6 Modelling the control of visual attention in Stroop-like tasks

Ardi Roelofs[1,2,3] **and Martijn Lamers**[3]
[1]*Max Planck Institute for Psycholinguistics, The Netherlands*
[2]*F. C. Donders Centre for Cognitive Neuroimaging, The Netherlands*
[3]*Nijmegen Institute for Cognition and Information, The Netherlands*

Introduction

In naming one of several objects in a visual scene or in reading aloud one of several words in a text, a speaker has to perform a number of interrelated attentional tasks. The first major task is to select one particular object or word for further processing. This usually involves moving the gaze to the spatial location of the relevant object or word. Objects and words are multi-dimensional entities. For example, they usually have a colour and shape, and they consist of several parts (i.e., written words are made up of letters). The second major task is therefore to select the action-relevant dimension, such as the shape of the objects and words rather than their colour. For words, responding to the shape rather than their colour is the default. Finally, if the objects and words are spatially close together, the speaker has to plan and execute the appropriate action in the face of distraction by the other objects and words. This constitutes the third major task.

In a number of influential publications, Posner and colleagues (e.g., Posner, 1994; Posner & Peterson, 1990; Posner & Raichle, 1994) argued that the three major tasks described above are achieved by two different attentional systems. They postulated a perceptual attention system that controls the orienting of attention and gaze in order to select the appropriate input (the first major task). Moreover, they postulated an executive attention system that helps to achieve selective perceptual processing of the target dimension (the second major task) and that controls the selection of appropriate actions (the third major task). It was hypothesized that the executive system also controls the orienting of attention.

According to Posner and colleagues, directing attention from one location

to another involves disengaging attention from the current location, moving attention to the new location of interest, and engaging it there. The disengage, move, and engage operations were associated with, respectively, the posterior parietal cortex of the human brain, the superior colliculus in the midbrain, and the pulvinar nucleus of the thalamus. Executive attention was associated with the lateral prefrontal and anterior cingulate cortices. It was assumed that the allocation of attention to a spatial location causes modulatory changes in extrastriate visual areas involved in the analysis of stimulus features. Selection of a dimension is less specified than the allocation of spatial attention in Posner's framework, but one presumes it is achieved through interactions between the anterior and posterior attentional networks (e.g., Posner & Raichle, 1994).

In this chapter, we provide an overview of recent work that examined these attentional systems in object-naming and word-reading tasks. We restrict ourselves to the visual domain and to vocal responding, in particular, the interplay between visual attention and vocal responding in Stroop-like tasks such as the classic colour-word Stroop task (MacLeod, 1991; Stroop, 1935) and various picture-word versions of it (e.g., Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Lupker, 1979; Smith & Magee, 1980). A major way to study attention is to introduce conflict. In the classic colour-word Stroop task (Stroop, 1935), participants are asked to name the ink colour of written colour words, such as the word RED in blue ink. Results have consistently shown that people are much slower in naming the ink colour of incongruent colour words than in naming the ink colour of a row of neutral Xs, an effect called Stroop interference, and people are fastest when colour and word are congruent (e.g., BLUE in blue ink), an effect called Stroop facilitation. The literature documents numerous manipulations of the basic Stroop interference and facilitation effects (reviewed by MacLeod, 1991), providing evidence on the nature of the attentional control systems.

A theory of attentional control is limited by the theories of the processes that are controlled. To know what attentional control does in Stroop-like tasks, it is essential to know what the subordinate processes of perception and vocal response planning do in these tasks. In the first section, we provide a short overview of a computationally implemented theory for vocal responding and its attentional control. The theory provides an account of the various processes underlying spoken word production and its relation with word and object/colour recognition (Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992, 1997) and an account of how attention controls word production and perception (Roelofs, 2003). The theory has been computationally implemented in the WEAVER++ model. In the next section, we address the issue of how visual orienting, selective stimulus processing, and vocal response planning are related. In particular, is visual orienting dependent only on visual processing (e.g., D. E. Meyer & Kieras, 1997; Sanders, 1998) or also on verbal response planning? We describe recent work that suggests that visual orienting in Stroop-like tasks is jointly determined by visual

processing and response planning. In the third section, we address the issue of selective perceptual processing in Stroop-like tasks. We review the literature on the role of space-based and object-based attention and on dimensional selection in the Stroop task. Moreover, we report on computer simulations of a paradoxical finding using WEAVER++. In the final section, we briefly discuss evidence on the neural correlates of executive attention in Stroop-like tasks.

An outline of WEAVER++

WEAVER++ is a computational model designed to explain how speakers plan and control the production of spoken words. The model falls into the general class of "hybrid" models of cognition in that it combines a symbolic associative network and condition-action rule system with spreading activation and activation-based rule triggering. The model plans spoken words by activating, selecting, and connecting (weaving together) types of verbal information. The model gives detailed accounts of chronometric findings on spoken word production (e.g., Levelt et al., 1999; Roelofs, 1992, 1997). Recently, WEAVER++ has also been applied to neuroimaging data (Roelofs & Hagoort, 2002).

In a classic paper, Norman and Shallice (1986) made a distinction between "horizontal threads" and "vertical threads" in the control of behaviour. Horizontal threads are strands of processing that map perceptions onto actions, and vertical threads are attentional influences on these mappings. Behaviour arises from interactions between horizontal and vertical threads. WEAVER++ implements specific theoretical claims about how the horizontal and vertical threads are woven together in planning verbal actions (Roelofs, 2003). A central claim embodied by WEAVER++ is that the control of verbal perception and action is achieved symbolically by condition-action rules rather than purely associatively (cf. Deacon, 1997; Logan, 1995).

