Psychological Review
1991, Vol. 98, No. 1, 122–142

0033-295X/91/$3.00

The Time Course of Lexical Access in Speech Production: A Study of Picture Naming

Willem J. M. Levelt and Herbert Schriefers
Max-Planck-Institut für Psycholinguistik
Nijmegen, The Netherlands

Dirk Vorberg
Technische Universität Braunschweig
Braunschweig, Federal Republic of Germany

Antje S. Meyer and Thomas Pechmann
Max-Planck-Institut für Psycholinguistik
Nijmegen, The Netherlands

Jaap Havinga
Nijmegen University
Nijmegen, The Netherlands

Lexical access in object naming involves the activation of a set of lexical candidates, the selection of the appropriate (or target) item, and the phonological encoding of that item. Two views of lexical access in naming are compared. From one view, the 2-stage theory, phonological activation follows selection of the target item and is restricted to that item. From the other view, which is most explicit in activation-spreading theories, all activated lexical candidates are phonologically activated to some extent. A series of experiments is reported in which subjects performed acoustic lexical decision during object naming at different stimulus-onset asynchronies. The experiments show semantic activation of lexical candidates and phonological activation of the target item, but no phonological activation of other semantically activated items. This supports the 2-stage view. Moreover, a mathematical model embodying the 2-stage view is fully compatible with the lexical decision data obtained at different stimulus-onset asynchronies.

One of a speaker's core skills is to lexicalize the concepts intended for expression. Lexicalization proceeds at a rate of two to three words per second in normal spontaneous speech, but doubling this rate is possible and not exceptional. The skill of lexicalizing a content word involves two components. The first one is to select the appropriate lexical item from among some tens of thousands of alternatives in the mental lexicon. The second one is to phonologically encode the selected item, that is, to retrieve its sound form, to create a phonological representation for the item in its context, and to prepare its articulatory program. An extensive review of the literature on lexicalization can be found in Levelt (1989). This article addresses only one aspect of lexicalization, namely its time course. In particular, we examine whether the selection of an item and its phonological encoding can be considered to occur in two successive, nonoverlapping stages.

This is by no means a novel concept. One should rather say that it is the received view in the psycholinguistic literature (see especially Butterworth, 1980, 1989; Fromkin, 1971; Garrett, 1975, 1976, 1980; Kempen, 1977, 1978; Kempen & Huijbers, 1983; Levelt, 1983, 1989; Levelt & Maassen, 1981; Morton, 1969; Schriefers, Meyer, & Levelt, 1990). The first stage, lexical selection, makes available a semantically specified lexical item with its syntactic constraints. Kempen (1977, 1978) called this a *lemma*. Lemmas figure in grammatical encoding, specifically in the creation of syntactic frames. During the second stage, phonological encoding, phonological information is retrieved for each lemma. These phonological codes are used to create the articulatory plan for the utterance as a whole. Both Garrett (1976) and Kempen (1978), following Fry (1969), have stressed that the grammatical encoding and phonological encoding of an utterance normally run in parallel. Grammatical encoding, of which lexical selection is a proper part, is just slightly ahead of phonological encoding. The phonological encoding of a given item overlaps in time with the selection of a subsequent item. Only at the level of individual lexical items can one speak of successive stages. An item's semantic–syntactic makeup is accessed and used before its phonological makeup becomes available.

Garrett (1975, 1976) argued for this separation of stages on the basis of speech error data. He distinguished between two classes of errors, word exchanges and sound exchanges, and could show that these classes differ in distributional properties. Word exchanges occur between phrases and involve words of the same syntactic category (as in *this spring has a seat in it*). Sound exchanges typically involve different category words in the same phrase (as in *heft lemisphere*). Word exchanges are

---

We acknowledge the invaluable contributions of John Nagengast and Johan Weustink, who programmed the computer-based experiments; of Ger Desserjer and Hans Fransen, who ran the experiments and assisted in data analysis; and of Inge Tarim, who provided graphical assistance. We also acknowledge Gary Dell's and Pienie Zwitserlood's detailed comments on an earlier version of this article, as well as the thorough comments of an anonymous reviewer.

