reversing the target–distractor relation by making the word the target and the picture the distractor, as is done in the current experiment, the direction of the semantic effect should also reverse: Semantic interference should turn into semantic facilitation. Thus, a reverse semantic gradient should be obtained for word reading with a determiner in the context of picture distractors.

Experiment 3: A Test of a Reverse Semantic Gradient in Word Reading With a Determiner

Method

Participants. The experiment was carried out with 36 paid participants from the pool of the Max Planck Institute. All participants were young adults who were native speakers of Dutch. None of them participated in any of the earlier experiments reported in this article.

Materials and design. From the picture gallery available at the Max Planck Institute, 32 pictured objects from eight different semantic categories were selected together with their basic-level terms and category names in Dutch. Appendix B lists the word materials. Two of the objects in a category had basic-level names with neutral grammatical gender (the *het* words), and the two remaining objects had names with non-neuter gender (the *de* words). In addition to the 32 pictured objects, a drawing of an empty picture frame was created to serve as a neutral control picture. Furthermore, four additional pictures from two semantic categories (different from the eight experimental categories) were selected as practice items. The pictures were white line drawings on black backgrounds. They were digitized and scaled to fit into a virtual frame of 15 cm × 15 cm. The words were on average 1.5 cm high × 6 cm wide.

There were two independent variables. The first independent variable was task, with three levels: word reading, word categorizing, and word reading with a determiner. Task was varied between participants, with 12 participants performing each of the three tasks. Participants who were given the word-reading task simply had to read aloud the written word and ignore the picture. Participants performing the word-categorizing task had to produce the semantic category name of the word and ignore the picture. For example, they had to say “wapen” (“weapon”) in response to the word PISTOOL (“pistol”). Finally, participants performing the reading-with-determiner task had to produce the word preceded by its gender-marked definite article and ignore the picture. For example, they had to say “het pistol” (“the pistol”) in response to the word PISTOOL (“pistol”).

The second independent variable was distractor, which was varied within participants. Each target word was combined with a picture from the same semantic category (the semantic condition), randomly with a picture from another semantic category (the unrelated condition), or the neutral picture frame (the control condition). The word and the picture name always had the same grammatical gender. A participant received 32 word–picture pairings in each of the three distractor conditions, yielding 96 trials in total. The word–picture stimuli were presented in random order.

Procedure and apparatus. The participants were tested individually with the same computer monitor and microphone as was used in the earlier experiments. Before the beginning of the experiment, the participants were familiarized with the words, the responses (in the categorizing task), and

the distractor pictures. After a participant had read the instructions, a block of 12 practice trials was administered, which was followed by the experimental trials. The structure of a trial was as follows. First, the participant saw an asterisk for 500 ms. Next, the screen was cleared for 500 ms, followed by the display of the distractor picture 150 ms before the presentation of the target word. The asterisk, picture, and word were presented in white on a black background. The distractor pictures appeared in the middle of the screen and the target words were shown in the center of the pictures. Before the start of the next trial there was a blank interval of 500 ms. The total duration of a trial was 3 s. A Hermac computer controlled the presentation of the stimuli and the collection of response times.

Analysis. The error coding was the same as in the previous experiments. The latencies and errors were submitted to by-participant and by-item ANOVAs with the crossed variables task and distractor (because 32 different words–pictures were tested, by-item analyses were also done). Task was tested between participants and distractor within participants.

Results and Discussion

Table 4 gives the means and standard deviations of the production latencies and the mean error percentages for the distractor by task cells. Overall, production onset latencies were slower for word categorizing than for word reading and reading with a determiner. The word reading latencies did not differ between distractor conditions, whereas there were clear distractor effects for word categorizing and reading with a determiner. Categorizing was slower in the unrelated than in the control condition and faster in the semantic than in the control condition. Reading with determiner was slower in the unrelated than in the control condition, but the semantic and control condition did not differ. Figure 18 shows how the latency effects (i.e., semantic and unrelated vs. control) varied with task.

Production latencies. The statistical analyses of the production times confirmed these observations. There were main effects of task, $F_1(2, 33) = 132.22$, $MSE = 13,934$, $p < .001$, $F_2(2, 93) = 1926.96$, $MSE = 2,525$, $p < .001$, and distractor, $F_1(2,$

Table 4

Means (RTs, in Milliseconds) and Standard Deviations (SDs) of the Production Latencies and Mean Error Percentages (E%) as a Function of Task and Distractor for Experiment 3

Distractor	RT	SD	E%
Task: Word reading			
Semantic	497	82	1.8
Unrelated	493	82	1.8
Control	495	86	1.3
Task: Word categorizing			
Semantic	862	126	2.1
Unrelated	983	158	8.4
Control	888	165	4.5
Task: Word reading with determiner			
Semantic	550	92	0.3
Unrelated	576	130	3.7
Control	539	105	1.6

Note. RT = response time.

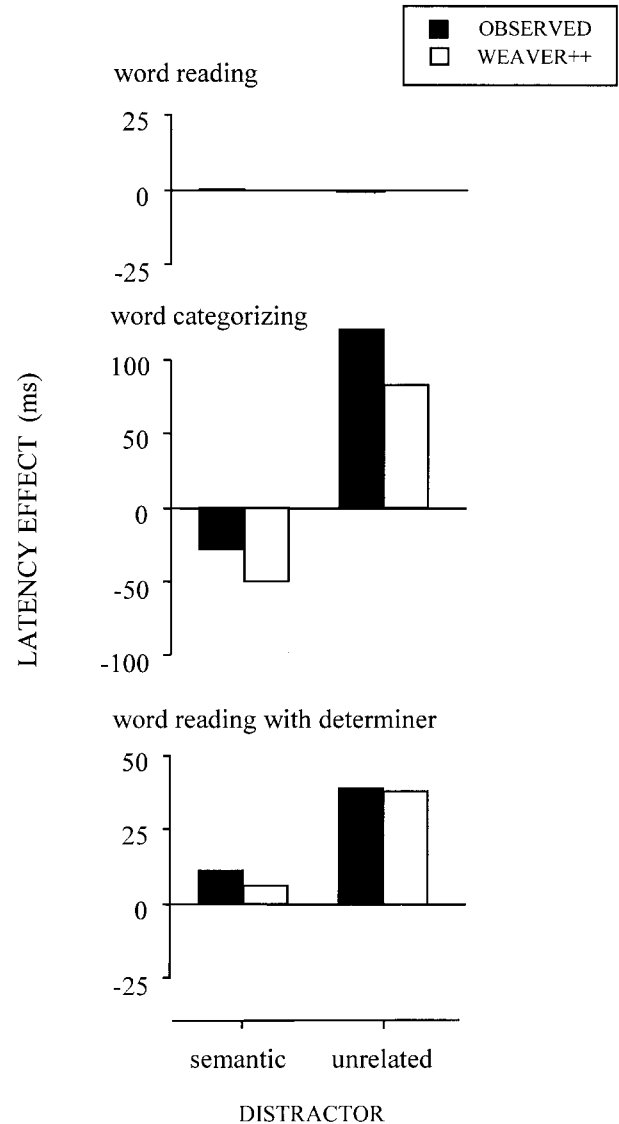


Figure 18. Latency effects of semantically related and unrelated distractor pictures relative to control stimuli as a function of task in Experiment 3.

66) = 61.29, $MSE = 421$, $p < .001$, $F_2(2, 186) = 91.60$, $MSE = 729$, $p < .001$. Task and distractor interacted, $F_1(4, 66) = 33.90$, $MSE = 421$, $p < .001$, $F_2(4, 186) = 50.20$, $MSE = 729$, $p < .001$. Distractor had no effect on word reading, $F_1(2, 22) = 0.62$, $MSE = 93$, $p > .50$, $F_2(2, 62) = 0.36$, $MSE = 294$, $p > .50$, but it had an effect on word categorizing and reading with a determiner, respectively: $F_1(2, 22) = 51.14$, $MSE = 975$, $p < .001$; $F_2(2, 62) = 91.97$, $MSE = 1,392$, $p < .001$; $F_1(2, 22) = 22.72$, $MSE = 195$, $p < .001$; and $F_2(2, 62) = 23.58$, $MSE = 500$, $p < .001$. Categorizing latencies were faster in the semantic than in the control condition, $F_1(1, 11) = 5.74$, $MSE = 873$, $p < .05$, $F_2(1, 31) = 14.77$, $MSE = 776$, $p < .001$, and latencies were slower in the unrelated than in the control condition, $F_1(1, 11) = 58.65$, $MSE = 911$, $p < .001$; $F_2(1, 31) = 73.67$, $MSE = 1907$, $p < .001$. The latencies of reading with determiner did not differ between the semantic and the control condition, $F_1(1, 11) = 3.23$, $MSE = 186$,

$p > .10$, $F_2(1, 31) = 6.26$, $MSE = 287$, $p < .05$, but latencies were slower in the unrelated than in the control condition, $F_1(1, 11) = 32.58$, $MSE = 255$, $p < .001$; $F_2(1, 31) = 34.94$, $MSE = 636$, $p < .001$. Furthermore, latencies were slower in the unrelated than in the semantic condition, $F_1(1, 11) = 30.41$, $MSE = 146$, $p < .001$; $F_2(1, 31) = 19.67$, $MSE = 578$, $p < .001$.

