2

THE STROOP TASK IN COGNITIVE RESEARCH

COLIN M. MACLEOD

The Stroop task is the one of the best known paradigms in cognitive psychology; it is also among cognitive psychology's most used contributions to clinical psychology. It is really quite extraordinary that a paradigm of such apparent simplicity has flourished as the discipline has grown and become more sophisticated. In fact, the Stroop task can trace its lineage directly to the origins of experimental psychology. In his dissertation, done under Wundt's supervision, Cattell (1886) made a fundamental observation. He measured response time to name a variety of stimuli and noted that it took longer to name objects (such as a picture of a horse) and properties of objects (such as their color) than it did to read the corresponding words. His explanation—that reading words was much more practiced than naming pictures or colors—introduced the concept of automaticity to psychology, a concept that has remained highly influential (see, e.g., Logan, 1988, 2002; Posner & Snyder, 1975; Shiffrin & Schneider, 1977).

My research cited in this chapter has been continuously supported by a discovery grant from the Natural Sciences and Engineering Research Council of Canada. I appreciate the helpful comments provided on the initial draft of this chapter by Uri Bibi, Ramona Bobocel, and Michael Dodd.

Almost 50 years later, working on his dissertation, John Ridley Stroop[1] (1935) combined word and color, creating a conflict situation. His first experiment contrasted reading color words in normal black ink on a white background (e.g., *green* in black, say "green": the control condition) with reading words in incongruent colors, ignoring the colors (e.g., *green* in red, say "green": the experimental condition). Stroop found little difference in oral word reading time, indicating that the incompatible ink colors were easily ignored during reading. It was his Experiment 2 that introduced the task that now goes by his name. The contrast here was between naming the colors of colored rectangles (e.g., for the green rectangle, say "green": the control condition) and naming the colors of incongruent color-word combinations (e.g., *red* in green, say "green": the experimental condition). Color naming was dramatically slower for the incompatible color words, and the Stroop effect was born.

Stroop accepted Cattell's differential practice account to explain the asymmetrical interference pattern between Experiments 1 and 2, augmenting it with the idea of his dissertation supervisor, Peterson (cf. Peterson, Lanier, & Walker, 1925), that words lead to a single reading response whereas colors lead to multiple reading responses. Stroop later put this idea to further experimental test (Stroop, 1938) and again concluded that it provided an adequate explanation. It is perhaps surprising that his task did not immediately catch on; for the next 30 years, use was limited, typically as a psychometric marker of attention (e.g., Thurstone, 1944).

A study by Klein (1964) opened the floodgates. Klein reasoned that the type of word that had to be ignored while color naming could be critical, suggesting that the word's meaning and its relation to the response of color naming were important in determining the size of the effect. To test this idea, he compared naming the colors (say, blue) of rows of asterisks (*****), nonwords (*evgjc*), low-frequency words (*helot*), high-frequency words (*heart*), color-related words (*grass*), color words not among the ink colors to be named (*black*), and incompatible color words in the response set (*green*). He found that interference increased across these conditions as the reading response to the item became stronger and more color-related. This study unleashed the potential of the Stroop effect as a research tool. Bibi and MacLeod (2004) have replicated the results of the classic Stroop and Klein studies using modern computerized presentation, providing contemporary benchmarks for these procedures.

Two more studies deserve places at the head table for their methodological innovations. Tecce and Dimartino (1965) introduced a tachisto-

---

[1] For a biographical sketch of John Ridley Stroop, see MacLeod (1991b). My Web site (http://www.arts.uwaterloo.ca/~cmacleod/Research/Stroopbiog.htm) contains an electronic version of this sketch and further pictures. Copies of my articles since 1991 can also be downloaded as PDF files from http://www.arts.uwaterloo.ca/~cmacleod/Research/Publications2.htm.

scopic version of the task. Instead of requiring the naming of the colors of multiple items on a single card, as Stroop (1935) and Klein (1964) had done, this new version presented single color-word trials for individual color-naming responses. To investigate this single item procedure more thoroughly, Dalrymple-Alford and Budayr (1966) introduced a new condition: They included congruent trials, in which the color and the word were compatible (e.g., *red* in red, say "red" to the color). Later studies demonstrated that this congruent condition often (but not always) resulted in faster responding than did the control condition—this effect is called Stroop facilitation—although the facilitation was virtually always much less than the interference (see pp. 174–175 in MacLeod, 1991a).[2] By this point, then, the standard ingredients were all in place, and the Stroop literature was just beginning to burgeon.

A number of review articles have since been published. Jensen and Rohwer (1966) reviewed the largely clinical and psychometric literature up to that point, emphasizing the utility of the task as a measure of attention in applied and clinical settings. Dyer (1973b) reviewed the early cognitive literature. MacLeod (1991a) provided an extensive review of the cognitive literature over the first 50-plus years of the task's existence. In 2000, MacLeod and MacDonald provided a sketch of the cognitive and especially the recent cognitive neuroscience research involving the Stroop task. All of these reviews are useful sources for pointers to and details about specific experimental questions that have been addressed.

In this chapter, my goal is to cover how Stroop experiments have been done in the past and how they are done now, with consideration of the many factors involved in conducting these experiments and in interpreting the data that they produce. I begin with the original paradigm and then move to contemporary procedures, along the way attempting to highlight issues especially relevant to the clinical researcher who uses variants of this task.

THE ORIGINAL MULTIPLE-ITEM CARD VERSION

In Stroop's (1935) original procedure, he used cards containing 100 items, arranged in 10 rows and 10 columns. In his Experiment 2, in which the task was color naming, the control card had 20 rectangles in each of five colors: red, blue, green, brown, and purple. He ensured that no color appeared more than twice in any row or column. For the incongruent card, he attempted the same arrangement, stating that no word or color was presented more than twice in any row or column. As it turns out, this is impossible:

---

[2] All references to "MacLeod" are to the work of Colin M. MacLeod, author of this chapter. All references to "C. MacLeod" refer to the work of Colin MacLeod, author of chapter 3.