Until the beginning of the twentieth century, the problem of the directedness of thought, perception, and action was largely ignored. Starting with Aristotle and revitalized by Locke, the prevailing theorizing was largely associationistic. A major credit for introducing a directional notion into theorizing goes to Ach and Watt of the Würzburg school (Ach, 1905; Watt, 1905), who experimentally investigated the effects of the task, or *Aufgabe*, and they demonstrated its vital importance in determining the course of associations (e.g., Mandler & Mandler, 1964). We refer to Humphrey (1951) for a thorough and sustained treatment of what the Würzburgers did, why they did it, and how they interpreted it.

Watt (1905) presented participants with written nouns and asked them to produce partially constrained associations, like producing a superordinate term (e.g., say the hyperonym "animal" in response to the word DOG) or a name for a part (e.g., say "tail" in response to DOG). Response times were measured. Watt observed that the speed of responding was determined

independently by the type of task instruction (e.g., naming a superordinate vs. naming a part) and the strength of the association between stimulus and response given a particular task. For example, if in free word associations, "tail" was more frequently given than "neck" in response to DOG, the production latency was smaller for "tail" than for "neck" in producing a name of a part in response to DOG. These findings led Watt to propose that the direction of our thoughts and actions is determined by associations among stimuli and goals (*Aufgaben*), on the one hand, and responses, on the other. In the 1910s, this view was firmly integrated within an associationist framework and fiercely defended by Müller (1913), whereas Selz (1913) held that mental rules rather than associations are critical in mediating between stimuli/goals and responses. All current theories are descendants of these ideas.

In the view that currently dominates the attention and performance literature (e.g., computationally implemented in GRAIN by Botvinick, Braver, Barch, Carter, & Cohen, 2001), goals associatively bias the activation of one type of response (e.g., colour naming in the Stroop task) rather than another (e.g., oral reading), following Müller (1913). WEAVER++ implements a third theoretical alternative (considered but rejected by Selz, 1913), in which both rules and associations play a critical role. WEAVER++'s lexical network is accessed by spreading activation while the condition-action rules determine what is done with the activated lexical information depending on the task. When a goal symbol is placed in working memory, the attention of the system is focused on those rules that include the goal among their conditions. These rules ensure that colour naming rather than word reading is performed and that the irrelevant perceptual input is suppressed (Roelofs, 2003). Both GRAIN (Botvinick et al., 2001) and WEAVER++ have been applied to the colour-word Stroop task. Elsewhere (Roelofs, 2003; Roelofs & Hagoort, 2002), WEAVER++ is compared in great detail with GRAIN and other implemented models of attentional control in the Stroop task. However, to provide such detailed model comparisons for the modelling of visual attention in Stroop-like tasks is outside the scope of the present chapter.

As indicated, to know what attentional control does, it is essential to know what the subordinate processes of perception and vocal response planning do. As concerns conceptually driven naming (e.g., object naming), the WEAVER++ model distinguishes between conceptual preparation, lemma retrieval, and word-form encoding, with the encoding of forms further divided into morphological, phonological, and phonetic encoding (Levelt et al., 1999). The model assumes that memory is an associative network that is accessed by spreading activation. Figure 6.1 illustrates the structure of the network. During conceptual preparation, concepts are retrieved from memory and flagged as goal concepts. In lemma retrieval, a goal concept is used to retrieve a lemma from memory, which is a representation of the syntactic properties of a word, crucial for its use in sentences. For example, the lemma of the word *blue* says that it can be used as an adjective. Lemma retrieval
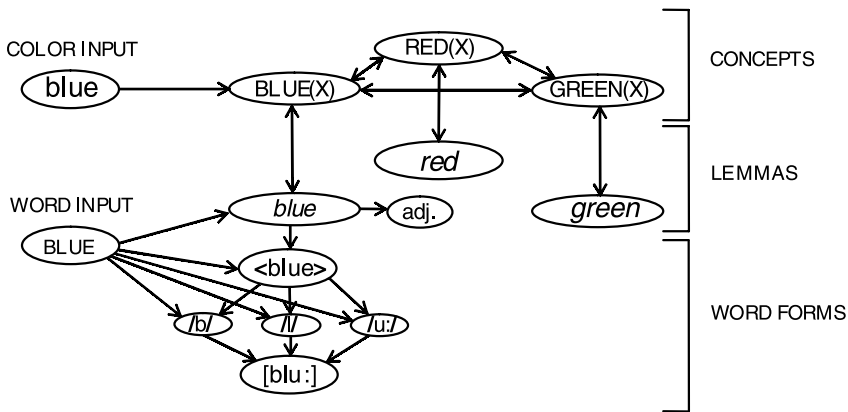
Figure 6.1 Fragment of the lexical network of WEAVER++ for colour terms (cf. Roelofs, 2003); adj. = adjective.

makes these properties available for syntactic encoding processes. In word-form encoding, the lemma is used to retrieve the morphophonological properties of the word from memory in order to construct an appropriate articulatory program. For example, for *blue*, the morpheme <blue> and the speech segments /b/, /l/, and /u:/ are retrieved, and a phonetic plan for [blu:] is generated. Finally, articulation processes execute the motor program, which yields overt speech. Perceived words (e.g., BLUE) may be read aloud by selecting a lemma (*blue*) and then encoding the corresponding word form (i.e., [blu:]) or by directly encoding a word form without first selecting a lemma (see Figure 6.1).

Let us assume that a participant in a colour-word Stroop experiment has to name the ink colour of the word RED in blue ink. This involves the conceptual identification of the colour based on the perceptual input and its designation as goal concept (i.e., BLUE(X)), the retrieval of the lemma of the corresponding word (i.e., *blue*), and the encoding of the form of the word (i.e., [blu:]). The final result is a motor program for the word "blue", which can be articulated. Perceived words activate their lemmas and word forms in parallel.