Herbert Schriefers is now at Freie Universität Berlin, Berlin, Federal Republic of Germany, and Thomas Pechmann is now at Universität des Saarlandes, Saarbrücken, Federal Republic of Germany.

Correspondence concerning this article should be addressed to Willem J. M. Levelt, Max-Planck-Institut für Psycholinguistik, Wundtlaan 1, NL-6525 XD Nijmegen, The Netherlands.

unaffected by phonological factors, whereas sound exchanges occur regardless of whether their products correspond to existing lemmas or not. Garrett (1975, 1976) adduced the genesis of these two classes of errors to the first and the second encoding stages, respectively. In Garrett (1988), this analysis was further qualified; Garrett made a distinction between two types of substitutions, namely those that have to do with the selection of lemmas (e.g., *toe* for *finger*) and those that have to do with the retrieval of sound forms (e.g., *mushroom* for *mustache*). These two error sources are quite independent: "Target and intrusion words related in form rarely show a meaning relation, and conversely" (Garrett, 1988, p. 73). So-called combined or mixed errors, where there are both form and meaning relations between target and intrusion (e.g., *lobster* for *oyster*), are rare. Still, they are more likely than chance, and we return to them later.

Further evidence for the claim that lexical access proceeds in two nonoverlapping stages has come from experimental work by Levelt and Maassen (1981) and by Kempen and Huijbers (1983). In both studies, subjects described events or scenes presented to them, and the voice onset latencies of their utterances were measured. In both studies, the obtained latencies for different kinds of syntactic forms could best be explained by assuming a strict succession of lemma selection and phonological encoding.

In a recent study, Schriefers et al. (1990) used an interference paradigm where subjects named pictures while they were auditorily presented with distracter words, which they had to ignore. These distracter stimuli could appear at different (negative and positive) stimulus-onset asynchronies (SOAs) with respect to the pictures. It was found that *semantic* distracters (e.g., *goat* when the naming target was *sheep*) only affected the naming latencies when presented before the picture appeared. However, *phonological* distracters (e.g., *sheet* when the target was *sheep*) affected naming latencies only when presented simultaneously with the picture or shortly after the picture was shown. This result supports the notion that semantic and phonological activations of the target word are strictly successive.

Other arguments for this two-stage view of lexical access have come from tip-of-the-tongue studies (Brown & McNeill, 1966; see Levelt, 1989, for a review of the subsequent literature). The speaker knows the word, arranges the appropriate syntactic context for it to appear, and then blocks partly or wholly on retrieving its phonological form. Jones and Langford (1987) have claimed that this blocking in the second stage can even be induced or aggravated by presenting the speaker with a word that is phonologically related to the target (e.g., *secant* when the target word is *sextant*). However, a semantic distracter item (e.g., *latitude*), was totally ineffective. In their study, however, Meyer and Bock (1990) found no evidence that a phonologically related nuisance word creates extra blocking of access.

Butterworth (1989), reviewing the spontaneous speech error evidence, the experimental evidence, as well as the evidence from hesitation pauses and from aphasiology, concluded

> that lexical access in speech production takes place in two temporally distinct stages. In the first stage, the speaker accesses a "semantic lexicon." . . . This, in essence, is a transcoding device, that takes as input a semantic code and delivers as output an address. The second stage takes the address as input to another transcoding device, the "phonological lexicon," . . . and delivers a phonological word form as output. (p. 110)