Error rates. The statistical analysis of the error rates yielded a significant result for distractor, $F_1(2, 66) = 7.88$, $MSE = 0.0012$, $p < .001$, $F_2(2, 186) = 13.14$, $MSE = 0.0023$, $p < .001$, but not for task, $F_1(2, 33) = 2.31$, $MSE = 0.01$, $p > .10$, $F_2(2, 93) = 12.61$, $MSE = 0.0023$, $p < .001$. Furthermore, task and distractor interacted, $F_1(4, 66) = 2.53$, $MSE = 0.0012$, $p < .05$; $F_2(4, 180) = 4.19$, $MSE = 0.0023$, $p < .005$. These error effects appeared to be due to the fact that more errors were made in the unrelated than the semantic condition for both categorizing, $F_1(1, 11) = 8.49$, $MSE = 0.0035$, $p < .05$, $F_2(1, 31) = 17.64$, $MSE = 0.0035$, $p < .001$, and for reading with a determiner, $F_1(1, 11) = 14.19$, $MSE = 0.0007$, $p < .005$, $F_2(1, 31) = 13.76$, $MSE = 0.0013$, $p < .001$. Most errors were made in the unrelated condition, which also had the longest latencies. Thus, there is no evidence of a speed-accuracy trade-off.

To conclude, the experiment replicates the findings of W. R. Glaser and D ngelhoff (1984) for word reading and word categorizing. Distractor pictures had no impact on reading aloud, whereas they yielded a reverse semantic gradient (i.e., faster latencies in the semantic than in the unrelated condition) in word categorizing. The critical new finding is that when participants had to generate the word together with a gender-marked article, a reverse semantic gradient was also obtained. This agrees with the predictions derived from WEAVERTT. Figure 18 shows the simulation results for the experiment. To fine-tune the fit, critdiff was set to 3.0 for reading with determiner and to 3.6 for word categorizing. All other parameter values were unchanged. There is a good agreement between model and real data.

Simulations revealed that the reverse semantic gradient in word reading with a determiner increases when the size of critdiff increases in the model. The model predicts that, different from word categorizing, the semantic condition in reading with a determiner should never be faster than the control condition. This prediction may be tested in future research, for example, by examining semantic effects across a range of SOAs.

Order-Based Color Naming With Word-Color Stimuli and Task Uncertainty (Finding 8)

In a standard color-word Stroop experiment, there is no task uncertainty. One group of participants is asked to name the color and ignore the word, and sometimes (e.g., M. O. Glaser & Glaser, 1982) another group of participants is asked to read aloud the word and ignore the color patch. Trials are blocked by task. W. R. Glaser and Glaser (1989) introduced task uncertainty by instructing one group of participants to respond to the second stimulus (the condition with distractor-first SOAs) and another group to respond to the first stimulus (the condition with distractor-second SOAs). If the designated stimulus was a word it had to be read aloud, and if it was a color it had to be named. WEAVERTT performs this task by production rules that set the goal to color naming or word reading depending on what the first stimulus is and depending on

the SOA condition. With distractor-first SOAs, for example, the rules say the following:

- P5. IF the first stimulus is a word
THEN the goal is to say the name of the color.
- P6. IF the first stimulus is a color
THEN the goal is to say the name of the word.

W. R. Glaser and Glaser (1989, Experiment 1) showed that the standard Stroop patterns are replicated with this task variant. Interference from incongruent stimuli increases when the preexposure and postexposure times become shorter and interference peaks around SOA = 0 ms. Facilitation from congruent stimuli does not vary much with the preexposure time. Figure 19 shows the SOA curves that were observed by W. R. Glaser and Glaser and those generated by WEAVERTT. Like the real data, interference increases in the model when the distractor-first and distractor-second SOAs become shorter, with interference peaking around SOA = 0 ms. Facilitation does not vary much with distractor preexposure. So, the model captures the empirical patterns obtained with task uncertainty.

The fact that the standard SOA patterns are obtained under task uncertainty poses a challenge to the models of Cohen et al. (1990) and Phaf et al. (1990), in which a task needs to be specified before the beginning of a trial so that the system can settle into a ready state for the task. Furthermore, proposals in the literature (e.g., Rogers & Monsell, 1995) have suggested that the essence of the Stroop conflict is competition between tasks per se (i.e., reading and color naming). The Stroop conflict is explained as interference on the "weaker" color-naming task from the supposedly "stronger" reading task, whereas the reverse does not occur. However, task competition fails to explain why the congruent condition (where the same tasks compete) yields facilitation. Also, this view would predict a greater cost for switching between weak tasks like color and numerosity naming than between strong tasks like word and numeral reading. However, Allport, Styles, and Hsieh (1994) observed that alternating between tasks that exhibit the Stroop

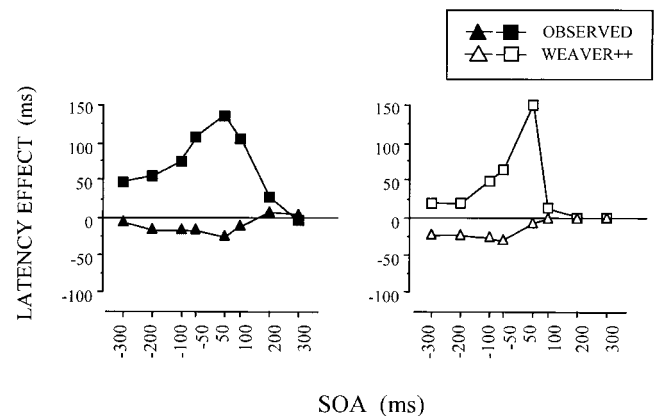


Figure 19. Time course of the Stroop phenomenon for order-based color naming with color-word stimuli and task uncertainty. The left-hand panel shows the empirical data of W. R. Glaser and Glaser (1989, Experiment 1) and the right-hand panel shows the WEAVERTT simulation results. □, ■ = incongruent; △, ▲ = congruent; SOA = stimulus onset asynchrony.

conflict (i.e., between color and numerosity naming trials) does not yield a greater task switch cost than alternating between tasks that do not yield the Stroop conflict (i.e., between word and numeral reading trials).

Recent brain imaging studies using the color-word Stroop task have shown that the anterior cingulate cortex is more active in the incongruent than in the congruent condition, but both conditions show more activation than the control condition. Furthermore, there is increased anterior cingulate activation in anticipation of a task (for reviews, see Bush et al., 2000; MacLeod & MacDonald, 2000). This suggests that the anterior cingulate plays a role in setting a task and securing task-relevant control (i.e., directing attention to a goal) rather than dealing with conflict per se, as has often been assumed (e.g., Posner & Raichle, 1994). In the congruent condition there is less of a need for exerting task-relevant control, because both naming the color and reading the word would yield the correct response. (In most brain imaging studies, the Stroop conditions were blocked, like in the original card version designed by Stroop, 1935.) This would explain why the anterior cingulate is more active in the incongruent than in the congruent condition. The finding that the pattern of anterior cingulate activation does not correspond to the response time ordering of the Stroop conditions agrees with the distinction between executive and slave aspects of performing a task in WEAVER++. Anterior cingulate activation reflects executive processing (setting and maintaining a goal), whereas the Stroop conflict arises during the planning of the target word (i.e., during lemma retrieval and word-form encoding). The latter happens in perisylvian areas of the brain (including the areas of Wernicke and Broca; for details, see Indefrey & Levelt, 2000; see Figure 9 in this article). This view predicts that the anterior cingulate should be active in word reading with a color distractor (which requires task-relevant control) even in the absence of reverse Stroop interference. Furthermore, task uncertainty should increase anterior cingulate activation compared with task certainty, even when Stroop interference does not differ. These predictions may be tested in future brain imaging studies.

Order-Based Color Naming With Color-Color Stimuli (Finding 9)

In the color-color variant of the Stroop task, participants are instructed to name the second color and ignore the first color (the condition with distractor-first SOAs) or to name the first color and ignore the second one (the condition with distractor-second SOAs). M. O. Glaser and Glaser (1982) and W. R. Glaser and Glaser (1989) showed that with this task variant (both with and without task certainty), the standard SOA patterns are replicated with one major difference. With color-color stimuli, the incongruent condition produces a flat rather than a steep interference curve. Van der Heijden (1981) presented the color-color stimuli tachistoscopically and indicated the relevant color by a bar marker; he also observed smaller interference effects for color-color than for word-color stimuli. Similarly, W. R. Glaser and Glaser (1989) observed smaller effects for picture-picture than for picture-word interference (W. R. Glaser & Dünghoff, 1984).

Figure 20 shows the color-color SOA curves as measured by W. R. Glaser and Glaser (1989) and those generated by WEAVER++. Because there was task uncertainty in the experiment,

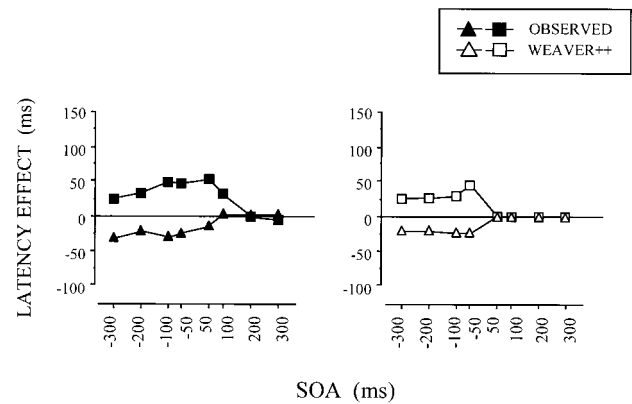


Figure 20. Time course of the Stroop phenomenon for order-based color naming with color-color stimuli. The left-hand panel shows the empirical data of W. R. Glaser and Glaser (1989, Experiment 2) and the right-hand panel shows the WEAVER++ simulation results. □, ■ = incongruent; △, ▲ = congruent; SOA = stimulus onset asynchrony.