It has been shown that WEAVER++ successfully simulates several classic data sets on Stroop, mostly taken from the review by MacLeod (1991), including response set, semantic gradient, stimulus, spatial, multiple task, manual, bilingual, training, age, and pathological effects (Roelofs, 2003). With only three free parameters taking two values each to accommodate task differences (colour naming, picture naming, word reading, and manual responding), the model accounted for 96% of the variance of 16 classic studies (250 data points). Moreover, WEAVER++ successfully simulated the human brain's blood-flow response during Stroop task performance in neuroimaging studies; in particular, the functional magnetic resonance

imaging (fMRI) BOLD response in the anterior cingulate cortex, one of the classic brain areas involved with Stroop task performance (Roelofs & Hagoort, 2002).

Visual orienting

Visual acuity is best at the centre of eye fixation. By 5° from the centre, acuity has diminished about 50%. Therefore, to bring aspects of the visual world into the focus of attention, eye fixations are directed to those visual aspects that are of most interest. This makes a shift of gaze an overt sign of the allocation of attention (e.g., Kustov & Robinson, 1996), although attention and eye movements can be dissociated in simple signal detection tasks (e.g., Posner, 1980).

Over the past few decades, the control of eye movements has been intensively investigated for visual search and other cognitive tasks, such as problem solving, typing, and reading for comprehension. We refer to Rayner (1998) for an extensive review of the literature. However, the control of eye movements during the production of spoken words has only recently become a topic of interest. Whereas it has long been assumed that we look at aspects of the visual world just as long as is needed to identify them and that response factors play no role (D. E. Meyer & Kieras, 1997; Sanders, 1998), recent research suggests that when we want to respond vocally to the visual aspects, the gaze durations depend on the time to plan the corresponding words (e.g., A. S. Meyer, Griffin, 2001; Sleiderink, & Levelt, 1998). For example, when speakers are asked to name two objects in a row, they look longer at first-to-be-named objects with two- than with one-syllable names even when the object-recognition times do not differ (A. S. Meyer, Roelofs, & Levelt, 2003). The effect of the number of syllables suggests that the shift of gaze from one object to another is initiated only after the phonological form of the object name has been encoded.

In terms of the theory of Posner and colleagues, the effect of the number of syllables suggests that the endogenous visual orienting of attention is determined by response factors. This does not exclude, however, a role for factors related to visual processing. In order to examine the interplay between visual orienting, selective visual processing, and vocal response selection, we performed a series of Stroop-like experiments (Roelofs, submitted). The experiments examined what the trigger is for moving the eyes from fixated stimuli that cause more versus less interference for vocal responding. In all experiments, picture-word versions of the Stroop task were used. Instead of naming colour patches with superimposed colour words, participants responded to pictured objects with superimposed distractor words.

Earlier research demonstrated that reading the word or naming the picture of a picture-word stimulus replicates the response time patterns known from colour-word Stroop experiments (e.g., Glaser & Düngelhoff, 1984; Glaser &

Glaser, 1989; see Roelofs, 2003, for a review). This suggests that the colour of a colour-word Stroop stimulus is the limiting case of a picture. As with colour-word stimuli, WEAVER++ assumes that pictures have direct access to concepts, whereas words have direct access to lemmas and word forms (see Figure 6.1).

In the eye-tracking experiments, speakers were presented with picture-word stimuli displayed on the left side of a computer screen and left- or right-pointing arrows (flanked by Xs on each side: XX<XX or XX>XX) displayed on the right side of the screen. Figure 6.2 illustrates the visual displays that were used in the experiments. The picture-word stimuli and the arrow were presented simultaneously on the screen. The participants' task was to respond vocally to the picture-word stimulus, and to shift their gaze to the arrow stimulus in order to indicate the direction in which the arrow was pointing by pressing a left or right button. Eye movements were recorded. In particular, it was measured how long participants looked at a picture-word stimulus before they moved their gaze to the arrow. The manual task involving the arrow stimuli was given in order to be able to assess the gaze durations for the picture-word stimuli. Moreover, it is possible that the eyes remain fixated on the left picture-word stimulus while attention is covertly moved to the right side of the screen to allow processing of the arrows before a gaze shift. If this were the case, the gaze durations would not reflect the attention given to the left object. However, when the manual response latencies parallel the gaze durations (i.e., when the differences among the gaze durations are preserved by the manual response latencies), this would be evidence that preprocessing of the arrow has not taken place and that the shift of gaze corresponds to a shift of attention.

The experiments were run with three basic tasks used within the picture-word interference paradigm (e.g., Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Lupker, 1979; Smith & Magee, 1980): picture naming, word reading, and word categorizing. These tasks were performed in the context of written distractor words (picture naming) or distractor pictures (word reading and categorizing). In all experiments, vocal response latencies, gaze durations, and manual response latencies were measured. The picture-word

Figure 6.2 Illustration of the visual displays used in the eye-tracking experiments.

stimuli had varying types of relatedness between picture and word: semantic, unrelated, identical, and control. For example, participants said "swan" in response to a pictured swan (picture naming), while trying to ignore the word DOG (the semantic condition), the word VEST (the unrelated condition), the word SWAN (the identical condition), or a series of Xs (the control condition). The word-reading experiment used displays that were identical to those used for picture naming except for the control condition. The string of Xs of the control condition in picture naming was replaced by an empty rectangle (cf. Glaser & Düngelhoff, 1984). The task for the participants was to read aloud the word while ignoring the picture and then to respond to the arrows, as in the picture-naming experiment. Finally, the word-categorizing experiment used the same displays as the word-reading experiment, while the vocal task was changed from word naming to word categorizing. That is, participants responded to the written words by categorizing them (i.e., producing hypernyms) while trying to ignore the picture distractors. For example, they said "animal" in response to the word DOG, while trying to ignore the pictured swan (semantic), the pictured vest (unrelated), the pictured dog (identical), or the empty rectangle (control).