Inspection of his original cards, on display at Vanderbilt University, confirms that he applied the constraint to the rows but not the columns. Each participant was given two cards in each condition, the second having the items in the reverse order from the first. Short practice sequences (10 items/card) prepared the participant for each condition. Although we do not know why Stroop did not use congruent cards, my guess is that he realized that, when the word and color agreed, participants would notice and switch to reading. As a result, the congruent condition would not necessarily involve the intended color-naming response and consequently would not be comparable to the other two conditions.

Stroop's instructions to participants were to respond as quickly as possible, leaving no error uncorrected. He started a stopwatch with the participant's first response and stopped it with his or her last response, obtaining one total time in seconds per card. The issue of how to handle errors is a significant one because errors are common, especially on the difficult incongruent cards. Requiring participants to correct errors results in two responses to those items. Thus, researchers often do not require error correction online which would ensure precisely 100 responses to all cards. Of course, more responses are errors on the incongruent card than on other cards, so there is a discrepancy across conditions regardless of the instructions concerning errors. Jensen (1965) discusses how to score the card version of the Stroop task.

The standard way to report data in the card version of the Stroop task is to sum the times for the (two) cards in a condition and then average them to provide a condition mean. The difference between the condition means on the control card and on the incongruent cards is taken to be a measure of interference, termed the *Stroop effect* or *Stroop interference.* The original data are highly replicable (cf. MacLeod, 1991a, pp. 164–165) in terms of both pattern and, remarkably, absolute times. In the card version, for with my data, it takes about 60 seconds to name the 100 control items and 102 seconds to name the 100 incongruent items. The 42-second difference constitutes an increase due to interference of about 70%, a truly huge increment when compared with virtually any other effect frequently investigated in the cognitive literature.

The card version has given way to the computerized version over the past 30 years. The card version does have disadvantages. First, errors are mixed in with correct responses because the single-card response time is a kind of "all in" measure. Second, previous or upcoming stimuli, which are visible simultaneously, can also influence processing of the current stimulus. Third, it is tedious to create multiple randomizations. Fourth, it is difficult to establish reliability, given limited numbers of observations per participant. Of course, the principal advantages of the card version are that it requires no equipment and is very portable.

THE MODERN SINGLE-ITEM COMPUTER VERSION

Today, cognitive psychologists conduct the vast majority of their studies under computer control. As a result, the single-item version of the Stroop task has supplanted the traditional multiple-item card version. Generally, the colored rectangle control used by Stroop has been replaced by a control item made up of keyboard characters (I discuss control items more in depth later in the chapter). In the single-item version, one item appears on each trial and is separately timed from its onset to the participant's response, which results in a time per trial in milliseconds. The single-item version allows error trials to be discarded, which solves one of the problems inherent in the card version of the task. The typical instructions for a cognitive response time task—"Respond as quickly as possible while avoiding errors"—are used despite their inherent ambiguity.

If one were to attempt to abstract a prototypical modern Stroop experiment, I believe that it would look something like this: There would be the standard four colors—red, blue, green, and yellow—and the corresponding words. Each of the 12 incongruent color-word combinations (given four colors) would be presented equally often. There would be at least two to three repetitions of the set, resulting in 24 to 36 trials. The control condition would be treated in the same way, although the precise control string (or strings) actually used varies widely. The trials from the incongruent and control conditions would be intermingled randomly. If a congruent condition was included, these trials would also be mixed in randomly. The congruent condition makes sense in the single-item procedure because, due to the random sequence, the participant cannot guess the condition and therefore cannot switch to a reading strategy on congruent trials. Note, however, that there would be only four congruent items, so they would be repeated more often than would individual incongruent items.

The total number of trials would range from 48 to 288, depending upon the number of conditions and the number of trials within each condition. Ordinarily, each participant would receive a unique random ordering of the trials. Administration of the task would take from 5 to 30 minutes, not counting initial instructions and practice trials. The number of practice trials varies, but 8 to 24 would not be uncommon (unless performance on initial trials is specifically of interest). The practice trials are especially important in the single-item procedure to acclimatize the participant to the display and especially to the response characteristics of the task.

Now consider a single trial. First, a blank screen would appear, perhaps lasting 500 ms and possibly broken into a fixation (e.g., ++++) and then a blank (250 ms each). Next, the color-word item would appear, usually centered on the screen in lower case (which is more like normal reading). The computer timing would be initiated with the item onset. The item would be removed and the timing would stop with the participant's response. Sometimes, the item is displayed for a fixed time (e.g., 150 ms) rather than having its offset contingent on the response. If scoring is done online, the temporal gap before the next trial might be a little variable due to the experimenter having to input a keypress to indicate accuracy. One can overcome this problem by programming a sufficiently long time between trials (e.g., 500 ms or more) to allow the experimenter to input accuracy easily within that interval.

Responding in the single-item procedure is done either vocally or manually (see pp. 182–183 in MacLeod, 1991a). Vocal responding into a microphone/voice key requires special circuitry to interface with the computer, but setting it up is not difficult.[3] When responding is via keypresses, often the index and middle fingers of each hand are used, with one color assigned to each finger. Common keys to use are "z", "x", ".", and "?". It is essential in manual responding to have a practice session to accustom the participant to the assignment of colors to keys. I recommend a minimum of 48 trials that perhaps use colored rows of asterisks (or the like) to conserve the word-based trials for the actual experiment. In light of the robustness of the task in the face of practice, though (see, e.g., MacLeod, 1998), it would do little harm to include actual word-based trials in practice. Keypress responses are typically slower than vocal responses, possibly because participants must translate a covert vocal response to an overt keypress.

Although keypresses are nonnatural responses to colors, with a little practice they work quite well. They have the advantage that no responses are lost due to not speaking loudly enough, or the like. Moreover, an experimenter need not be present during testing because accuracy can be scored by the computer; thus multiple participants can be tested simultaneously. Keypress responses have the disadvantage that the size of the Stroop effect is considerably smaller than with vocal responding (e.g., Redding & Gerjets, 1977), although it is still a large effect in comparison with other cognitive response time measures. It is worth noting that there is continuing debate about what processing differences underlie the response modality difference in effect size (Brown & Besner, 2001; Sharma & McKenna, 1998).