WEAVER++ performed this task by using production rules that set the goal to color naming or word reading depending on what the first stimulus was. With negative SOAs, for example, the rule underlying color naming is as follows:

- P7. IF the first stimulus is a color
THEN the goal is to say the name of the second color.

Thus, unlike the model of Cohen et al. (1990), WEAVER++'s attentional mechanism also works when a single pathway subserves both target and distractor processing. The figure shows that, like the real data, there is an increase of interference in WEAVER++ with decreasing preexposure time of the distractor. The increment rate of the interference is lower than normal, and the interference does not reach its normal steep height, as empirically observed. The reason for the attenuated effect in the model is that color distractors do not have direct access to the response level (the lemma level), unlike distractor words, which directly activate their lemmas. Thus, the effect of incongruent colors is reduced somewhat compared with incongruent words because of network distance.

Spatial Certainty (Finding 10)

When the components of a color-word Stroop stimulus are spatially separated, the interference is reduced (e.g., Kahneman & Henik, 1981; MacLeod, 1998). Gatti and Egeth (1978) increased the spatial distance between color and word from 1° to 5° and observed that interference diminished from 90 ms to 40 ms. Moreover, M. O. Glaser and Glaser (1982) observed that when spatially separated color-color or word-word stimuli were presented under spatial certainty (e.g., the target color patch always appeared above the distractor color) interference around SOA = 0 ms disappeared, which suggests that under this condition, the distractor may be filtered out in an early stage of visual processing. M. O. Glaser and Glaser (Experiment 4) examined the time course of the Stroop effect for spatially separated color-color and word-word stimuli whose location was certain. They observed some

interference and facilitation at the most negative SOAs, but no effect at all around the SOA of 0 ms, where normally the largest Stroop conflict occurs. This suggests that under spatial certainty, expectancy-induced priming is present but automatic priming does not occur.

Enhanced input control is simulated in WEAVER++ by reducing the duration of distractor input (changed from 100 to 25 ms in the simulations). That is, it is assumed that if an input control rule (like production rule P3) makes reference to spatial locations, it works faster because spatial information is available faster. This is supported by the event-related brain potential evidence that suggests earlier attentional disengagement for spatial than for nonspatial stimulus dimensions (Hillyard et al., 1995). Figure 21 shows the SOA curves measured by M. O. Glaser and Glaser (1982) and those generated by WEAVER++. Like the real data, there is some interference and facilitation for color naming at long distractor-first SOAs but no effect at all around the SOA of 0 ms. For reading, there is some facilitation at distractor-first SOAs but, again, no effect around SOA = 0 ms. However, different from what holds for color-color and word-word stimuli, the interference does not disappear in the model with color-word stimuli, as shown in Figure 14. With enhanced input control (with *du* set to 25 ms) there is still interference and facilitation for color-word stimuli at SOA = 0 ms, in agreement with the color-word findings of MacLeod (1998).

Semantic Variation (Finding 11)

Klein (1964) was the first to examine how “semantic distance” affects color naming in the Stroop task. Measuring stimulus-card

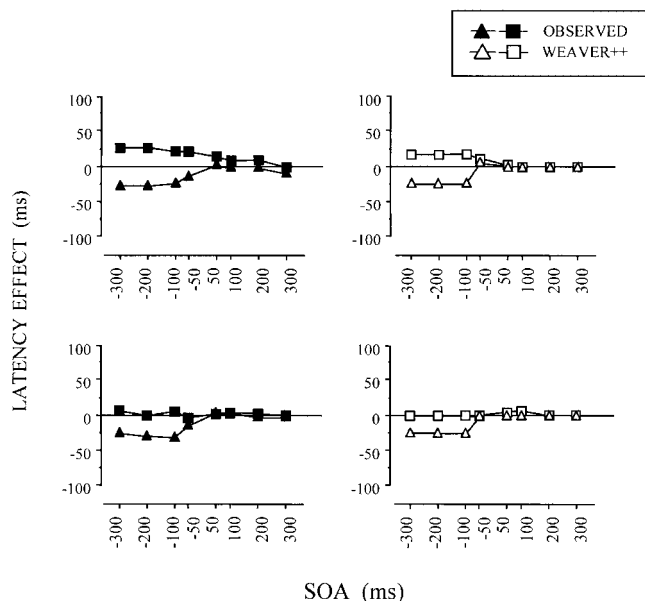


Figure 21. Time course of the Stroop phenomenon with color-color stimuli (top panel) and word-word stimuli (bottom panel) under spatial certainty: Observed and simulated findings. The left-hand panels show the empirical data of M. O. Glaser and Glaser (1982, Experiment 4). The right-hand panels show the WEAVER++ simulation results. □, ■ = incongruent; △, ▲ = congruent; SOA = stimulus onset asynchrony.

completion times, he observed that the standard Stroop condition produced the largest interference effect. This effect was reduced when color words were used that were not possible responses, which is the response set effect. The interference was further reduced when incongruent color associates were used such as SKY (related to blue). Rare words such as SOL reduced the interference still further, and nonsense syllables produced almost no interference. This difference in interference effect depending on how close the distractor is to the target word was referred to as the *semantic gradient*. The gradient is not restricted to incongruent distractors. Dalrymple-Alford (1972) not only observed that “near” incongruent color words (e.g., BLUE in naming the color red) yielded more interference than “far” incongruent color associates (e.g., SKY in naming red), but also that, compared with unrelated control words, “near” congruent color words (RED in naming red) yielded more facilitation than “far” congruent color associates (BLOOD in naming red).

W. R. Glaser and Glaser (1989, Experiment 5) examined the time course of the semantic gradient using near and far incongruent and congruent conditions. For example, a to-be-named red color patch was paired with the words FIRE (far congruent, FC), RED (near congruent, NC), LAWN (far incongruent, FI), or GREEN (near incongruent, NI). The SOAs tested were -300, -200, -100, -75, -50, 0, 50, 75, 100, 200, and 300 ms. Figure 22 shows the empirical results. In agreement with the findings of Klein (1964) and Dalrymple-Alford (1972), the word GREEN yielded more interference than LAWN, and RED yielded more facilitation than FIRE. Interestingly, the FC words (e.g., FIRE) produced facilitation at distractor-first SOAs, but they tended to yield interference around SOA = 0 ms. At distractor-first SOAs, FC distractors act like NC ones (RED), whereas at short distractor-second SOAs they behave like FI distractors (LAWN).

The figure also shows the simulation results. The simulations made use of additional semantic fields linked up with the color field (e.g., RED(X) was connected to FIRE(X)), following Roelofs (1992). As can be seen, the Stroop conflict in WEAVER++ exhibits a semantic gradient. However, the model clearly underestimates the amount of interference in the NI condition (e.g., from GREEN in naming a red patch). Although the interference increases with decreasing distractor-first SOAs in the real experiment, the magnitude of the interference appears to be unusually high, in particular, some 50–75 ms higher at all SOAs compared with the interference scores in the studies of M. O. Glaser and Glaser (1982) and Long and Lyman (1987). It is unclear why the interference is so exceptionally large in Experiment 5 of W. R. Glaser and Glaser (1989). Furthermore, facilitation is equal for far and near in the model at longer SOAs, unlike the real data where facilitation is almost absent for FC. For present purposes, however, it is important that the Stroop conflict in the model exhibits the semantic gradient in color naming. This is in contrast to word reading, where a lack of a semantic gradient is predicted and empirically observed (W. R. Glaser & Glaser, 1989; see also Experiment 3 above). In agreement with the empirical data on color naming, the FC condition (FIRE) produces facilitation at distractor-first SOAs, but it tends to yield interference around SOA = 0 ms.

Manual Responding (Finding 12)

Stroop interference remains with typing the word (e.g., Logan & Zbrodoff, 1998) and with button responses (i.e., pressing a colored

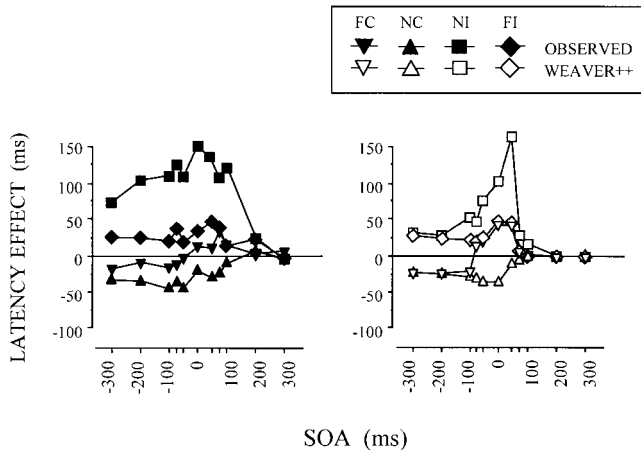


Figure 22. Time course of semantic gradient for color naming with color-word stimuli: Observed and simulated findings. The left-hand panel shows the empirical data of W. R. Glaser and Glaser (1989, Experiment 5). The right-hand panel shows the WEAVERTT simulation results. FC = far congruent; NC = near congruent; NI = near incongruent; FI = far incongruent; SOA = stimulus onset asynchrony.

button or a button labeled with a word). Manipulating response modality (oral vs. manual) and type (word vs. color) has been important in testing translation models of Stroop performance (e.g., Virzi & Egeth, 1985), which hold that colors and words are mapped from one internal code into another by a translation device. Such models predict interference when the relevant stimulus type (e.g., a color stimulus) does not match the response type (e.g., a word response), that is, when the task requires a translation between types. This explains why irrelevant words interfere with color naming (which requires a color-to-word translation) and why irrelevant colors do not interfere with word reading (not requiring a translation). However, contrary to the translation account, interference has been obtained from irrelevant words on color responses with an untranslated color-color task in several studies, though not all. In reviewing these studies, Sugg and McDonald (1994) argued that interference for untranslated color-color responses has been obtained in these studies because their design failed to prevent verbal mediation. However, remedying this design problem does not save the translation account, as Sugg and McDonald's own time-course findings show.