Previous research (e.g., Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Lupker, 1979; Smith & Magee, 1980) showed that participants are slower in naming a picture with an incongruent word superimposed (e.g., saying "swan" to a pictured swan with the written word DOG superimposed) than in naming the picture with a series of Xs superimposed in the control condition. Furthermore, participants are faster than control when picture and word agree in the congruent condition (e.g., saying "swan" to a pictured swan with the identical word SWAN superimposed). This corresponds to what is obtained with the colour-word Stroop task (e.g., MacLeod, 1991), where speakers are slower in naming a colour patch with an incongruent colour word superimposed (e.g., the word RED superimposed onto a blue patch of colour) than in naming a colour patch with a row of neutral Xs, and speakers are fastest when colour patch and word are congruent (e.g., the word BLUE superimposed onto a blue colour patch). In addition, a semantic effect is obtained with picture-word stimuli. Participants are slower in naming a picture (e.g., saying "swan" to a pictured swan) with a semantically related word superimposed (e.g., DOG) than with an unrelated word (e.g., VEST). When the task is to read aloud the words and to ignore the pictures, there is no interference from incongruent pictures or facilitation from congruent pictures relative to control, and there is also no semantic effect (e.g., Glaser & Düngelhoff, 1984). However, when the words have to be categorized (e.g., saying "animal" to the word DOG), the response latencies with picture distractors are shorter in the semantic than in the unrelated condition, whereas the latencies in the semantic, identical, and control conditions do not differ (e.g., Glaser & Düngelhoff, 1984). The semantic, incongruency, and congruency effects were computationally modelled with WEAVER++ by Roelofs (Levelt et al., 1999, 1992, 2003).

The results for the vocal response latencies in our eye-tracking experiments replicated earlier findings in the literature (e.g., Glaser & Düngelhoff, 1984; Lupker, 1979; Roelofs, 2003; Smith & Magee, 1980). The gaze durations generally paralleled the vocal response latencies, but not always. This indicates that the signal to move the eyes is not simply the completion of the planning of the vocal response. The manual response latencies always paralleled the gaze durations, suggesting that the gaze shifts indexed attention shifts in the experiments.

Picture naming was slowed by semantically related compared to unrelated word distractors, whereas naming was facilitated by identical relative to control distractors. The gaze durations and the manual response latencies closely followed this pattern. These results suggest that the response planning latency is a major determinant of gaze shift, unlike what has been previously suggested (D. E. Meyer & Kieras, 1997; Sanders, 1998). Why do the results implicate response planning (i.e., word planning including phonological encoding), as opposed to lemma retrieval or any other stage? It is generally assumed that Stroop-like effects reflect response selection processes (e.g., MacLeod, 1991; Roelofs, 2003). Thus, given that the distractor effects on the naming latencies were reflected in the gaze durations for the pictures, gaze shifts must have been triggered after response selection (i.e., lemma retrieval in WEAVER++). Moreover, earlier research (e.g., A. S. Meyer et al., 2003) and (yet) unpublished picture-word interference experiments with written distractor words from our own laboratory showed that phonological effects in picture naming are reflected in the gaze durations. This suggests that gaze shifts are dependent on the completion of phonological encoding. To conclude, the results suggest that response-planning latencies are a major determinant of gaze shifts in picture naming.

In reading aloud the words while ignoring the pictures, participants moved their eyes from words with rectangles around them sooner than they moved their eyes from words with actual pictures around them, even though the vocal response latencies did not differ among conditions (see upper panel of Figure 6.3). This suggests an independent role for lower-level stimulus-related factors in determining the gaze shift. The data suggest that the actual pictures held the attention of the participants longer (in spite of their task irrelevance) than the rectangles, and so speakers could move their eyes away sooner with the rectangles. To conclude, Stroop-like stimuli appear to have attentional effects on reading that are reflected in the eye fixation durations but not in the vocal reading latencies.

In categorizing the word while ignoring the pictures (e.g., saying "animal" to the word DOG), vocal responding was slowed by unrelated picture distractors relative to the other distractors, and the gaze durations followed this pattern. Relative to the control condition, the effect of response-incongruent distractors (e.g., the unrelated picture of a vest activating the incongruent response "clothing") and congruent distractors (e.g., the picture of a dog activating the response "animal") was now the same for the vocal
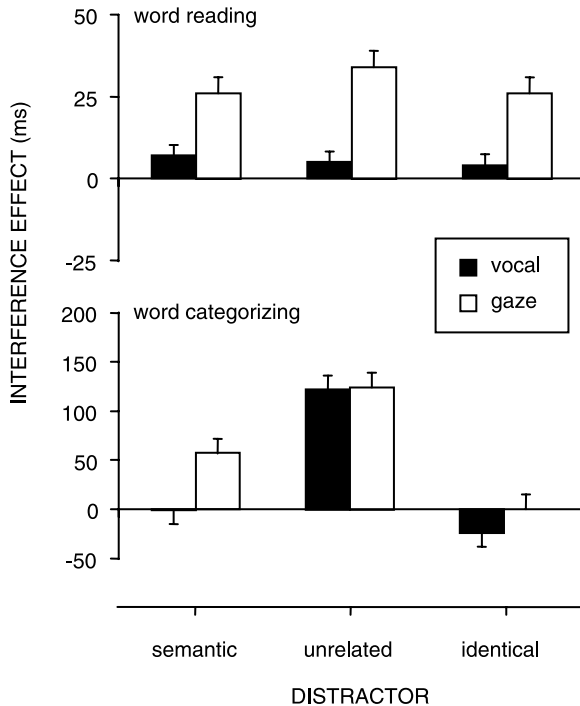
Figure 6.3 Mean effect sizes of the semantic, unrelated, and identical distractors relative to the control stimuli for the vocal responses and gaze shifts in word reading (upper panel) and word categorizing (lower panel) obtained by Roelofs (submitted). The error bars indicate the standard error of the mean.