Vocal responding raises the issue of how to handle response errors. Options include tape-recording responses and retrospectively locating and marking errors or, more commonly, scoring responses online, which requires an experimenter to be present to record accuracy following each trial. It is important to realize that experimenters are also vulnerable to the Stroop effect, so they should simply record the participant's response without trying to judge its accuracy online; the computer can then make the necessary comparison. Note that there are at least two types of errors. Most obvious is making the wrong oral response, which usually means reading the word instead of

---

[3] Commercial experimental packages such as e-Prime (http://www.pstnet.com/) can provide both the software and the input hardware at a not-exorbitant cost.

naming the color (although occasionally a participant faced with *green* in red will respond "yellow" or "blue" and then look rather sheepish!). Recording actual responses on error trials, rather than just the fact that an error was made, also preserves the entire record of the participant's responses. The other main type of error is a voice key error, such as speaking too quietly on a trial or coughing or saying "uh" before the response. Participants should be instructed to speak in a quite loud voice and to be careful not to precede their response with any other sound. Such voice key errors are best scored with a different code than are true errors so that they can be differentiated and reported separately.

Reporting of data in the single-item version of the task is straightforward. First, all errors are removed and counted. It is important to report both error proportions and mean correct response times to expose any possible trade-off between speed and accuracy (cf. Pachella, 1974). Fortunately, response time and accuracy are usually positively correlated in the Stroop task, with most errors in the slow incongruent condition, so no trade-off clouds interpretation. The individual response times from each condition are then assembled. Often, investigators define upper and lower criteria to trim extreme response times, known as outliers, out of the data. There are quite a few techniques for this (see, e.g., Miller, 1991; Ratcliff, 1993; Ulrich & Miller, 1994; Van Selst & Jolicoeur, 1994), but the basic goal is to remove anticipations (less than, say, 300 ms) and lapses of attention (more than, say, 1,500 ms). The proportion of outliers removed should also be reported and ordinarily should not exceed 3% to 5%. At this point, the mean response time for each condition can be calculated; these values form the primary data. Trimmed means are much more common than are medians as central tendency measures in the Stroop literature, although both measures should produce the same qualitative data pattern.[4]

Data in the typical single-item Stroop task are rather different from data in the multiple-item version. Estimating the mean response time per item from the card version produces estimates of about 600 ms for the control condition and about 1,020 ms for the incongruent condition (based on the means in the previous section), which suggests 420 ms of interference. In fact, on the basis of data from my laboratory over the years (see, e.g., MacLeod, 1998), mean response times in the single-item version are on the order of 700 ms for the control condition and 820 ms for the incongruent condition, assuming vocal responding.[5] Thus, the interference effect is estimated here

---

[4] Researchers have occasionally reported transformed response time data, using a ratio (e.g., [Incongruent–Neutral]/Neutral), the log, speed, or the like. My position is that, unless there is a powerful, theory-driven reason to transform the data, response times are best left untransformed. In this way, the data represent actual processing time.

[5] A reasonable estimate of the congruent condition mean for the single-item procedure with vocal responding would be on the order of 680 ms, or about a 20-ms facilitation effect, although facilitation can be rather variable and is quite often not significant. Response times in the keypressing version of

at 120 ms. This represents a 17% increase as opposed to a 70% increase from the control to the incongruent condition. Why should the single-item method produce so much less interference? No doubt the ability to look ahead in the multiple-item version speeds overall responding (hence the low control condition estimate) but also permits more interference from nearby items, and the additional problem of recovering from errors slows total response time more in the multiple-item version. In this regard, Salo, Henik, and Robertson (2001) have begun the important job of directly comparing the two test formats.

STROOP VARIANTS AND OTHER INTERFERENCE TASKS

Methods other than vocal and keypress responding have been used to measure interference over the years, though much less frequently. Most prevalent among these is the card-sort method, in which individual color–word stimuli on cards are sorted into bins by color or by word, and sorting time for a deck of cards is the dependent measure (Chmiel, 1984; Tecce & Happ, 1964; see pp. 166–167 in MacLeod, 1991a). This method has been largely abandoned with the rise of computerized methods of conducting cognitive experiments, but still has potential value for settings (or individuals) where computerized testing is not possible.

There are also many other interference tasks, in which different types of materials are used. Most prevalent among these is the picture-word task (Glaser & Düngelhoff, 1984; Hentschel, 1973; see pp. 167–168 in MacLeod, 1991a), in which the participant names a line-drawing picture while ignoring the word typed within it. A version that is seeing more use, notably in the brain imaging literature, is the *counting Stroop task*, in which the participant counts the number of digits displayed, ignoring their identity (Bush et al., 1998; Windes, 1968). And, of course, there is the *emotional Stroop task* (Williams, Mathews, & C. MacLeod, 1996) considered at length in chapter 3. Numerous other interference tasks exist as well, as I have described elsewhere (pp. 168–170 in MacLeod, 1991a). Conceptual cousins are also common in the literature. These include the flanker task (Eriksen & Eriksen, 1974), in which a response-relevant target is surrounded by irrelevant but competing distracters, and the global-local task (Navon, 1977), in which a large letter is constructed from different smaller letters and the individual must attend to one of the sizes, ignoring the other. In my view, researchers are often too willing to refer to any task that involves interference as a Stroop task. The term *Stroop task* should probably be reserved for the color-word

---

the task depend on the amount of practice given with the color-key assignments before the experiment proper begins, so these are harder to estimate. With a little practice, control response times could be around 650 ms and incongruent times could be 700–720 ms, for interference of about 60 ms, half that of the vocal version but still quite a large and robust effect.

interference task, and the more superordinate term *interference task* should be used to designate the class of which the Stroop task is only one member.

THE THREE CONDITIONS

As already discussed, three conditions see frequent use in Stroop experiments: incongruent, congruent, and control. Although these conditions may appear straightforward, complexities lurk just beneath the surface, as so often is the case in experimental psychology. Consider the incongruent condition. How many words and colors should be included? What perceptual characteristics are important to take into consideration?