Sugg and McDonald (1994) examined the SOA curves for all four combinations of relevant stimulus type (color or word) and response type (color or word) in a button-pressing version of the standard Stroop task. Color responses required pressing a colored button and word responses required pressing a button labeled with a word. SOAs varied from -300 ms to $+300$ ms in steps of 100 ms. For the translated color and word responses (word-color and color-word, respectively), the interference predicted by translation models was obtained, and the untranslated word responses (word-word) showed no effect, also as predicted. For the untranslated color-color responses, however, Sugg and McDonald obtained interference at distractor-first SOAs (in agreement with earlier studies that showed interference) but no effect at $SOA = 0$ ms (in agreement with the earlier studies that showed no effect). Thus, the SOA manipulation appears to be important in obtaining interfer-

ence for untranslated responses. Most importantly, interference is obtained for untranslated color-color responses, which is difficult to explain under the translation account (Virzi & Egeth, 1985).

According to WEAVERTT, Stroop interference lies within the language production system. Interference should remain if lexical entries are needed to mediate a button-press response. In applying the model to the findings on button pressing, I included nodes in the network that represented responses outside the language production system proper but that were mediated by the lexical network. Following the proposal by Sugg and McDonald (1994), I assume that responses are generated using the shortest route possible from stimulus to response while respecting the response type. Thus, response nodes are connected to concept nodes for color responses, to lemma nodes for color-driven word responses, and to form nodes for word-driven word responses. Verification was again the means of selection. Thus, although button responses are involved in all conditions, how lexical entries mediate the response critically differs between conditions. Figure 23 shows the time-course findings from Sugg and McDonald (1994) together with the simulation results.⁶ For the translated color responses (WC) and word responses (CW), interference was obtained, in agreement with the empirical findings. The untranslated word responses (WW) showed no effect, as empirically observed. And for the untranslated color-color responses (CC), the model produces interference at distractor-first SOAs and no effect at $SOA = 0$ ms and later ones, in full agreement with the empirical data. To conclude, unlike the translation models (e.g., Virzi & Egeth, 1985), WEAVERTT captures the data patterns.

Stroop Interference in Bilingual Individuals (Finding 13)

Performance of bilingual individuals in a bilingual Stroop experiment creates an additional selection problem, namely that selection of the response should be from the appropriate language. For example, if a Dutch-English speaker has to use English in naming the color of the Dutch word GROEN ("green") in red ink, not only must the word *green* be ignored but also the translation equivalent of the English target, Dutch *red* ("red"). Stroop interference occurs between the languages of bilinguals (e.g., Dutch GROEN interferes with the production of "red" in response to GROEN in red ink), although the effect is not as great as that within either one of the languages. The interference between languages appears to be about 60%–70% of that within a single language (e.g., Dyer, 1971a; see Francis, 1999, for a review).

Selection from one language rather than the other may be achieved by specifying the target language in the production rules for lemma selection (Roelofs, 1998a). In a bilingual Stroop experiment, the task specifies the target language. For example,

- P8. IF RED(X) is flagged as goal concept,
and the target language is English,
and the activation of *red* exceeds threshold,
THEN select *red*.

⁶ The button nodes are not critical. Equivalent results are obtained when response selection involves the corresponding nodes within the lexical network.

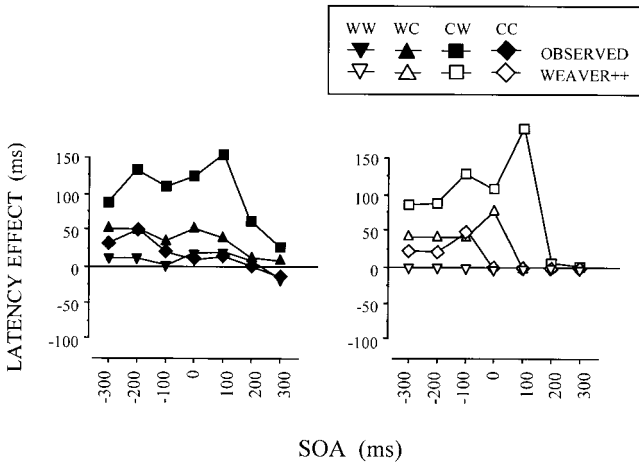


Figure 23. Time course of the Stroop phenomenon with manual responding: Observed and simulated findings. The left-hand panel shows the empirical data of Sugg and McDonald (1994, Experiment 1). The right-hand panel shows the WEAVERTT simulation results. WW = word-word response; WC = word-color response; CW = color-word response; CC = color-color response; SOA = stimulus onset asynchrony.

Figure 24 shows the findings on bilingual Stroop performance of a study by Dyer (1971a), which are typical. Participants were bilingually balanced in English and Spanish. The figure shows the percentage of the standard interference (when English is both the distractor and naming language) for the three other language combinations. When Spanish is both the distractor and naming language (labeled “second-second” in the figure), the standard amount of interference is obtained. However, interference reduces when the distractor and target are from different languages, both when the distractor is in the second language and naming is in the first language, and vice versa (labeled “second-first” and “first-second,” respectively). The figure also shows the WEAVERTT simulation results. The simulations used a network that included lemmas for the second language words *rajo*, *verde*, and *azul* (the Spanish words used by Dyer, 1971a). The second-language lemmas were connected to the concept nodes RED(X), GREEN(X), and BLUE(X), respectively. Figure 24 shows that the model captures the results. In performing a bilingual Stroop task, the words in the other language are not permitted responses. So, according to the model, they behave like nonresponse set members in the standard, monolingual version of the task, which explains the attenuated effect.

Interference as a Function of Training (Finding 14)

MacLeod and Dunbar (1988) examined Stroop interference as a function of training on saying color words in response to arbitrary shapes. Participants learned to call each of four different polygons by a different color name (“green,” “pink,” “orange,” “blue”). In Experiment 2, participants received 20 practice sessions on shape naming, on 20 separate days. After 1, 5, and 20 sessions of practicing shape naming (respectively, 72, 504, and 2,520 trials per stimulus), performance was tested for interference from “incongruent” colors on shape naming (e.g., saying “blue” to the green polygon named “blue”) and from “incongruent” shapes on color

naming (e.g., saying “green” to the color green of the polygon named “blue”). On Day 1, colors interfered with shape naming, but shapes not with colors. On Day 5, colors interfered with shape naming and shapes interfered with color naming. Finally, on Day 20, shapes interfered with color naming, but colors no longer interfered with shape naming. Thus, training on a task may reverse the patterns of interference between tasks. Importantly, with moderate practice, there is equivalent interference in both tasks.

What does training accomplish? Evidence from a long tradition in psychology (e.g., Anderson, 1983; Cattell, 1886; Posner & Raichle, 1994; Posner & Snyder, 1975; Shiffrin & Schneider, 1977) suggests that practice in performing a novel task may lead to a change in processing route. For example, in a positron-emission tomography (PET) study of S. E. Peterson, Fox, Posner, Mintun, and Raichle (1989), participants had to say a verb when visually presented with a noun (e.g., HAMMER—say “hit” or another appropriate verb). PET images revealed that brain activity during verb generation increased in frontal and temporal areas relative to a control condition that required reading the nouns aloud. However, after practice on verb generation, the differential activation between verb generation and noun reading disappeared. This suggests that practice may alter the route underlying task performance: Practicing verb generation seems to change conceptually mediated responses into word-level responses.

Similarly, practice on shape naming may lead to a change in the route used for performing the task. Like verb generation, shape naming initially may be conceptually mediated. Initially (on Day 1), participants produce a color name in response to a visual shape by retrieving the newly learned association between the shape and a color concept. Because the association is still weak, colors interfere with shape naming but not vice versa. After more practice on shape naming (Day 5), the association between shape and concept has been strengthened, but also a direct association between the visual shape and the word may have been established.

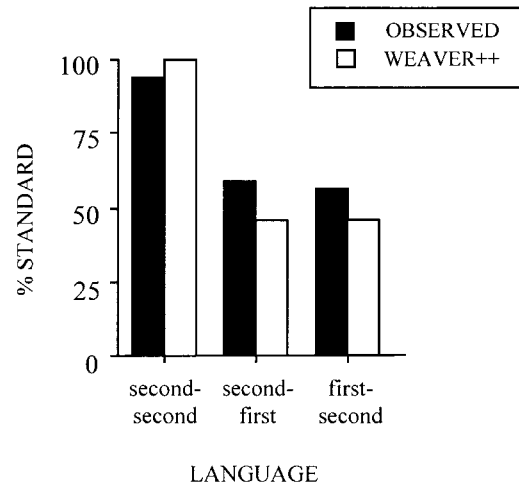


Figure 24. Latency effects for different language combinations expressed as a percentage of the standard effect: Real data from Dyer (1971a) and WEAVERTT simulations. The language conditions are labeled according to what the distractor and naming language was. For example, “second-first” means that the distractor is from the second language and the target from the first language of the participants.

Consequently, colors still interfere with shape naming, but shapes also interfere with color naming. Finally, after extensive practice on shape naming (Day 20), the direct association between the visual shape and the corresponding color word may have become so strong that it has gained full control of the shape-naming response (i.e., shape naming becomes similar to reading aloud). Consequently, shapes interfere with color naming, but colors no longer interfere with shape naming.