responses and the gaze durations. This is different from what was observed for reading the words. It seems that when the attentional demands of the task are high, as with word categorizing compared with word reading, the difference in effect between actual pictures and rectangles in determining the shift of gaze disappears. However, with word categorizing, the difference in effect between semantically related and unrelated pictures was much smaller for the gaze shifts than for the vocal responses (see the lower panel of Figure 6.3). Participants fixated the picture-word stimuli in the semantic condition much longer than to be expected on the basis of the vocal categorizing latencies. The difference in semantic effect between the gaze shifts and vocal responses again suggests that the signal to move the eyes is not simply a flag that the planning of the vocal response is completed. Rather, the dissociation implies an independent role for semantic factors. The data suggest that the picture-word stimuli held the attention of the participants longer in the semantic than in the unrelated condition, and so speakers could move their eyes away only later in the semantic condition. To conclude, Stroop-like

stimuli appear to have attentional effects on word categorizing that are reflected in the eye fixation durations, but not in the vocal categorizing latencies.

The findings from the experiments support the following model of the orienting of attention and gaze in Stroop-like tasks (Figure 6.4). The model assumes that gaze shifts and attention shifts are closely related in Stroop-like tasks. An attention shift issues a gaze shift command to the saccadic gaze control system, which executes the saccade (e.g., Van Opstal, 2002). The critical assumption for explaining the described findings is that the decision to shift attention is based on input from both perception/conceptualizing and response planning, whereby the former biases only for an attention shift (i.e., lowers or heightens the threshold) and the latter actually triggers the shift (i.e., leads to an actual exceeding of the threshold). The findings suggest that a rectangle lowers the threshold for an attention shift relative to actual pictures in reading words. Consequently, gaze durations are shorter for
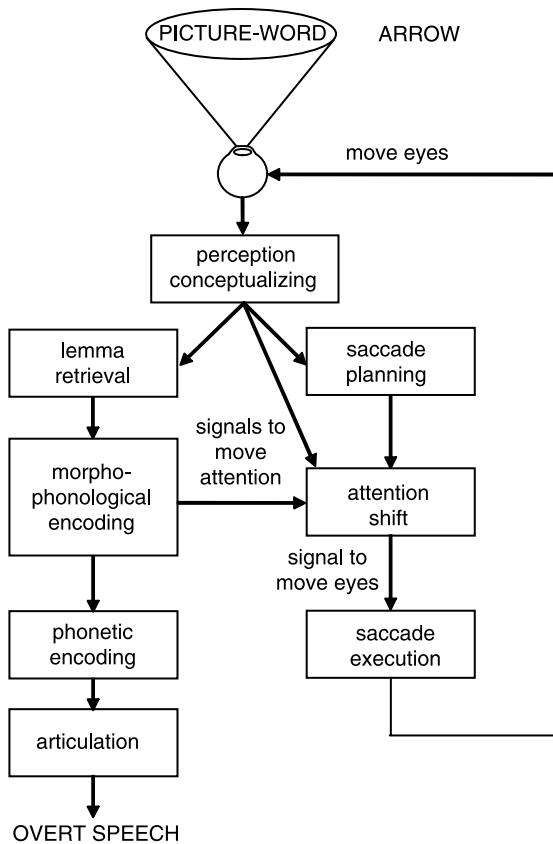

Figure 6.4 A model for the orienting of attention and gaze in the picture-word inter-ference paradigm.

rectangles than actual pictures, even when the word-reading latencies do not differ. The findings also suggest that semantically related pictures as distractors may heighten the threshold relative to other pictures and rectangles in word categorizing. Consequently, semantically related pictures are looked at longer than is to be expected on the basis of the vocal categorizing latencies. Therefore, the difference in effect between semantically related and unrelated pictures is smaller for the gaze shifts than for the vocal responses. Moreover, gaze durations are longer for semantically related than identical pictures and rectangles, even though the word-categorizing latencies do not differ.

To conclude, our findings suggest that the time it takes to prepare a vocal response is a major determinant of gaze shifts in Stroop-like tasks. However, the finding that the distractor effects on the gaze shifts and the vocal responding may dissociate suggests that the signal to move the eyes is not simply the completion of the planning of the vocal response. Instead, the shift of gaze is independently determined by stimulus-related factors. In all experiments, the distractor effects on the gaze shifts were propagated into the manual responses, indicating that gaze shifts index attention shifts in Stroop-like tasks.