The typical number of colors and words is from three to five, although other quantities have been used (see p. 177 and pp. 183–184 in MacLeod, 1991a). Using only two colors/words can be problematic because participants can develop special strategies (e.g., in the incongruent condition, instead of naming the color, covertly read the word but respond aloud with the other word). Thus, I recommend always using more than two colors and words. Going beyond five words/colors begins to introduce potential hue discrimination problems, so the range three to five generally seems best. With odd numbers of color responses in the keypress situation, it may be best to have participants respond with one finger to multiple keys, rather than having an imbalanced key distribution across hands. The colors should be chosen to be readily discriminable, on the basis of careful pilot research. It may be useful to consult a set of color norms, such as those of Solso (1971). Klopfer (1996) has shown that interference, whether measured by response time or error rate, increases as the color–word similarity of individual items increases, something the investigator may wish to control.

In the incongruent condition, typically all of the combinations are used equally often, so for a set of four colors and the corresponding words, there are 12 possible incongruent items. Of course, for this same set of items, there are only four possible congruent items, where word and color coincide. It may be important that there are therefore fewer different congruent items than incongruent items; Melara and Mounts (1993) and Sabri, Melara, and Algom (2001) have begun to investigate this, focusing on the important topic of discriminability. Melara and Algom (2003) have in fact developed a theory of Stroop interference based on discriminability. The issue of number of different items may also apply to the control condition, depending on how it is constructed.

The most difficult decisions have to be made in constructing the control condition. Exactly what is to be controlled for, and what does this demand in terms of choosing the to-be-ignored word dimension? It has been quite common to use a string of characters as the control, such as ***** or xxxxx in red. This configuration uses a repeated keyboard character with no word-like properties to convey the color, so there should be no interference with color naming. Repeated character strings are closest to the colored rectangle that Stroop used with the goal of measuring pure (i.e., unidimensional) color naming while avoiding making the control condition too distinctive from the other conditions. Sometimes, a different letter string is constructed to correspond to each of the color words (e.g., www, xxxx, sssss, and mmmmmm, to correspond to red, blue, green, and yellow in length; see MacLeod & Hodder, 1998). Care should be taken not to use letters that are themselves the first letters of color words (even those not in the response set), as Regan (1978) has shown that the first letter of a color word (e.g., *g* in green, say "green") by itself causes substantial interference. Although letter strings do produce a small amount of interference relative to nonletter strings, these can all be treated as relatively pure estimates of color-naming time without verbal distraction.

When possible controls are considered, the next candidate is a to-be-ignored verbal item more like a word. Pronounceable nonwords (e.g., *dral*) cause measurable interference relative to unpronounceable nonwords (e.g., *hbnw*), which behave like letter strings (see Bibi & MacLeod, 2004; Dalrymple-Alford, 1972; Dalrymple-Alford & Azkoul, 1972; Klein, 1964). These might be suitable control items when verbal item pronounceability is of concern. When lexical status is considered crucial, switching to actual (noncolor) words as control items makes sense. There is some question as to whether the frequency of the word in the language is (Burt, 1999, 2002; Klein, 1964) or is not (Bibi & MacLeod, 2004; Monsell, Taylor, & Murphy, 2001) a factor. Although other control items could be and have been suggested (e.g., the color nonwords such as *blat* and *grend* of Besner, Stolz, & Boutilier, 1997, but these are problematic because they use the first few letters of color words), the standard ones are nonletter strings, letter strings, nonwords, and noncolor words.

And then there is the congruent condition (see pp. 174–175 in MacLeod, 1991a). The received view is that congruent trials produce facilitation for the same reason that incongruent trials produce interference, with the smaller effect size (on the order of 20 ms of facilitation vs. the 120 ms of interference) viewed as a consequence of it being more difficult to speed up a response than to slow it down (see MacLeod, 1998). However, MacLeod and MacDonald (2000) briefly describe research indicating that there may not be true facilitation, but rather that on some proportion of congruent trials participants slip and accidentally read the word instead of naming the color. Given the faster reading response (Cattell, 1886), there appears to be facilitation, but it is not true facilitation. This issue is not yet completely worked out, but unless a researcher has a particular reason for including congruent trials, I recommend omitting them. Of course, the congruent condition does not make sense in the context of noncolor-word experiments (e.g., there is no color with which *table* or *spider* is congruent or incongruent).

Sometimes, researchers compare the congruent and incongruent conditions directly, without a neutral control condition. In this case, the congruent condition is essentially treated as the control condition, according to the argument that the words and colors are the same in the two conditions and only congruency is manipulated. This argument has been made most strongly by Sabri et al. (2001) and by Melara and Algom (2003), who maintain that there is an advantage to having the same words in both the incongruent and the congruent/control condition, and that the difference score in this case can be meaningfully interpreted as the total effect of congruency/incongruency.

One more factor should be considered regarding conditions: the randomization of trials and conditions. Although fully randomized trial sequences are often used, some researchers prefer to program their experiments to prevent the occurrence of consecutive trials from the same condition or of the same entire item, or of the same color or word. Sequence effects do occur in the task (see pp. 177–178 in MacLeod, 1991a), and it may be desirable to avoid them. Of course, this rule could be relaxed slightly to allow runs of two or three but not longer, for example. The smaller the set of items, the more this issue becomes salient.

PRIMING IN THE STROOP TASK

In the same article in which they introduced the congruent condition, Dalrymple-Alford and Budayr (1966) also reported a new phenomenon hidden in the trial sequence. Imagine the following sequence of two trials: The first trial is the word *red* in green (respond "green") and the second is the word *yellow* in red (respond "red"). In this situation, the response "red" must be avoided on the first trial and then produced on the second trial. Dalrymple-Alford and Budayr reported an additional slowing of the response on the second trial in such a sequence, relative to sequences without repetition. This delay has come to be called *negative priming* and has been widely studied (see Fox, 1995; May, Kane, & Hasher, 1995, for reviews), although more recently it has usually been studied in contexts outside the Stroop task. It may be desirable under some circumstances to preclude or to systematically manipulate these ignored repetition trials that can be used to measure negative priming. The relation between the amounts of negative priming and Stroop interference observed within the paradigm should also be considered (see, e.g., Mari-Beffa, Estevez, & Danziger, 2000); changes in one can help in interpreting changes in the other.