Simulations showed that WEAVER++ captures the reversal of interference as a function of training when extensive training on saying "red" to an arbitrary polygon shifts control from concept-route responses (shape \rightarrow RED(X) \rightarrow red \rightarrow "red") to word-route responses (shape \rightarrow red \rightarrow "red"). The association strength between a shape, on the one hand, and a concept and word, on the other hand, was assumed to increase through Hebbian learning with practice from 0.0 (no association) to 0.004 (weak association), and finally to 0.02 (strong association). The association strength is the proportion of extin that is sent from the shape representation to the target. Figure 25 shows how interference changed as a function of training in the study of MacLeod and Dunbar (1988) and in WEAVER++ simulations. The model captures the reversal of interference as a function of training: Initially, colors interfere with shape naming but shapes not with color naming; with moderate practice, there is equivalent interference in both tasks; and with extensive training, only shapes interfere with color naming and not vice versa. Thus, unlike the model

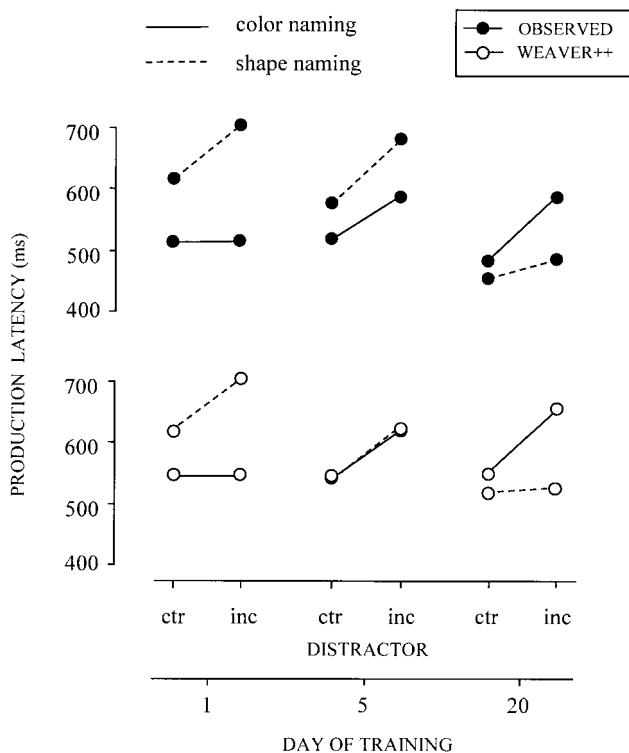


Figure 25. Interference as a function of training. Observed production latencies from MacLeod and Dunbar (1988) and WEAVER++ simulation results. The absolute production latencies were obtained for the model by adding a constant (i.e., 447 ms, estimated earlier) to the simulation latencies. ctr = control; inc = incongruent.

of Cohen et al. (1990), WEAVER++ yields interference in both tasks with moderate practice (note that WEAVER++ yields equivalent interference even with equal speed of the tasks).

There is initially no architectural difference between color naming and shape naming. However, I argue that such a difference does exist (from reading onset) between color naming and word reading, and that therefore it is not possible to reverse the patterns of interference for color naming and word reading, not even with extensive practice on color naming (e.g., Brown, 1915; MacLeod, 1998). According to WEAVER++, even if extensive training on naming four colors would establish direct links between these colors and the corresponding words so that color naming would become like word reading, Stroop interference should persist: The model yields interference for word-word stimuli, as we saw.

Stroop Interference in Childhood, Adulthood, and Aging (Finding 15)

Ligon (1932) measured color-word reading and color-naming times for hundreds of schoolchildren whose ages ranged from 6 to 16 years. He observed that reading and naming times decrease with age but that reading remains faster than color naming at a constant difference. But what happens to the Stroop conflict in color naming? Comalli et al. (1962) tested participants from ages 7 to 80 years on the Stroop task and confirmed the finding of a constant difference between reading and naming. However, they observed that the youngest (7–13) and oldest (65–80) groups exhibited most Stroop interference. The larger interference effect for the oldest than for the middle group (17–44) was confirmed by Cohn, Dustman, and Bradford (1984) and, recently, by Spieler et al. (1996). The interference is larger than one might predict solely on the basis of generally slower processing in the youngest and oldest groups. Comalli et al. observed that performance in the control condition was similar for the middle and the oldest group but that interference for the oldest group was larger. Also, the youngest group was slower in the control condition than the oldest group, but the size of the interference effect was similarly high. This suggests that the amount of interference is not just a function of the general processing speed. Schiller (1966) observed a rise and fall of interference with age in children. Interference is minimal for children in Grade 1 and becomes maximal in Grades 2 and 3 as reading skill develops, after which the amount of interference declines. Comalli et al. observed exactly the same for their youngest group: Interference was maximal at the age of 7 and monotonically decreased until the age of 13. The decline was steep, going from 1,380 ms to 680 ms.

These findings argue against strength of processing and automaticity accounts of the Stroop conflict. On these accounts, words interfere with color naming because reading is a more practiced and hence a more automatic (e.g., MacLeod & Dunbar, 1988; Posner & Snyder, 1975; Logan, 1980) or stronger (Cohen et al., 1990) process than color naming. The more automatic a process is, the more it is capable of interfering with a less automatic process. In adults, reading is assumed to be very automatic and color naming much less so. Hence, words interfere with color naming but colors do not interfere with reading. However, this predicts that the Stroop conflict should increase as reading becomes more practiced in children in the course of learning to read. Or, if degree of automaticity is directly reflected in processing speed, the con-

flict should remain constant, given the findings of Ligon (1932) and Comalli et al. (1962) on reading and naming times. However, the empirical findings show exactly the opposite pattern: The Stroop conflict for color naming decreases with increased practice.

WEAVER++ captures the Stroop performance across the life span if one assumes a strengthening and weakening with age of control structures rather than of processing pathways. Children in Grade 1 are learning to read, so their color naming is minimally affected by written words. When reading becomes more practiced, Stroop interference arises. And when the attentional control over the reading and naming pathways becomes fully established, the amount of interference decreases. The amount of interference increases again when at older ages control structures become less effective. (For a review of the evidence on the effect of aging on attentional control, see, e.g., Hasher, Zacks, & May, 1999.)

Figure 26 shows the U-shaped function observed by Comalli et al. (1962). Their participants from ages 7 to 13 were drawn from a grammar school; participants ages 17–19 were undergraduate students; age groups 25–34 and 35–44 were drawn from an evening college; and participants from the age group 65–80 were drawn from a community senior club. The study used the original form of the Stroop task. The Stroop conflict is the difference between the completion time of color naming for the card with color patches and for the card with color words each written in a conflicting ink color. The interference scores represent the average difference in completion times (in seconds) between the incongruent and control condition for each age group. Figure 26 also shows the simulation results. The conditions are matched for baseline performance. When the duration of the word input (parameter du) is increased by a factor of 1.75 from normal to weakened attentional control, the model fits the performance across age groups.

Pathological Stroop Interference in Schizophrenia (Finding 16)

Several studies have addressed the increased interference observed with “damaged” Stroop performance. There are neuropsy-

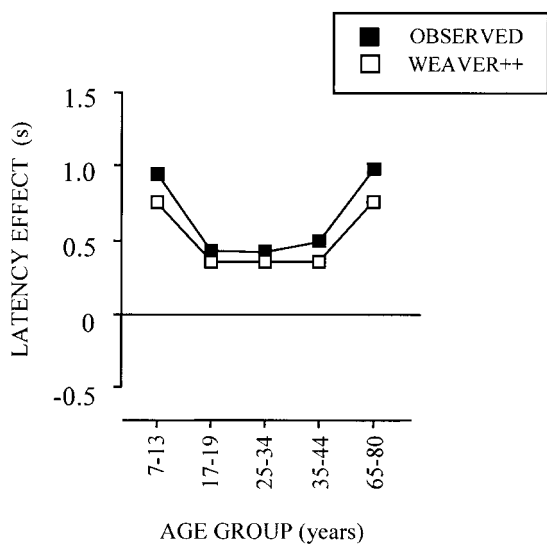


Figure 26. Changes in Stroop interference from 7 to 80 years of age: Empirical data (Comalli et al., 1962) and WEAVER++ simulations.

chological studies testing the performance of persons with frontal lobe lesions (e.g., Janer & Pardo, 1991) and individuals diagnosed with dementia of the Alzheimer type (e.g., Spieler et al., 1996) as well as neuropsychiatric studies examining the performance of individuals diagnosed with schizophrenia (e.g., Abramczyk et al., 1983; Wapner & Krus, 1960; Wysocki & Sweet, 1985). Common among these groups is that the Stroop interference observed is much higher than in control groups. The increase is disproportionate because it is higher than one would expect on the basis of slower processing speed, similar to the increased interference that is observed with young children and older adults (e.g., individuals with dementia of the Alzheimer type show interference proportionate to older adults; see Spieler et al., 1996). Here, I address Stroop performance by persons with schizophrenia, but the general line of reasoning applies to the other groups as well.

It has been claimed that selective attention is one of the mechanisms that is deficient in persons with schizophrenia. One line of evidence for this claim comes from their performance in Stroop experiments. Persons with schizophrenia show much increased interference from incongruent stimuli compared with normal controls (e.g., Abramczyk et al., 1983; Wapner & Krus, 1960; Wysocki & Sweet, 1985). Schizophrenia is often tied to hypometabolism of the frontal cortex due to a reduction of its dopaminergic supply (e.g., Meltzer & Stahl, 1976). Apparently, losing their capabilities to orient their actions to goals renders persons with schizophrenia more reliant on stimulus input activity in the posterior regions of the brain. Because persons with schizophrenia are easier to distract, they fail to screen out irrelevant representations, which gives rise to symptoms such as hallucinations, and they have difficulty in suppressing the irrelevant component of a Stroop stimulus.