In this article, we call this the *discrete two-stage model*, or shorter, the *two-stage model*. To compare this model to alternative views, it is important to distinguish between *activation* and *selection*. Many theories of lexical access assume that items are activated before they become selected. Among the two-stage models, this is most explicitly the case in Morton's (1969, 1979) model. Lexical items are mentally represented as *logogens* in this theory, which are devices that collect evidence for semantic, pragmatic, or other appropriateness of "their" word. All logogens are simultaneously active in collecting the specific information to which they are sensitive. When one of them reaches a certain threshold activation, it fires (i.e., is selected) and makes its phonological code available for articulation. At that moment, the logogen's activation level drops back to zero. This is a strict two-stage model; phonological encoding follows lexical selection. Of all activated items, only the selected one becomes phonologically encoded. It is characteristic of all discrete two-stage models that phonological encoding is restricted to the selected item. The time-course prediction is that there is no phonological, but only semantic, activity during Stage 1. From the most stringent view, there is no semantic, but only phonological, activity during Stage 2. This holds, for instance, for Morton's model, where the logogen's activation drops back to zero after it fires. Although most two-stage theories are not explicit with respect to the issue of whether a selected item's semantic activation decays sharply or gradually, we test the most stringent view because it is the most vulnerable one. Less stringent versions would allow both semantic and phonological activation at the beginning of Stage 2.[1]

The main experimental tests reported in this article involved picture naming. We probed the semantic and phonological activation evolving between the presentation of a picture and the onset of overt articulation of the picture's name. Figure 1 depicts the two-stage theory that we tested. Figure 1a shows the two stages of access. During the first stage, right after seeing the picture (more precisely, right after conceiving of the concept to be lexically expressed), there is semantically driven activation of a set of lemmas. We call this set the *semantic cohort*. This is a set of one or more meaning-related items that receive activation from the input concept. Eventually, only one of these semantic alternatives survives the selection process; we call it the *target item*. During the second stage, this target item, and only this item, becomes phonologically encoded. That is, an articulatory plan is constructed for just that item.

Figure 1b shows three activation functions predicted by this schema: semantic activation and two types of phonological activation. During the first stage of semantic activation, the activation of the target item increases until the moment of selection. Thereafter, it drops back to zero and stays there during the second stage. The target item's phonological activation is at zero level during the first stage and increases after the moment of selection, that is, during the second stage. Also shown is the

---

[1] If the stringent version finds support in this study on naming, it does not follow that in other tasks (e.g., semantic category decision) there will not be more sustained semantic activation.

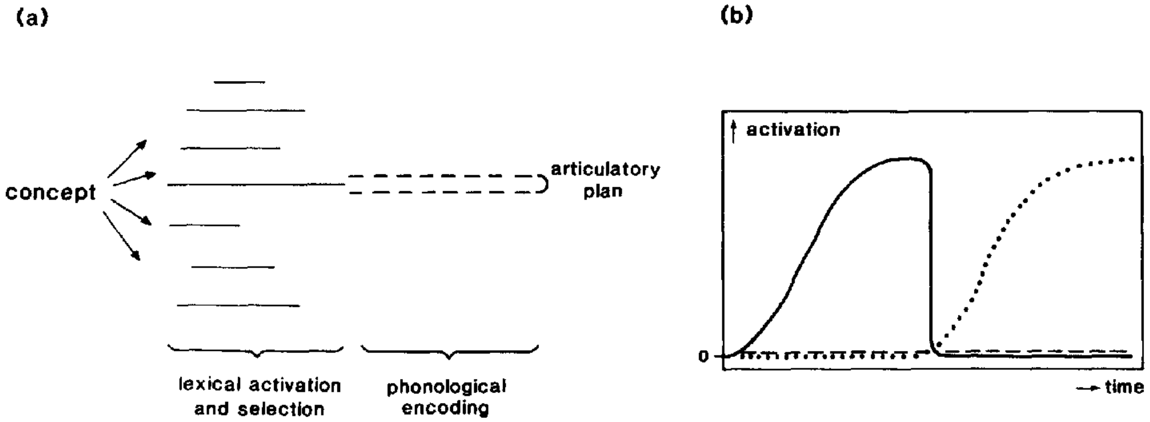

Figure 1. The discrete two-stage theory of lexical access. (a: Stages of lexical activation–selection and of phonological encoding. b: Schematic diagram of the time course of semantic [solid line] and phonological [dotted line] activation of target and of phonological activation of semantic alternatives [dashed line].)

phonological activation of semantic alternatives. The theory predicts that these items will at no time become phonologically activated; hence, the flat, zero-level shape.