The more standard priming procedure, so familiar in cognitive psychology, has also been used quite extensively in Stroop studies (see pp. 173–174 in MacLeod, 1991a). In the Stroop case, priming usually involves presenting one or a few words prior to the color-naming trial, varying the relation between the prime and the ignored item in the critical color-naming trial. However, in the priming work (unlike in standard Stroop studies) the words are noncolor words. This work originated with Warren (1972, 1974) and has continued in the studies of Henik, Friedrich, and Kellogg (1983), Whitney (1986; Whitney & Kellas, 1984), Burt (1994, 1999, 2002), MacLeod (1996), and others.

Warren (1972) originally showed greater interference when *aunt, uncle, cousin* preceded either *aunt* in red or *relative* in red (the response being to say "red" to the color) than when three unrelated words or the phrase "no list this trial" preceded the colored target. The intuitive idea that priming the word itself (called *repetition priming* or *identity priming*)—or its semantically related superordinate (called semantic priming)—would make it interfere more with color naming was supported. Warren (1974) obtained a similarly enhanced interference effect when the related word was a semantic associate. This pattern was later extended to the role of semantic context in sentences by Whitney and his colleagues (Whitney, 1986; Whitney & Kellas, 1984; Whitney, McKay, Kellas, & Emerson, 1985).

Burt (1994, 1999, 2002) has questioned whether the straightforward story that priming increases interference is as simple as was first believed. She has shown that identity/repetition priming actually tends to facilitate color naming, making it faster, a pattern that I have also repeatedly observed in unpublished research. Burt has argued that facilitation is, in fact, normal for identical words except when phonological activation of the word increases response competition. In contrast, non-identity priming tends to produce interference, as originally shown by Warren and others. Identity primes may allow the to-be-ignored word on the next color-naming trial to be more easily or more completely ignored, whereas related (but not identical) primes lead the individual to check back to determine the prime-target relation, which absorbs additional time and thereby appears to increase interference.

THE USE OF NONCOLOR WORDS

I noted that priming studies use noncolor words. Thus, these studies are not truly Stroop experiments, but rather use the color-naming task as a way to indirectly measure the activation of words and/or concepts. This use of priming is highly relevant, of course, to the applied domain. In clinical studies, the goal is generally to explore the state of activation of certain concepts relating to an individual's diagnosis or to the investigator's dimension of concern. The innovation in this work is in the type of words used. This innovation goes back to Ray (1979) who investigated test anxiety by using exam-related words as the to-be-ignored dimension. Shortly after Ray's study, other work began to explore anxiety (Mathews & C. MacLeod, 1985), phobias (Watts, McKenna, Sharrock, & Trezise, 1986), eating disorders (Ben-

Tovim, Walker, Fok, & Yap, 1989), and numerous other diagnosed disorders and dimensions of concern (for a review, see Williams et al., 1996). This clinical Stroop research has now become a very large domain of research.

In essence, the studies in the clinical literature that use what is now widely known as the emotional Stroop task also rely on priming. Their goal is to use color naming as an indirect (presumably nonstrategic) measure of the activation of a concept. Here, it is useful to make a distinction between acute (short-term, within-session) and chronic (long-term, extended over time) priming. The cognitive literature has examined acute priming from a just-presented prime that is related to the word whose color is to be named. In the clinical literature, in contrast, activation is assumed to be chronic because of the history and experience of the individual. There may be no within-experiment priming manipulation; instead, the manipulation is one of individual differences. This distinction has several important implications that will be considered in chapter 3; I consider only one here.

When priming is acute, one can either prime or not prime a given word. By way of illustration, the target word *table* in red could be preceded by a neutral prime (*****), an unrelated prime (*horse*), a related prime (*chair*), or an identity prime (*table*). This acute priming procedure permits tight experimental control in that each item serves as its own control, either within or across participants. Moreover, investigators can select as many different words as they need from a very large set in the language, controlling selected item characteristics. In contrast, in the chronic situation, the words must be relevant to the diagnosed disorder or dimension of concern, which limits the investigator to a small set of items that may not be easily balanced with respect to their properties. Furthermore, the concern-related words necessarily are the primed items, so the problem then becomes finding a suitable unprimed set. Words ordinarily cannot serve as their own controls in the clinical setting, except perhaps in the case where both a pretest and a posttest are used to determine whether a treatment has been effective in reducing a problem (cf. Watts et al., 1986). Even in that case, however, it may be best to use two counterbalanced sets, to avoid item repetition.

What should the control items be in the clinical setting? The answer hinges on the composition of the experimental item set. Consider an example in which the concern is eating. The experimental set consists of eating-related words, most likely including foods (*cake*), but also eating-related activities (*lunch*), body-shape-related terms (*obese*), and possibly instances of other eating-related categories. Creation of the control set now becomes complicated. One can begin by selecting items from the Thorndike and Lorge (1944) corpus, perhaps finding a length-matched word for each of the critical items. One can also control for category membership (using the Battig & Montague, 1969, norms), word frequency (using the Kučera & Francis, 1967, and Francis & Kučera, 1982, norms), and perhaps imageability (using the Paivio, Yuille, & Madigan, 1968, norms). Still, it is often difficult to find a suitable set, let alone to match the members of that set to the members of the critical set. Nonetheless, this is a very important aspect of the design of the experiment, given that the amount of interference—or even its presence—is based on the deviation of the experimental set from the control set. It is worth noting here that using a separate control group can help to solve this problem (e.g., spider phobics or normal controls should not show enhanced interference on words related to eating).

OTHER FACTORS IN EXPERIMENTAL DESIGN

Even a procedure as superficially uncomplicated as the Stroop paradigm has its intricacies when one reaches the concrete stage of actually setting up and programming the experiment. In addition to the stimuli and the response format, other elements must be selected and fine-tuned as well. Consider the instructions. Participants avoid reading (or even listening to) instructions, expecting to figure it out on the fly. In my laboratory, after giving the participants the opportunity to read the instructions, we ask them (without forewarning) to summarize the task. After they admit that they did not pay attention to the instructions, the experimenter explains the task and then again asks for a summary. We find that participants pay attention this second time, so the repetition is well worth doing. A few sample trials on large index cards are also useful to demonstrate stimulus format and the desired response. A series of practice trials—eight or more, depending on available time—is valuable as well.