Cohen and Servan-Schreiber (1992) showed that their model could mimic the Stroop performance of persons with schizophrenia. In “lesioning” the model, they compared two possibilities, namely reducing the rate of spreading activation in the network (mimicking a general deficit in information processing) and reducing the input from the task nodes to the rest of the network (mimicking an attention deficit). Reducing the spreading rate slowed down the production latencies in the model but did not yield the pathological increase in interference, which was over 50% in all three studies examined (Abramczyk et al., 1983; Wapner & Krus, 1960; Wysocki & Sweet, 1985). By contrast, reducing the task input to the network gave the pathological result.

WEAVER++ behaves similarly. Decreasing the spreading rate increases the Stroop interference in the model, but this increase is never more than about 25% of the standard interference effect. It appears that a pathological increase can only be obtained in the model by assuming an attentional deficit. Figure 27 shows the results of simulations assuming a problem in controlling the gain of the word input relative to the color input. When the duration of the word input is increased by a factor of 1.75 from normal to pathological performance, the model captures the performance of participants with schizophrenia. Thus, as in the model of Cohen et al. (1990), it is more plausible in WEAVER++ to attribute the deficit shown by persons with schizophrenia in the Stroop task to impaired selective attention rather than to a general deficit in information processing.

The parameter manipulation that yielded the pathological results for schizophrenia is the same as the manipulation proposed for the

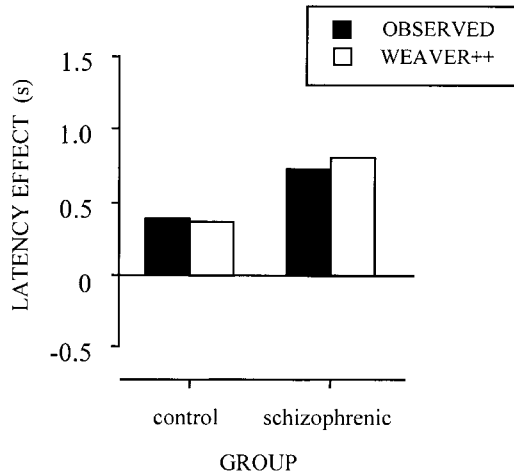


Figure 27. Stroop performance by persons with schizophrenia and by control participants. Observed effects (mean data from Abramczyk et al., 1983; Wapner & Krus, 1960; Wysocki & Sweet, 1985) and WEAVER++ simulations.

fall and rise of Stroop interference with age in the previous section. This suggests that the functional basis (a problem in the attentional control of action) is the same in both cases, but it does not, of course, imply a common underlying biological cause. Schizophrenia is presumably the result of frontal lobe dysfunction due to a reduction in dopaminergic supply, whereas the increased conflict in children and old age is due to not having fully installed or practiced appropriate control structures or to diminished efficiency, respectively.

Agreement Between Model and Data

Now that I have applied the model to the 16 key findings from Table 1, the data fit may be more formally evaluated. Overall, there appears to be good agreement between the Stroop latency effects produced with simulations of WEAVER++ and the real data. With a few exceptions, the simulated and observed effects are even close in magnitude. In Figure 13 (showing the classic SOA patterns), for example, the facilitation effect is maximally about 30 ms, both simulated and observed. The interference effect is maximally about 80 ms, both simulated and observed. Nonetheless, it seems useful to compute a measure of goodness of fit to guide our intuitions about the agreement between model and data.

Figure 28 shows the relation between the simulated latency effects and those observed in the literature that are discussed in this article. The figure displays only the latencies from the individual-item version of the Stroop task, because including the effects from the card version (which are on the order of 1 s) would reduce the resolution of the graph too much. For the same reason, the absolute production latencies from the training study are omitted. As a measure of goodness of fit between model and data, I computed a Pearson product-moment correlation coefficient. The correlation between simulated and observed latency effects (including those from the card version and the training data) is $r = .98$ ($N = 250$), with $p < .0001$. This means that the model accounts for 96% of the variance in the real data. Thus, even though I did not attempt to fit

the data quantitatively but only the ordinal properties that are specified in Table 1, the model can readily stand a quantitative test.

Good data fits are very important, but fits alone do not constitute strong support for a particular model. Therefore, I have made an effort in this article to show that extant Stroop models (i.e., a translation model, a task-competition model, models without lemmas, etc.; Cohen et al., 1990; Phaf et al., 1990) cannot handle critical aspects of the data whereas WEAVER++ can. It is, of course, possible that a model does better than other models because it is more flexible. However, this appeared not to be the case for WEAVER++. I have shown that the model's critical behavior persists when the values of its parameters are varied. Thus, WEAVER++ is constrained and can fit only certain data patterns. In addition, the new empirical work in this article falsified an assumption of another model (Experiment 1), supported an assumption of WEAVER++ (Experiment 2), and confirmed a critical prediction (Experiment 3).

The data addressed in this article concerned robust findings from over half a century of Stroop research. These Stroop effects, given the richness of the literature, exhibit considerable variability in size, but they only rarely differ in their nature. MacLeod (1991) provided an in-depth discussion of the range of the data. For example, interference typically falls within the 75–150 ms range and facilitation within the 20–50 ms range. However, although there are size differences, maximal interference is always observed when the target and distractor begin within 100 ms of each other, and interference always increases with decreasing preexposure time. Size differences are readily accommodated by the model. As Figure 14 shows, parameter variations in WEAVER++ yield the same time course of interference with different magnitudes. Also, the model accommodates differences in effect size between the individual item and the card version of the Stroop task.

One may argue that if many of the specific assumptions that WEAVER++ implements were to be made in similar fashion for the other models, they might fit the data equally well. But this is exactly the point of this article. I have argued that Stroop perfor-

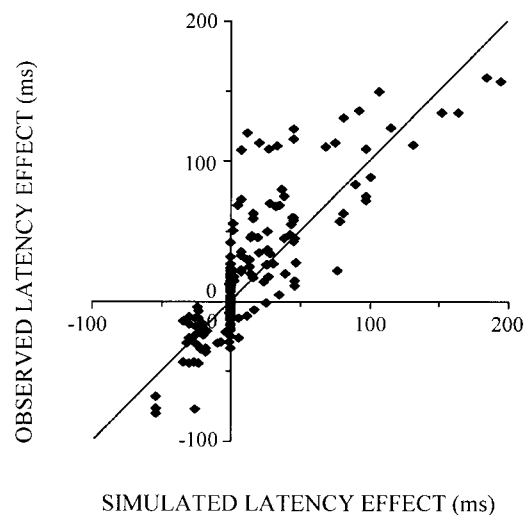


Figure 28. Relation between simulated latency effects and the corresponding effects observed in the literature for the individual-item version of the Stroop task. The diagonal indicates where simulated = observed.

mance can best be understood as arising from the specific organization of the spoken-word production architecture and task-relevant control achieved by explicit reference to goals, as implemented in WEAVER++. For example, I have argued that color naming and word reading differ because of an architectural distinction rather than a difference in pathway strength. The architectural distinction in WEAVER++, together with the discreteness assumption, helps to explain the absence of a reverse Stroop effect in word reading, even after extensive training on color naming. In addition, verification rather than an activation bias helps to explain maximal interference in color naming around SOA = 0 ms. An architectural distinction alone between color naming and word reading appears to be insufficient to account for the data: The model of Phaf et al. (1990) implements such a distinction but incorrectly yields a reverse Stroop effect in word reading and fails to account for the time course of interference in color naming. A specific architecture together with verification accounts for the data.

To conclude, in this article I have shown that extant Stroop models cannot handle critical aspects of known findings, whereas WEAVER++ can. Importantly, WEAVER++ does not account for the data because it is more flexible than the other models. Overall, the analyses in this article support the view that Stroop performance emerges from the specific organization of the speech production architecture and task-relevant control achieved through explicit reference to goals. The details of the architecture that lead to the effects are independently motivated to account for various word-production phenomena.

Final Remarks

One of the key tasks of the human cognitive system is to select one appropriate action at any given moment and to focus the machinery of movement on that action. The “singleness of action” (Sherrington, 1947, p. 235) forces us to make choices. Some theoreticians (e.g., Allport, 1993) even go so far as to claim that there is selectivity of attention only because of the need for coherent control of action (“selection for action”; Allport, 1993, p. 195) rather than to protect a limited-capacity system from information overload, as is traditionally assumed (e.g., Broadbent, 1958). The combination of a spreading activation network with a parallel system of production rules yields a simple but powerful and efficient device for selecting one line of action among the available options (e.g., Anderson, 1983). The crucial role of spreading activation is to provide a relevance heuristic. This was also Quillian’s (1969) seminal argument for why it is a good artificial intelligence mechanism.

Spreading activation attempts to solve the so-called frame problem that confronts any cognitive system, which has been considered to be among the fundamental problems in cognitive science (e.g., Pylyshyn, 1987). In his *Consciousness Explained*, Dennett (1991) referred to it as Plato’s problem (the problem of Plato’s aviary). In making decisions, a cognitive system can, in principle, draw on all the information available, but the amount may be indefinitely large in that everything may potentially be relevant. The frame problem is how to get at the relevant information (“how to get the right bird out of the aviary”) and when to stop thinking and start acting. Spreading activation is a parallel mechanism for making relevant information available and triggering relevant

computations, following the heuristic that information connected to the current information is likely of direct relevance, too. Triggering production rules by spreading activation prevents the problem of all production rules having to test their conditions at any moment in time (e.g., Miller & Johnson-Laird, 1976, p. 286, proposed a hierarchical system, “table linkage,” to solve this problem for word retrieval; for a discussion of problems with their proposal, see Levelt, 1989). Only the production rules that are linked to a sufficiently activated piece of memory become evaluated. In naming a color, no more than a dozen or so production rules (for color terms) will test their conditions rather than all production rules in a speaking lexicon of some 30,000 words (Levelt, 1989).