Selective stimulus processing

In visual orienting, preattentive processes provide a map of spatial locations through which attention moves (i.e., from the location of the picture-word stimulus to the location of the arrow). In naming one of several objects in a visual scene or in reading aloud one of several words in a text, visual orienting may help to separate the target in one location from irrelevant distractors in other locations. However, selecting a spatial location does not separate the target and distractor in the Stroop task, because the word and colour are spatially integrated in a classic Stroop stimulus. Object-based attention does not help either. In object-based attention, preattentive processes segment the visual scene into figures and ground, whereby attention selects one of the figures for further processing (e.g., Kanwisher & Wojciulik, 2000). Object-based attention does not help to separate the word from the colour word in the Stroop task, because evidence for this type of attention suggests activation enhancement for all information within the spatial boundaries of the object. Rather than separating the word from the colour, space-based and object-based attention lead to a "Trojan horse effect" in the Stroop task (cf. Kahneman & Henik, 1981; Neumann, 1986). According to legend, the Greeks won the Trojan War by hiding in a huge, hollow, wooden horse to sneak into the fortified city of Troy. By attending the ink colour of a Stroop colour-word stimulus, the colour word also receives attention and sneaks into the word-production system. This is also evident from manipulations that increase the spatial distance between a colour patch and colour word and that make the colour and word part of different objects.

It has been shown that increasing the spatial distance between a colour patch and colour word reduces Stroop interference. Gatti and Egeth (1978) increased the spatial distance between colour and word from 1° to 5° and observed that interference diminished from 90 to 40 ms. Merikle and Gorewich (1979) observed no decrease of Stroop interference with increasing distance from 0.5° to 2.5° when letter size was increased to compensate for acuity loss. Brown, Gore, and Carr (2002) presented colour patches and words at different locations. Before stimulus onset, the location of either the colour patch or the word was cued. Manipulation of the locus of attention modulated the magnitude of Stroop interference. Stroop interference was larger when the locations of the colour and word were close together than when they were far apart (e.g., 13°) and attention was drawn to the location of the colour patch. To conclude, the manipulations of the spatial distance between colour and word show that increasing the distance reduces but does not eliminate Stroop interference. This supports the idea that the spatial integration of the colour and word in the classic Stroop stimulus leads to a Trojan horse effect.

It has been observed that making the colour and the word part of different objects also reduces Stroop interference. Kahneman and Henik (1981) presented colour-word Stroop stimuli and coloured colour-neutral words in a circle and square that were about 9° apart on the screen. The task was to name the colour of the word in the circle, which could be the colour word or the neutral word. Stroop interference was much larger when the colour-word Stroop stimulus was part of the relevant object (the circle) than when it was not. Van der Heijden, Hagenaar, and Bloem (1984) replicated this result with a much smaller spatial distance between the objects (about 1°). In the experiments of Kahneman and Henik (1981) and Van der Heijden et al. (1984), object membership and spatial distance are confounded. Wühr and Waszak (2003) controlled for spatial distance by having participants name the colour of one of two partly overlapping objects and ignore colour words that appeared in the relevant object, in the irrelevant object, or in the background. Words produced much larger Stroop interference in the relevant object than in the irrelevant object or background. Irrelevant objects and background did not differ in the magnitude of interference produced. To conclude, the manipulations of object membership suggest that Stroop interference is greater when the colour and incongruent word are part of the same object than when they are not. This supports the idea that the integration of the colour and word into one object in the classic Stroop stimulus leads to a Trojan horse effect.

Space-based and object-based attention lead to the selection of a location or object that contains the relevant information for performing the task, the colour-word stimulus in the Stroop task. Next, attention needs to be engaged to the task-relevant stimulus dimension (the colour) and disengaged from the irrelevant dimension (the word). In WEAVER++, this is achieved by a condition-action rule that blocks out the irrelevant input depending on the

task. Consequently, the system receives perceptual input much longer for the relevant than for the irrelevant stimulus dimension.

Whereas the onset of an event (e.g., a flash of light or the appearance of an object or word) often attracts attention, the offset of a significant event may lead to a disengagement of attention. La Heij, Van der Heijden, and Plooij (2001) reported the paradoxical finding that an early removal of the Stroop stimulus from the screen reduces rather than increases Stroop interference. The interference was larger when the colour-word Stroop stimulus stayed on the screen until the response was made (the "continuous" condition) than when the stimulus was presented only briefly (the "short" condition) or the colour was replaced by neutral white colour after 120 ms (the "replaced" condition). This suggests that an early offset helped to disengage the space-based or object-based attention to the Stroop stimulus, thereby reducing Stroop interference. Interestingly, when the colour was replaced by neutral white after 120 ms while the word stayed on the screen until the response (the replaced condition), the Stroop interference was even less than when both colour and word were presented briefly (the short condition). This suggests that a temporal segregation of colour and word helps to disengage attention from the irrelevant dimension.

In order to provide some form of test of this theoretical analysis ("a proof of principle"), we tested the effect of early stimulus offset in computer simulations with the WEAVER++ model. The simulations showed that the model correctly produced the paradoxical result. Reducing the duration of both colour and word input in the model reduced the magnitude of Stroop interference, as shown in Figure 6.5. The simulation of the continuous condition followed Roelofs (2003). Colour input was provided to the lexical network until response, whereas word input was given for 100 ms. The processes of colour and word perception were not included in the simulation. The duration of these processes is estimated to be around 100–125 ms (Roelofs, 2003). This means that attention disengaged around 200–225 ms after stimulus onset in the model. This corresponds to estimates of attention to colour between 150 and 350 ms after stimulus onset (e.g., Hillyard, Mangun, Woldorff, & Luck, 1995). In the replaced condition, the duration of both the colour and word input were set at 75 ms. This implies that attention in this condition disengaged early, namely, around 175–200 ms after stimulus onset.