When the participant is responding vocally, we usually follow the response with a brief blank screen and then the word "ready?" The participant is told that this provides a chance to ask for a short break, but in fact it allows the experimenter to input a keypress indicating the participant's response. It is worth noting that the time between a response to one trial and the onset of the next trial—the response-stimulus interval (RSI)—can be influential and should be taken into account. Sharma and McKenna (2001) have presented evidence that time pressure, achieved primarily with a short RSI, can substantially impact the amount of interference observed. Indeed, they found emotional interference only when there was time pressure.

One decision that may have to be made is whether to present the word in color—the integrated version—or to present the color separately from the word—the separated version (see pp. 175–176 in MacLeod, 1991a). Separation would be necessary, for example, if the two dimensions were to be presented simultaneously to different hemispheres (e.g., Dyer, 1973a; see pp. 185–186 in MacLeod, 1991a) or sequentially to investigate the time course of the effect (e.g., Glaser & Glaser, 1982; see pp. 179–180 in MacLeod, 1991a). Interference certainly declines with practice in the separated case (MacLeod, 1998), although the large interference seen in the integrated version appears to diminish even more quickly with practice. I have argued that, with practice, participants may learn to separate the two dimensions of the integrated stimulus.

In the separated case, not surprisingly, interference declines with distance of separation (Gatti & Egeth, 1978). When the first and second dimensions are presented sequentially, the time between their appearance is referred to as the stimulus onset asynchrony (or SOA; e.g., Glaser & Glaser, 1982; see pp. 179–180 in MacLeod, 1991a). Interference maximizes when the two dimensions appear within about 100 ms of each other and falls off sharply outside this window. It is very difficult to obtain a reverse Stroop effect, which occurs when an incongruent color interferes with reading the word, even when the color is given a substantial head start (Glaser & Glaser, 1982). With simultaneous exposure, Dunbar and MacLeod (1984) even showed that interference was almost identical for normally oriented words and for upside-down words, despite the dramatic slowing of word reading when words were upside down. The word exerts a powerful influence on color naming even when intuition—and reading time—might lead one to expect otherwise.

In discussing separation of word and color, it is worth noting that subliminal presentation of the word prior to the color has a checkered history in terms of whether interference occurs and, if so, under what conditions (e.g., Cheesman & Merikle, 1984, 1986; Severance & Dyer, 1973). Part of the difficulty is in defining the criteria for *subliminal*. Because this technique has been popular in the clinical literature (e.g., C. MacLeod & Rutherford, 1992; Mogg, Bradley, Williams, & Matthews, 1993; Mogg, Kentish, & Bradley, 1993)—it provides another way to reduce intentional or conscious processing—this important consideration is given more attention in Williams et al. (1996) and in chapter 3.

There are two additional somewhat related factors to weigh in constructing the trials in a Stroop experiment. First, it is well established that the probability of various trial types can affect the amount of interference observed (see pp. 176–177 in MacLeod, 1991a). Increasing the proportion of congruent trials (Lowe & Mitterer, 1982) or of neutral/control trials (Tzelgov, Henik, & Berger, 1992) leads to increased interference on incongruent trials, but interestingly does not appear to affect facilitation on congruent trials. It is most common to include equal proportions of trial types, but in some situations, deviation from this should be carefully considered, for both applied and theoretical reasons (see the argument made by Melara & Algom, 2003).

The second trial composition factor is repetition. In any Stroop-like experiment, the to-be-named colors are presented repeatedly across trials, given the limited set. In the traditional Stroop experiment, the words are also repeated because they are the corresponding color words. This repetition may actually heighten interference, given that the words are eligible responses and are being frequently presented across trials, so priming builds up. But what should be done when the words are not color words, as in studies of acute or chronic priming? This can be a critical question in clinical studies, in which the set of relevant words may be quite finite. My recommendation is to block repetitions, so that all words in the set are presented before any word is presented again. This blocking also permits analysis of changes in interference due to repetition across blocks, which may help to provide additional insight.

INDIVIDUAL DIFFERENCES IN INTERFERENCE

There have been fairly extensive studies of how certain individual differences affect Stroop interference. In my review (MacLeod, 1991a, pp. 184–187), I singled out three prominent differences: sex, age, and language. I concluded that sex differences were negligible; I have not seen evidence that would lead me to revise that conclusion. With regard to developmental trend, interference emerges with the onset of reading, rises in the early grades of school as reading skill develops, and then very slowly decreases over the adult years (Comalli, Wapner, & Werner, 1962). Comalli et al. also suggested that interference then rises again after age 60, but a meta-analysis by Verhaeghen and De Meersman (1998) suggests that this effect may result from overall cognitive slowing, not increased interference. As for the case of bilingualism, interference is almost as strong between languages as within language, although the dominant language tends to produce more interference (Chen & Ho, 1986; Dyer, 1971; Mägiste, 1984). I also included a subsection in my review article on a within-individual parameter—hemispheric differences; I concluded that the left hemisphere is subject to more interference, not surprising given its central role in language (MacLeod, 1991a, pp. 185–186). More clinically relevant individual differences were considered by Williams et al. (1996) and will be examined in chapter 3.

THEORY: WHAT CAUSES INTERFERENCE?

At the end of the day, what we are always aiming to do in cognitive experiments is to isolate and characterize the processes that underlie our routine information processing. In clinical studies, the goal is to characterize the disruption that occurs when processes are distorted, are replaced, or become altogether unavailable. We wish to understand the stream of processing that permits us to perform basic cognitive functions such as attending and remembering. So it is reasonable to ask: What does Stroop interference measure?