Before the cognitive (counter-) revolution in psychology in the late 1950s, associationist and behaviorist theories accounted for action selection by postulating connections between stimuli and responses (S-R associations; cf. Skinner’s, 1957, work on verbal behavior). However, if all our actions were determined exclusively by S-R associations, we would not be free to choose which action to make to a certain stimulus, because the strongest association would determine the action. If we ask participants to perform one particular task rather than another, they are able to do it, usually without much practice. In one way or another, goals can be set to control actions. The Würzburg school (with Bühler, a pioneer of psycholinguistics) extended the idea of S-R associations to associations between stimuli and an internally represented task (*Aufgabe*), on the one hand, and responses, on the other. The connectionist accumulator and attractor models of Cohen et al. (1990) and Phaf et al. (1990) can be regarded as direct descendants of these ideas. The models achieve goal-directed performance by wiring a task specification as a node into the network. However, if all possible tasks had to be wired into the network beforehand, the system would run into a combinatorial explosion. Furthermore, tasks such as “name the first color” and “name the second color” share part of their specification. Thus, models that represent tasks by single nodes are missing a generalization.

Verification models such as WEAVER++ also bear a resemblance to these ideas, relating a set of conditions, including a goal, to an action that occurs when the conditions are satisfied. But there is also an important difference: In securing task-relevant control, verification models use goal symbols, whose presence constitutes one of the conditions for the firing of a production rule (cf. Anderson, 1983; Anderson & Lebiere, 1998; Newell, 1990). When the goal symbol is placed in working memory, the attention of the system is focused on those production rules that include this goal among their conditions. Actions are selected with explicit reference to the goals, whereas in an associationist approach, selection is “blind” in that the first element that reaches the response criterion is selected without further ado (cf. Pashler, Johnston, & Ruthruff, 2001).

Production rules provide flexibility in the response selection system by allowing tasks to be specified through the combination of task-relevant constructs such as “name,” “first,” and “color” rather than through a unitary task node. As shown here, the production rules for Stroop task performance do not need to be complex. The condition of a production rule always began with a test for a goal, which was followed by maximally two other simple tests. The tests were combined into a logical conjunction; there were no disjunctions and negations. The tests referred to constructs

from a small set. Clearly, not all tasks that can be specified by production rules can actually be performed; similarly, although a task node for any imaginable task may be included in an accumulator or attractor network, this does not mean that the task can be done. Apart from combinatorial restrictions, the operation of production rules is constrained by other systems providing relevant information, such as which color came first, and by the amount of information that can simultaneously be held in working memory (e.g., Anderson & Lebiere, 1998; Baddeley, 1986).

Production rules are sometimes criticized for not being brainlike in their computation (see some of the peer commentaries on Levelt et al., 1999b). However, production rules mean nothing more than the operations that they specify. Crucial for the issue of “neural plausibility” is whether we can exclude that the brain performs such if-then operations, and the criticisms do not bring forward evidence for that. Clearly, there is no direct way to observe whether verbal actions are selected by purely associationist means or by verification. We can only make use of indirect evidence coming from behavioral and neural studies. In this article, I have applied the WEAVER++ model to one widely used task and analogs of it, the Stroop task. The model accounted for the basic Stroop phenomenon and for performance in variants of the task that manipulate the SOA, response-set membership, semantic distance, stimulus dimensions, task and location uncertainty, response mode and type, basis of responding, and amount of training. Also, findings from the picture-word task, different age groups, bilingual people, and clinical groups were explained. Models are moving targets: It remains to be seen whether a purely associationist approach can ultimately be saved and whether it then can account for these findings. In the meantime, WEAVER++ provides a clearly testable alternative account.

References

- Abramczyk, R. R., Jordan, D. E., & Hegel, M. (1983). “Reverse” Stroop effect in the performance of schizophrenics. *Perceptual and Motor Skills*, *56*, 99–106.
- Allport, D. A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of 25 years. In D. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: A silver jubilee* (pp. 183–218). Cambridge, MA: MIT Press.
- Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. London: Erlbaum.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Brown, W. (1915). Practice in associating color-names with colors. *Psychological Review*, *22*, 45–55.
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*, 215–222.
- Caplan, D. (1992). *Language: Structure, processing, and disorders*. Cambridge, MA: MIT Press.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.
- Carr, T. H. (1999). How does WEAVER pay attention? *Behavioral and Brain Sciences*, *22*, 39–40.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.
- Charniak, E., & McDermott, D. (1985). *Introduction to artificial intelligence*. Reading, MA: Addison-Wesley.
- Cohen, J., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.
- Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and non-conscious information processing* (pp. 453–476). Cambridge, MA: MIT Press.
- Cohen, J., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in Schizophrenia. *Psychological Review*, *99*, 45–77.
- Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity: Principles and models. *American Journal of Psychology*, *105*, 239–269.
- Cohen, J. D., Usher, M., & McClelland, J. (1998). A PDP approach to set size effects within the Stroop task: Reply to Kanne, Balota, Spieler, and Faust (1998). *Psychological Review*, *105*, 188–194.
- Cohn, N. B., Dustman, R. E., & Bradford, D. C. (1984). Age-related decrements in Stroop color test performance. *Journal of Clinical Psychology*, *40*, 1244–1250.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.
- Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology*, *100*, 47–53.
- Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 318–344.
- Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the Stroop color-word task. *Perception and Psychophysics*, *11*, 274–276.
- Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the Stroop color-word test. *Perceptual and Motor Skills*, *23*, 1211–1214.
- Dennett, D. C. (1991). *Consciousness explained*. London: Penguin Books.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. (Original work published 1868)
- Dunbar, K. N., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 622–639.
- Duncan, J. (1986). Disorganization of behavior after frontal lobe injury. *Cognitive Neuropsychology*, *3*, 271–290.
- Duncan-Johnson, C. C., & Kopell, B. S. (1980). The locus of interference in a Stroop task: When you read “blue,” do you see “red”? *Psychophysiology*, *17*, 308–309.
- Duncan-Johnson, C. C., & Kopell, B. S. (1981, November 20). The Stroop effect: Brain potentials localize the source of interference. *Science*, *214*, 938–940.
- Dyer, F. N. (1971a). Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *10*, 297–302.
- Dyer, F. N. (1971b). The duration of word meaning responses: Stroop interference for different preexposures of the word. *Psychonomic Science*, *25*, 229–231.
- Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, *1*, 106–120.
- Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and

- encoding factors in "ignoring" irrelevant information. *Memory & Cognition*, 7, 86–94.
- Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychologica*, 30, 96–103.
- Francis, W. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin*, 125, 193–222.
- Gatti, S. V., & Egeth, H. E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society*, 11, 181–184.
- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience: The biology of the mind*. New York: Norton.
- Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 875–894.
- Glaser, W. R. (1992). Picture naming. *Cognition*, 42, 61–105.
- Glaser, W. R., & Dünghoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 640–654.
- Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118, 13–42.
- Goolkasian, P. (1981). Retinal location and its effect on the processing of target and distractor information. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1247–1257.
- Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriati (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.
- Hillyard, S. A., Mangun, G. R., Woldorff, M. G., & Luck, S. J. (1995). Neural systems mediating selective attention. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 665–681). Cambridge, MA: MIT Press.
- Hinton, G., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74–95.
- Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, 83, 299–303.
- Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 845–865). Cambridge, MA: MIT Press.
- Jahanshahi, M., & Frith, C. D. (1998). Willed action and its impairment. *Cognitive Neuropsychology*, 15, 483–533.
- Janer, K. W., & Pardo, J. V. (1991). Deficits in selective attention following bilateral anterior cingulotomy. *Journal of Cognitive Neuroscience*, 3, 231–241.
- Just, M. A., & Carpenter, P. N. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 497–509.
- Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.
- Kanne, S. M., Balota, D. A., Spieler, D. H., & Faust, M. E. (1998). Explorations of Cohen, Dunbar, and McClelland's (1990) connectionist model of Stroop performance. *Psychological Review*, 105, 174–187.
- Kanwisher, N., & Driver, J. (1992). Objects, attributes, and visual attention: Which, what, and where. *Current Directions in Psychological Science*, 1, 26–31.
- Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, 77, 576–588.
- Lassen, G. L. (1975). Extension of the Stroop interference effect to pictures and words. (Doctoral dissertation, University of Arizona, 1974). *Dissertation Abstracts International*, 35, 0223-B.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42, 1–22.
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3, 223–232.
- Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, 10, 553–567.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999a). Multiple perspectives on word production. *Behavioral and Brain Sciences*, 22, 61–75.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999b). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122–142.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.
- Ligon, E. M. (1932). A genetic study of color naming and word naming. *American Journal of Psychology*, 44, 103–122.
- Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, 12, 523–553.
- Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, 7, 166–174.
- Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 978–992.
- Long, G. M., & Lyman, B. J. (1987). Foveal and parafoveal processing of asynchronous Stroop stimuli. *British Journal of Psychology*, 78, 151–162.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203.
- MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General*, 121, 12–14.
- MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, 26, 201–211.
- MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 126–135.
- MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, 4, 383–391.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.
- McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309–360). New York: Wiley.
- Meltzer, H. Y., & Stahl, S. M. (1976). The dopamine hypothesis of schizophrenia: A review. *Schizophrenia Bulletin*, 2, 19–76.
- Mewhort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response time distributions and the Stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 872–882.

- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects. *Cognition*, *66*, B25–B33.
- Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review*, *104*, 3–65.
- Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.
- Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.
- Neumann, O. (1980). *Informations-selektion und Handlungssteuerung* [Selection of information and control of action]. Unpublished doctoral dissertation, University of Bochum, Germany.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum Press.
- Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, *52*, 629–651.
- Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 539–557.
- Peterson, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, *1*, 153–170.
- Peterson, S. E., Fox, P. T., Snyder, A. Z., & Raichle, M. E. (1990, August 31). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science*, *249*, 1041–1044.
- Phaf, R. H., Van der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, *22*, 273–341.
- Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.
- Posner, M. I., & DiGirolama, G. J. (2000). Attention in cognitive neuroscience: An overview. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 623–631). Cambridge, MA: MIT Press.
- Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. New York: Freeman.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.
- Proctor, R. W. (1978). Sources of color–word interference in the Stroop color-naming task. *Perception and Psychophysics*, *23*, 413–419.
- Pylyshyn, Z. W. (Ed.). (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.
- Quillian, M. R. (1969). The teachable language comprehender. *Communications of the ACM*, *12*, 459–476.
- Rafal, R., & Henik, A. (1994). The neurology of inhibition: Integrating controlled and automatic processes. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language*. San Diego, CA: Academic Press.
- Regan, J. E. (1978). Involuntary automatic processing in color-naming tasks. *Perception and Psychophysics*, *24*, 130–136.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.
- Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*, 59–87.
- Roelofs, A. (1994). On-line versus off-line priming of word-form encoding in spoken word production. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 772–777). Hillsdale, NJ: Erlbaum.
- Roelofs, A. (1996a). Computational models of lemma retrieval. In T. Dijkstra & K. De Smedt (Eds.), *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 308–327). London: Taylor & Francis.
- Roelofs, A. (1996b). Morpheme frequency in speech production: Testing WEAVER. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp. 135–154). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Roelofs, A. (1996c). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, *35*, 854–876.
- Roelofs, A. (1997a). A case for nondecomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research*, *26*, 33–67.
- Roelofs, A. (1997b). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, *12*, 657–693.
- Roelofs, A. (1997c). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.
- Roelofs, A. (1998a). Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition*, *1*, 94–95.
- Roelofs, A. (1998b). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 904–921.
- Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*, 173–200.
- Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 922–939.
- Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 246–251.
- Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma–lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition*, *69*, 219–230.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.
- Schiller, P. H. (1966). Developmental study of color–word interference. *Journal of Experimental Psychology*, *72*, 105–108.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, England: Cambridge University Press.
- Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*, 727–741.
- Sherrington, C. (1947). *The integrative action of the nervous system*. Cambridge, England: Cambridge University Press.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture–word processing. *Journal of Experimental Psychology: General*, *109*, 373–392.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance

- in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 461–479.
- Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 896–918.
- Stemberger, J.-P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). London: Erlbaum.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Styles, E. A. (1997). *The psychology of attention*. Hove, England: Psychology Press.
- Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 647–675.
- Taylor, D. A. (1977). Time course of context effects. *Journal of Experimental Psychology: General*, 106, 404–426.
- Thomas, J. K. (1977). Stroop interference with word or hue pre-exposure (Doctoral dissertation, University of Michigan, 1977). *Dissertation Abstracts International*, 38, 1450-B.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, England: Cambridge University Press.
- Van der Heijden, A. H. C. (1981). *Short-term visual information forgetting*. London: Routledge.
- Van der Heijden, A. H. C. (1992). *Selective attention in vision*. London: Routledge.
- Van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 787–806.
- Van Turennout, M., Hagoort, P., & Brown, C. M. (1998, April 24). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, 280, 572–574.
- Virzi, R. A., & Egeth, H. E. (1985). Toward a translation model of Stroop interference. *Memory & Cognition*, 13, 304–319.
- Vorberg, D. (1985). Unerwartete Folgen von Zufälliger Variabilität: Wettlauf-Modelle für den Stroop-Versuch [Unexpected consequences of random variability: Race models of the Stroop task]. *Zeitschrift für experimentelle und angewandte Psychologie*, 32, 494–521.
- Wapner, S., & Krus, D. M. (1960). Effects of lysergic acid diethylamide, and differences between normals and schizophrenics on the Stroop color-word test. *Journal of Neuropsychiatry*, 2, 76–81.
- Wysocki, J. J., & Sweet, J. J. (1985). Identification of brain-damaged, schizophrenic, and normal medical patients using a brief neuropsychological screening battery. *International Journal of Clinical Neuropsychology*, 7, 40–49.

Appendix A

Details of the Computer Simulations

This appendix gives the equations for the spreading of activation, the selection ratio, and the expected latencies in WEAVER++. Activation spreads according to

$$a(m, t + \Delta t) = a(m, t)(1 - d) + \sum_n ra(n, t),$$

where $a(m, t)$ is the activation level of node m at point in time t , d is a decay rate ($0 < d < 1$), and Δt is the duration of a time step in ms. The rightmost term denotes the amount of activation that m receives between t and $t + \Delta t$, where $a(n, t)$ is the output of neighbor n (equal to its level of activation). The factor r indicates the spreading rate.

The probability that a target node m will be selected at $t < T \leq (t + \Delta t)$, given that it has not been selected at $T \leq t$, and provided that the conditions for its selection are satisfied, is given by the ratio

$$p(\text{selection } m \text{ at } t < T \leq (t + \Delta t) \mid \text{no selection } m \text{ at } T \leq t) = \frac{a(m, t)}{\sum_k a(k, t)}.$$

The index k ranges over the lemma nodes and the syllable program nodes in the network (i.e., the lemmas *red*, *green*, etc., and the syllable program nodes [red], etc., as illustrated in Figure 8). The selection ratio equals the hazard rate $h(s)$ of lemma retrieval and word-form encoding at time step s (cf. Luce, 1986; McGill, 1963; Townsend & Ashby, 1983), where $t = (s-1)\Delta t$, and $s = 1, 2, \dots$. Roelofs (1992) gave a derivation of the mathematical expected lemma retrieval time in the model, which is

$$E(T) = \sum_{s=1}^{\infty} \left[h(s) \prod_{j=0}^{s-1} [1 - h(j)] s \Delta t \right].$$

The expected latency, including form encoding, $E(T)$, for monosyllabic words, equals

$$E(T) = \sum_{s=1}^{\infty} \left\{ \left\{ \sum_{u=s}^{\infty} \left(h_x(u) \prod_{j=s}^{u-1} [1 - h_x(j)] u \Delta t \right) \right\} h_L(s) \times \left\{ \prod_{j=0}^{s-1} [1 - h_x(j)] \right\} \right\},$$

where $h_L(s)$ and $h_x(s)$ are the hazard rates of lemma retrieval and word-form encoding, respectively (all details of form encoding can be found in Roelofs, 1997c). The parameters for lemma retrieval are

$$r_C = 0.0101 \text{ ms}^{-1},$$

$$r_L = 0.0074 \text{ ms}^{-1},$$

$$d = 0.0240 \text{ ms}^{-1},$$

$$\text{extin} = 0.1965 \text{ activation units ms}^{-1},$$

$$du = 100 \text{ ms},$$

$$\text{critdiff} = 1.6 \text{ activation units, and}$$

$$\Delta t = 25 \text{ ms}.$$

Scaling up the model by using a larger network with all basic color terms in English and by including word-form encoding required a few small adjustments to parameters. By increasing the size of the conceptual network four times, the amount of activation arriving at a node also increased considerably. To compensate for this increase, the spreading

rate was reduced somewhat (set to 0.7 r_C), and to compensate for the fact that the denominator of the selection ratio ranged over four times as many nodes, the target activation in the ratio was multiplied by four. Including word-form encoding in the simulations led to a small overestimation of congruency effects at short stimulus onset asynchronies (the extra stage amplified effects somewhat), which was remedied by increasing the onset latency of the task production rule (P1) by one time step.

Also, a few parameter values were adjusted to accommodate task changes. In the simulations of picture naming, critdiff was set to 3.3 in the

incongruent and to 3.0 in the congruent condition and bias was set to 4.0; in the simulations of manual responding, critdiff was set to 0.8 and bias was set to 8.0 for the word distractors; and in the simulations of oral reading, critdiff was set to 4.0, du to 200 ms, and the goal activation was reduced by half (a to-be-read word already receives considerable activation from its orthography).

The reported WEAVER++ simulations have been programmed in the C/C++ language using the Microsoft Visual C++ 6.0 environment. WEAVER++ may be obtained from Ardi Roelofs or retrieved at <http://www.mpi.nl/world/persons/private/ardi/weaver++.htm>

Appendix B

Materials of Experiment 3

dier (animal): de zwaan (swan) de schildpad (tortoise) het konijn (rabbit) het hert (deer)	wapen (weapon): de dolk (dagger) de speer (spear) het kanon (cannon) het pistool (pistol)
kleding (clothing): de trui (sweater) de rok (skirt) het hemd (shirt) het vest (vest)	servies (service): de beker (cup) de kom (bowl) het glas (glass) het bord (plate)
vervoer (transportation): de fiets (bike) de trein (train) het schip (ship) het vliegtuig (airplane)	meubel (furniture): de tafel (table) de kast (cupboard) het bed (bed) het bureau (desk)
gebouw (building): de molen (windmill) de fabriek (factory) het kasteel (castle) het paleis (palace)	lichaamsdeel (body part): de arm (arm) de neus (nose) het been (leg) het oor (ear)

Received May 9, 2000
Revision received November 15, 2001
Accepted January 22, 2002 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://watson.apa.org/notify/> and you will be notified by e-mail when issues of interest to you become available!