This strict two-stage theoretical picture has not remained unchallenged. The connectionist or activation-spreading models of lexical access proposed by Dell (1986, 1988, 1989), MacKay (1987), Stemberger (1985), and others predict that not only the selected (i.e., target) item becomes phonologically activated, but also any activated semantic alternative items. This is a consequence of the mechanics of a connectionist network. The lexicon, according to this conception, involves (at least) three levels of nodes. At the top (conceptual) level, nodes represent concepts or conceptual features of some kind. When they become activated, they spread their activation to the middle (lexical) level. Here, nodes represent lexical items or, in Kempen's terminology, lemmas. For instance, if the conceptual node representing animateness is active, it will spread its activation to all animate lexical nodes, such as the ones representing *bear, sheep,* or *lion.* The set of activated lexical nodes is equivalent to the semantic cohort introduced earlier. What pattern of lexical activation will result from a given pattern of conceptual activation depends, of course, on the pattern of connections between conceptual and lexical levels. Eventually, the most activated node will be selected as the target item. In turn, the activated lexical items will spread their activation to the bottom (phonological) level. In this case, nodes represent various aspects of a word's phonological structure. There are, in particular, phoneme nodes and (dependent on the theory) nodes for phoneme clusters and for phonological features.

A natural property of the sketched pattern of connectivity between levels is that any *activated* lexical item or lemma will spread its activation to its constituent nodes at the phonological level. Contrary to the two-stage theories, the activation-spreading models predict that not only the one selected (target) item but also the coactivated semantic alternatives become phonologically active (see Figure 2a). The phonological units of all semantically activated lexical items receive some activation, but only the units of the target item eventually become selected for articulation. Figure 2b gives the schematic time-course predictions for semantic and phonological encoding for this simplest case. As activation spreads from the conceptual to the lexical level, semantic activation of the target item increases up to some critical level. In most theories, there is an assumption that the target item's activation reduces to resting level shortly after its selection. Thereafter, various things may happen, but in the simplest case where there is only forward spreading of activation, semantic activation will stay at resting level. The phonological activation of the target (dotted line) begins shortly after its semantic activation and overlaps with it in time. Phonological activation increases until articulation sets in. Especially remarkable is the curve (dashed line) for the phonological activation of semantic alternatives. It is not the flat curve of the two-stage theory. Rather, the target item and its semantic alternatives simultaneously spread their activation to the phonological level. After selection of the target item, its phonological activation increases, whereas the phonological activation of the semantic alternatives decays.

This, however, is not the whole story for the activation-spreading theories. Although some of them, such as Humphreys, Riddock, and Quinlan's (1988) cascade model, only assume forward spreading of activation, most also allow for backward spreading, in particular from the phonological to the lexical level. This is precisely quantified in Dell's (1986, 1988) model of speech production. A main reason for introducing a mechanism of backward spreading of activation was the finding by Dell and Reich (1981), Stemberger (1985), and others that phonological speech errors result in real words more often than should be statistically expected. This is called the *lexical bias effect.* Lexical bias effects have also been demonstrated experimentally, initially by Baars, Motley, and MacKay (1975) and later by Dell (1985). In these experiments, the probability is about three times higher that a word pair such as *darn bore* would slip to *barn door* than a pair like *deal back* would slip to *beal dack;* only in the former case does the slip produce real words.