Neural correlates of executive attention

Neuroimaging studies have revealed that colour-word Stroop performance engages the anterior cingulate and dorsolateral prefrontal cortices for attentional control, the left lingual gyrus for colour processing, the left extrastriate cortex for visual word-form processing, and the left-perisylvian language areas, including the areas of Broca (posterior inferior frontal) and Wernicke (posterior superior temporal), for word planning (see Roelofs & Hagoort,
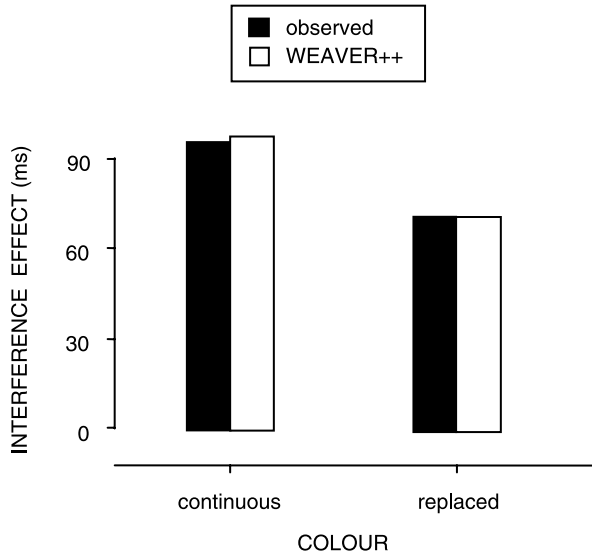
Figure 6.5 The paradoxical colour duration effect in the Stroop task: Real data (means across experiments; La Heij, et al., 2001) and WEAVER++ simulation results.

2002, for a review). Much evidence suggests that the dorsolateral prefrontal cortex serves to maintain the goals in working memory (e.g., Miller, 2000). Anterior cingulate cortex (ACC) involvement in attentional control agrees with the idea that attention is the principal link between cognition and motivation. For action control, it is not enough to have goals in working memory, but one should also be motivated to attain them. Extensive projections from the thalamus and brainstem nuclei to the anterior cingulate suggest a role for drive and arousal. Extensive reciprocal connections between the anterior cingulate and dorsolateral prefrontal cortex suggest a role for working memory. The motor areas of the cingulate sulcus densely project to the spinal cord and motor cortex, which suggests a role of the anterior cingulate in motor control. Paus (2001) argued that the cingulate motor areas contain subregions controlling vocal responses, manual responses, and eye movements. However, based on a meta-analysis of the existing literature and the results from a new neuroimaging experiment, Barch, Braver, Akbudak, Conturo, Ollinger, and Snyder (2001) argued that response conflict activates the rostral cingulate zone regardless of response modality (spoken, manual).

Previous neuroimaging studies showed that the presence of conflicting response alternatives increases ACC activity, indicating that the ACC is involved in attentional control. However, the exact nature of the ACC function is still under debate. The prevailing conflict detection hypothesis maintains that the ACC is involved in performance monitoring (Botvinick et al., 2001). According to this view, ACC activity reflects the detection of response

conflict and acts as a signal that engages regulative processes subserved by lateral prefrontal brain regions. According to an alternative view, the ACC plays a role in regulation itself. For example, the ACC may be involved in the top-down regulation of response selection processes (e.g., Holroyd & Coles, 2002; Posner & DiGirolamo, 1998, 2000; Posner & Raichle, 1994). Roelofs and Hagoort (2002) implemented a version of the regulative hypothesis about ACC function in WEAVER++. They showed that the model did a good job in fitting data on ACC activity derived during Stroop task performance.

It should be noted that assumptions about neural correlates (e.g., ACC function) are not a necessary part of functional models like GRAIN and WEAVER++. For example, it is not critical to the operation of WEAVER++ that the ACC be involved in regulation, and not in conflict detection. In order to apply a functional model like WEAVER++ to neuro-imaging findings, new assumptions need to be made about how functional aspects of the model map onto aspects of brain functioning. Thus, models like WEAVER++ may be used to formalize hypotheses about brain func-tion, but assumptions about neural correlates are not a necessary part of the model (see Roelofs, 2005, for an extensive discussion).

To test between the conflict-detection and regulative hypotheses about ACC function, Roelofs, Van Turennout, and Coles (submitted) conducted a fMRI experiment. A critical prediction made by the conflict-detection hypothesis is that ACC activity should be increased only when conflicting response alternatives are present (e.g., the word RED in blue ink). ACC activity should not differ between congruent trials (e.g., the word BLUE in blue) and neutral trials (e.g., XXX in blue), because competing response alternatives are absent on both trial types. In contrast, the regulative hypoth-esis predicts not only more ACC activity on incongruent than on neutral trials, but also less ACC activity on congruent than on neutral trials. More ACC activity is predicted for incongruent than for neutral trials, because more top-down regulation is required for incongruent than for congruent stimuli. Less ACC activity is predicted for congruent than for neutral trials, because the correct response is already activated by the distractor on congruent trials and therefore less regulation is required.

Participants were scanned while performing a manual arrow-word version of the Stroop task (e.g., Baldo, Shimamura, & Prinzmetal, 1998), in which they were presented with arrow-word combinations. The arrows were point-ing to the left or right, and the words were LEFT or RIGHT. The participants indicated by a left or right button press the direction denoted by the word (word task) or arrow (arrow task). Trials were blocked by task. On incongru-ent trials, the word and the arrow designated opposite responses. On con-gruent trials, the word and arrow designated the same response. On neutral trials in the word task, a word was presented in combination with a straight line, so only one response was designated by the stimulus. On neutral trials in the arrow task, an arrow was presented in combination with a row of Xs;

therefore, on these trials also only one response was designated by the stimulus. Congruent, incongruent, and neutral trials were presented rapidly, in a randomly intermixed order to prevent participants from anticipating and changing strategies for the different event types.