At the outset of this chapter, I mentioned several early accounts of interference. Cattell (1886) theorized that stimuli differ in the degree to which their processing has been practiced, with more practiced dimensions having greater automaticity. From this premise, one can derive the prediction that whenever the processing of one dimension is more practiced than that of another dimension, the more practiced dimension will be more likely to interfere with the less practiced one, and interference will be asymmetric. This is precisely what Stroop (1935) observed, and he adopted a differential practice account that was closely related to Cattell's theory. Efforts to directly control practice have also supported the automaticity account, as in the training experiments of MacLeod and Dunbar (1988). Using newly learned and practiced arbitrary shape-to-color-name connections, we observed a shift in interference over practice, such that the more practiced dimension did interfere more with the less practiced dimension. Automaticity explains these results and is readily extended to explain priming effects.

Stroop (1935) added the idea that greater practice led to less variance in the set of possible responses, and this may well be true, as automaticity becomes hyper-specific. Over the years, it became fairly standard to see the asymmetric interference of the Stroop effect as being due to a kind of horse race or speed-of-processing account, in which the faster-to-process dimension interfered with the slower-to-process dimension, and not vice versa (see Dyer, 1973b). This speed-of-processing is very intuitive, and is widely invoked in textbooks. But it is almost certainly wrong. Studies have shown that (a) the slower dimension can interfere with the faster to the same extent as the faster interferes with the slower (e.g., Dunbar & MacLeod, 1984), (b) giving the slower dimension a head start, even a substantial one, does not reverse the pattern of interference (Glaser & Glaser, 1982), and (c) dimensions that are processed at different speeds can nevertheless interfere with each other equivalently (MacLeod & Dunbar, 1988). Relative speed accounts cannot accommodate these results, and their reliance on a serial model of processing is now also seen as dated.

Instead, more contemporary explanatory ideas emphasize the strength of the disposition to make particular responses. They also move away from the older view of processing as involving the sequential queuing of processes, rather like a relay race (cf. Dyer, 1973b), and favor instead parallel processing occurring along multiple dimensions simultaneously. Thus, Logan (1980) painted the Stroop effect as the result of competition between ongoing processing of the word and the color dimensions at the same time, with differential rates of gain of evidence along the two dimensions. The strength idea grows out of automaticity and has been most explicitly realized in connectionist models, also known as parallel distributed processing models. Cohen, Dunbar, and McClelland (1990; see also Phaf, Van der Heijden, & Hudson, 1990) built a simple model that captured many of the results that

MacLeod (1991a) specified as pivotal for any account of the Stroop effect to explain. To account for interference, the Cohen et al. model relied on the buildup of practice for the word pathway being greater than that for the color pathway, a process tuned by attention. Like Logan's model, evidence accrual occurred in parallel along the two dimensions.

There are numerous other accounts of the Stroop effect (e.g., Virzi & Egeth, 1985), with new ones appearing regularly. Both Melara and Algom (2003) and Roelofs (2003) have presented powerful new theoretical frameworks. One conclusion that can be drawn on this basis is that assuming one of the existing accounts to be the correct one at this point is premature. For this reason, interpreting the results of Stroop experiments as evidence for a particular type of processing or for a particular process is suspect. One illustration of this principle is the concept of inhibition. Does the word have to be inhibited for the color to be named? This idea is widespread and quite intuitive, but the evidence in favor of it is not compelling. MacLeod, Dodd, Sheard, Wilson, and Bibi (2003) have tackled the question of inhibition in cognition more broadly and do not find strong support for inhibitory processing.[6] So it is important to devote considerable thought to the nature of the processing that is happening and not to assume a kind of one-to-one mapping between task and process, as is too often done.

CONCLUSION

In this chapter, I have set out how Stroop experiments are usually done, indicating along the way some of the methodological hurdles that are likely to be encountered. My goal has been to identify the important experimental parameters that should be considered in constructing a Stroop experiment. Such considerations are especially salient in the always complex realm of abnormal behavior. No experiment makes sense outside the context of theory, so I have also sketched out some of the theoretical ideas that have been proposed for explaining interference. My hope is that bringing these ideas together in one place will be helpful to researchers who wish to use the Stroop task as a tool for understanding attention, memory, and other basic cognitive processes, including their clinical ramifications. Studies of disrupted cognitive processing should inform our understanding of the normal operation of cognitive processing, just as the reverse should be true.

---

[6] As a consequence, I would prefer to banish the designation *Stroop inhibition* and always refer instead to the performance cost in the incongruent condition relative to the control condition as *Stroop interference*.

REFERENCES

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, 80(3, Pt. 2), 1–46.

Ben-Tovim, D. I., Walker, M. K., Fok, D., & Yap, E. (1989). An adaptation of the Stroop Test for measuring shape and food concerns in eating disorders: A quantitative measure of psychopathology? *International Journal of Eating Disorders*, 8, 681–687.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*, 221–225.

Bibi, U., & MacLeod, C. M. (2004.) *Stroop, Klein, and interference: Modern benchmarks for the color-word interference task.* Manuscript in preparation.

Brown, M., & Besner, D. (2001). On a variant of Stroop's paradigm: Which cognitions press your buttons? *Memory & Cognition*, *29*, 903–904.

Burt, J. S. (1994). Identity primes produce facilitation in a colour naming task. *Quarterly Journal of Experimental Psychology*, *47*A, 957–1000.

Burt, J. S. (1999). Associative priming in color naming: Interference and facilitation. *Memory & Cognition*, *27*, 454–464.

Burt, J. S. (2002). Why do non-color words interfere with color naming? *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1019–1038.

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., & Rauch, S. L. (1998). The Counting Stroop: An interference task specialized for functional neuroimaging: Validation study with functional MRI. *Human Brain Mapping*, 6, 270–282.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36*, 387–395.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology*, *40*, 343–367.

Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 397–401.

Chmiel, N. (1984). Phonological recoding for reading: The effect of concurrent articulation in a Stroop task. *British Journal of Psychology*, *75*, 213–220.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.

Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology*, *100*, 47–53.

Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the Stroop color-word task. *Perception & Psychophysics*, *11*, 274–276.

Dalrymple-Alford, E. C., & Azkoul, J. (1972). The locus of interference in the Stroop and related tasks. *Perception & Psychophysics, 11*, 385–388.

Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the Stroop Color-Word Test. *Perceptual & Motor Skills, 23*, 1211–1214.

Dunbar, K., & 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.