Dell's (1986, 1988, 1989) model accounts for lexical bias in sound form errors by assuming backward spreading of activation from lower level phonological nodes to higher level lexical

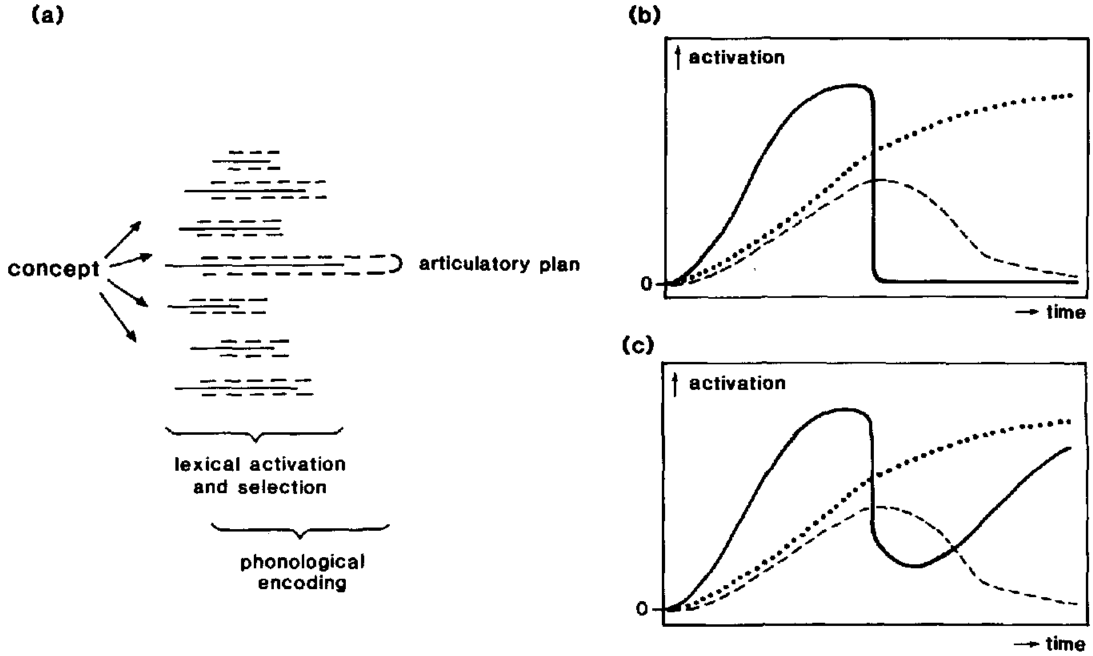

Figure 2. Activation-spreading (network) theories of lexical access. (a: Overlapping lexical activation and phonological encoding. b and c: Schematic diagrams of the time course of semantic [solid lines] and phonological [dotted lines] activations of target and of phonological activation of semantic alternatives [dashed lines] for forward-only and backward-spreading activation models, respectively.)

nodes. So, for instance, when the lexical node *darn* spreads its activation to the phoneme nodes /d/, /a/, /r/, and /n/, the last three will, in turn, spread their activation backward not only to the lexical node *darn,* but also to *barn, yarn,* and so on. This increases the probability that these words will appear as slips. Because there are no nonlexical nodes (e.g., for *beal* or *dack*), the likelihood of a lexical slip exceeds that of a nonlexical slip (everything else being equal).

The lexical bias effect can be understood not only in terms of backward spreading of activation, but also some other phenomena. One is the so-called *repeated phoneme effect* (MacKay, 1970). Two phonemes are more likely to exchange when their neighboring phonemes are identical. The error *kit to fill* (for *fit to kill*) is a more likely error than, say, *kit to fall* (for *fit to call*). In the former case, the lemma *fit* activates its vowel /I/, which in turn activates the lemma *kill* (by backward spreading). The increased activation of *kill* then spreads to its constituent phoneme /k/, which may then be erroneously selected, producing the error /kit/. This chain of forward–backward–forward activation will not arise in the case of *fit to call* because *fit* and *call* have no phonemes in common.