Reaction-time data showed that, consistent with earlier findings, responses to words were much slower on incongruent than on neutral trials, and fastest on congruent trials. Responses to arrows were only slightly slower on incongruent than on neutral and congruent trials, while no difference between neutral and congruent trials was obtained. FMRI data demonstrated that activity in the ACC was larger on incongruent than on congruent trials when participants responded to the words. Importantly, ACC responses were larger for neutral than for congruent stimuli, in the absence of response conflict. This result demonstrates that the ACC plays a role in regulation itself. WEAVER++ simulations, instantiating a version of the regulative hypothesis, showed that the model accounted for the empirical findings concerning response latencies and ACC activity.

Summary and conclusions

We discussed the issue of how visual orienting, selective stimulus processing, and vocal response planning are related in Stroop-like tasks. The evidence suggests that visual orienting is dependent on both visual processing and verbal response planning. We also discussed the issue of selective perceptual processing in Stroop-like tasks. The evidence suggests that space-based and object-based attention lead to a Trojan horse effect in the classic Stroop task, which can be moderated by increasing the spatial distance between colour and word and by making colour and word part of different objects. Reducing the presentation duration of the colour-word stimulus or the duration of either the colour or word dimension reduces Stroop interference. This paradoxical finding was correctly simulated by the WEAVER++ model. Finally, we discussed evidence on the neural correlates of executive attention, particularly the ACC. The evidence suggests that the ACC plays a role in regulation itself rather than only signalling the need for regulation.

Acknowledgements

References

Ach, N. K. (1905). *Über die Willenstätigkeit und das Denken* ["On the activity of the will and on thinking"]. Göttingen: Vandenhoeck & Ruprecht.

Baldo, J. V., Shimamura, A. P., & Prinzmetal, W. (1998). Mapping symbols to response modalities: Interference effects on Stroop-like tasks. *Perception and Psychophysics*, *60*, 427–437.

Barch, D. M., Braver, T. S., Akbudak, E., Conturo, T., Ollinger, J., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of response modality and processing domain. *Cerebral Cortex*, *11*, 837–848.

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

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition "automatic"? *Journal of Experimental Psychology: General*, *131*, 220–240.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: Norton.

Gatti, S. V., & Egeth, H. E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society*, *11*, 181–184.

Glaser, W. R., & Düngelhoff, 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.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1–B14.

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.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.

Humphrey, G. (1951). *Thinking: An introduction to its experimental psychology*. London: Methuen.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Neuroscience Reviews*, *1*, 91–100.

Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*, 74–77.

La Heij, W., Van der Heijden, A. H. C., & Plooij, P. (2001). A paradoxical exposure-duration effect in the Stroop task: Temporal segregation between stimulus attributes facilitates selection. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 622–632.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, *28*, 103–174.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition*, *7*, 485–495.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203.

Mandler, J. M., & Mandler, G. (Eds.) (1964). *Thinking: From association to Gestalt*. New York: Wiley.

Merikle, P. M., & Gorewich, N. J. (1979). Spatial selectivity in vision: Field size depends upon noise size. *Bulletin of the Psychonomic Society*, *14*, 343–346.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance. I. Basic mechanisms. *Psychological Review*, *104*, 3–65.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*, 131–147.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects. *Cognition*, *66*, B25–B33.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, *1*, 59–65.

Müller, G. E. (1913). Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufs ["On the analysis of memory and the flow of images"]. *Zeitschrift für Psychologie*, *8*, 475–489.

Neumann, O. (1986). *How automatic is Stroop interference?* (Rep. No. 109/1986). Bielefeld, Germany: University of Bielefeld.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwarts, & D. Shapiro (Eds.), *Consciousness and self-regulation. Advances in research and theory, Vol. 4* (pp. 1–18). New York: Plenum Press.

Paus, T. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews Neuroscience*, *2*, 417–424.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Posner, M. I. (1994). Attention: The mechanisms of consciousness. *PNAS*, *91*, 7398–7403.

Posner, M. I., & DiGirolama, G. J. (1998). Executive attention: Conflict, target detection, and cognitive control. In R. Parasuraman (Ed.), *The attentive brain* (pp. 401–423). Cambridge, MA: MIT Press.

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., & Peterson, S. E. (1990). The attentional system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. New York: W. H. Freeman.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125.

Roelofs, A. (2005). From Popper to Lakatos: A case for cumulative computational modeling. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 313–330). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Roelofs, A. (submitted). Saccadic eye movements and the attentional control of spoken word production.

Roelofs, A., & Hagoort, P. (2002). Control of language use: Cognitive modeling of the hemodynamics of Stroop task performance. *Cognitive Brain Research*, *15*, 85–97.

Roelofs, A., Van Turennout, M., & Coles, M. G. H. (submitted). *Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks*.

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. Hove, UK: Lawrence Erlbaum Associates, Ltd.

Selz, O. (1913). *Über die Gesetze des geordneten Denkverlaufs* ["On the laws of directed thinking"]. Stuttgart: Spemann.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General*, *109*, 373–392.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Van der Heijden, A. H., Hagenaar, R., & Bloem, W. (1984). Two stages in postcategorical filtering and selection. *Memory and Cognition*, *12*, 458–469.

Van Opstal, A. J. (2002). The gaze control system. In L. J. Van Hemmen, J. D. Cowan, & E. Domany (Eds.), *Models of neural networks, Vol. IV: Early vision and attention* (pp. 47–95). Heidelberg: Springer Verlag.

Watt, H. J. (1905). Experimental contribution to a theory of thinking. *Journal of Anatomy and Physiology*, *40*, 257–266.

Wühr, P., & Waszak, F. (2003). Object-based attentional selection can modulate the Stroop effect. *Memory and Cognition*, *31*, 983–994.