Dyer, F. N. (1971). Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 10, 297–302.

Dyer, F. N. (1973a). Interference and facilitation for color naming with separate bilateral presentations of the word and color. *Journal of Experimental Psychology*, 99, 314–317.

Dyer, F. N. (1973b). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition, 1*, 106–120.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Fox, E. (1995). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review, 2*, 145–173.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

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, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 875–894.

Henik, A., Friedrich, F. J., & Kellogg, W. A. (1983). The dependence of semantic relatedness effects upon prime processing. *Memory & Cognition, 11*, 366–373.

Hentschel, U. (1973). Two new interference tests compared to the Stroop Color-Word Test. *Psychological Research Bulletin, Lund University, 13*, 1–24.

Jensen, A. R. (1965). Scoring the Stroop test. *Acta Psychologica, 24*, 398–408.

Jensen, A. R., & Rohwer, W. D. (1966). The Stroop color-word test: A review. *Acta Psychologica, 25*, 36–93.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77*, 576–588.

Klopfer, D. S. (1996). Stroop interference and color-word similarity. *Psychological Science, 7*, 150–157.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*, 523–553.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109*, 376–400.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology, 36*, 684–700.

MacLeod, C., & Rutherford, E. M. (1992). Anxiety and the selective processing of emotional information: Mediating roles of awareness, trait and state variables, and personal relevance of stimulus materials. *Behaviour Research & Therapy, 30*, 479–491.

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

MacLeod, C. M. (1991b). John Ridley Stroop: Creator of a landmark cognitive task. *Canadian Psychology, 32*, 521–524.

MacLeod, C. M. (1996). How priming affects two speeded implicit tests of remembering: Naming colors versus reading words. *Consciousness & Cognition, 5*, 73–90.

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., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (vol. 43, pp. 163–214). San Diego, CA: Academic Press.

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., & Hodder, S. L. (1998). Presenting two incongruent color words on a single trial does not alter Stroop interference. *Memory & Cognition, 26*, 212–219.

MacLeod, C. M., & MacDonald, P. A. (2000). Inter-dimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*, 383–391.

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 304–315.

Mari-Beffa, P., Estevez, A. F., & Danziger, S. (2000). Stroop interference and negative priming: Problems with inferences from null results. *Psychonomic Bulletin & Review, 7*, 499–503.

Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research & Therapy, 23*, 563–569.

May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of negative priming. *Psychological Bulletin, 118*, 35–54.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*, 422–471.

Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21*, 627–645.

Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *Quarterly Journal of Experimental Psychology, 43A*, 907–912.

Mogg, K., Bradley, B. P., Williams, R., & Mathews, A. M. (1993). Subliminal processing of emotional information in anxiety and depression. *Journal of Abnormal Psychology, 102*, 304–311.

Mogg, K., Kentish, J., & Bradley, B. P. (1993). Effects of anxiety and awareness on color identification latencies for emotional words. *Behaviour Research & Therapy, 31*, 559–567.

Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition, 29*, 137–151.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.

Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Hillsdale, NJ: Erlbaum.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs, 76* (1, Pt. 2), 1–25.

Peterson, J., Lanier, L. H., & Walker, H. M. (1925). Comparisons of white and negro children in certain ingenuity and speed tests. *Journal of Comparative Psychology, 5*, 271–283.

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., & 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.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.

Ray, C. (1979). Examination stress and performance on a color-word interference test. *Perceptual & Motor Skills, 49*, 400–402.

Redding, G. M., & Gerjets, D. A. (1977). Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual & Motor Skills, 45*, 11–17.

Regan, J. E. (1978). Involuntary automatic processing in color-naming tasks. *Perception & Psychophysics, 24*, 130–136.

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

Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 515–537.

Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology, 15*, 462–471.

Severance, L. J., & Dyer, F. N. (1973). Failure of subliminal word presentations to generate interference to color naming. *Journal of Experimental Psychology, 101*, 186–189.

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition, 26*, 1033–1040.

Sharma, D., & McKenna, F. P. (2001). The role of time pressure on the emotional Stroop task. *British Journal of Psychology, 92*, 471–481.

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.

Solso, R. L. (1971). Meaningfulness of colors. *Psychonomic Science, 23*, 301–303.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662. (Reprinted in 1992 in the *Journal of Experimental Psychology: General, 121*, 15–23.)

Stroop, J. R. (1938). Factors affecting speed in serial verbal reactions. *Psychological Monographs, 50*, 38–48.

Tecce, J. J., & Dimartino, M. (1965). Effects of heightened drive (shock) on performance in a tachistoscopic color-word interference task. *Psychological Reports, 16*, 93–94.

Tecce, J. J., & Happ, S. J. (1964). Effects of shock-arousal on a card-sorting test of color-word interference. *Perceptual & Motor Skills, 19*, 905–906.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College Press.

Thurstone, L. L. (1944). A *factorial study of perception*. Chicago: University of Chicago Press.

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effect by manipulating expectation for color related stimuli. *Memory & Cognition, 20*, 727–735.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*, 34–80.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology, 47A*, 631–650.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology & Aging, 13*, 120–126.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13*, 304–319.

Warren, R. E. (1972). Stimulus encoding and memory. *Journal of Experimental Psychology*, *94*, 90–100.

Warren, R. E. (1974). Association, directionality, and stimulus encoding. *Journal of Experimental Psychology*, *102*, 151–158.

Watts, F. N., McKenna, F. P., Sharrock, R., & Trezise, L. (1986). Colour naming of phobia-related words. *British Journal of Psychology*, *77*, 97–108.

Whitney, P. (1986). Processing category terms in context: Instantiations as inferences. *Memory & Cognition*, *14*, 39–48.

Whitney, P., & Kellas, G. (1984). Processing category terms in context: Instantiation and the structure of semantic categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 95–103.

Whitney, P., McKay, T., Kellas, G., & Emerson, W. A. (1985). Semantic activation of noun concepts in context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 126–135.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, *120*, 3–24.

Windes, J. D. (1968). Reaction time for numerical coding and naming of numerals. *Journal of Experimental Psychology*, *78*, 318–322.