Another phenomenon that might find its explanation in a mechanism of backward spreading of activation is the case of *mixed errors,* such as *oyster* for *lobster,* or *cat* for *rat.* Here, feedback from the phonological level increases the likelihood that a (already active) semantic alternative will be selected instead of the target. Without further assumptions, a strict two-stage model would predict that the probability of a mixed error, *Pm,* is the product of the probabilities of a semantic error, *Ps,* and of a phonological error, *Pp;* that is, $Pm = Ps \cdot Pp$. But Dell and Reich (1981) showed that this estimation of the rate of mixed errors was substantially surpassed in their collection of naturally occurring speech errors. This finding was the main impetus for the development of an activation-spreading model. Meanwhile, this higher-than-chance occurrence of mixed errors repeatedly has been shown to arise in both natural error data (Harley, 1984; Stemberger, 1983) and experimentally obtained error data (Martin, Weisberg, & Saffran, 1989). The backward-spreading mechanism gives a natural account of these findings.

One consequence of this theory is that the backward spreading leads to late lexical–semantic activation. Hence, there will be both early and late semantic activation, as well as late phonological activation. These predictions are depicted in the activation curves of Figure 2c. The phonological activation curves are qualitatively the same as the ones in Figure 2b, but the course of semantic activation is different. There is a rebound of semantic activation during the later access phase. We call this the *backward-spreading* predictions, to distinguish them from the *forward-only-spreading* predictions in Figure 2b.

In this article, we compare the three sets of predictions schematized in Figures 1b and 2 (b and c)—that is, the activation predictions generated by the two-stage model and by the forward-only and the backward-spreading connectionist theories. In particular, we set out to test the course of semantic and phonological activation of the target item in naming tasks. We also tested the phonological activation of semantic alternatives.

The main experimental task was one of naming. Subjects were asked to name pictures one after another. Occasionally, an acoustic test probe was presented shortly after presentation of a picture, although before the naming response had set in. Test probes were words or nonwords. The subject's secondary task was to give a manual lexical decision response to such a test probe. In the critical trials, the test probe was semantically or phonologically related to the target name. So, if the target object was a sheep, the acoustic test probe could be *wool* (semantically related) or *sheet* (phonologically related). Although our object of study was the evolution of the naming response, lexical decision latencies were the dependent measures in these experiments. We assumed that semantic activation of the target item would affect the lexical decision latency for a semantically related test probe, whereas phonological activation would affect the decision latency for a phonologically related probe. In other words, we expected that the lexical decision would probe the current state of activation in the preparation of the naming response. By varying the SOA between presentation of the picture and presentation of the acoustic test probe, semantic and phonological activation could, we hoped, be traced over time.

It is a priori not obvious how lexical decision latencies will depend on different states of activation. If there is semantic or phonological activation in the preparation of the naming response, how will this affect the lexical decision response? Will there be interference or facilitation in case of a semantic or phonological relation between naming target and acoustic test probe? Whichever effect is found, it will be indicative of a particular type of informational contact between the production and comprehension systems. Only a null effect (predicted by theories that make a strict separation between input and output lexicon) would be really problematic. Later in this article, we develop a model of the interaction between the naming and lexical decision tasks that will specify the direction of the effects. Initially, however, the approach is purely experimental: Can we find effects of semantic and phonological activation on lexical decision at different stages in the preparation of a naming response, and if so, what do they tell us about the time course of semantic and phonological activation?

This experimental procedure required careful preparation of materials. The first section following this introduction describes these preparations. The second section presents the main lexical decision experiments, the results of which are on first, informal analysis supportive of neither the two-stage model nor the feedback connectionist model. In the third section, two experiments are reported that test the phonological activation of semantic alternatives. A main conclusion from these experiments is that semantic alternatives to the target item are not phonologically activated, contrary to the predictions of connectionist models but in agreement with the two-stage model. In the fourth section, we develop a mathematical account that reconciles the results of the main lexical decision experiments with the two-stage model. Whether these data can also be reconciled with activation-spreading models is at issue in the General Discussion section.

PREPARATION OF MATERIALS

The main experiment contained 16 critical target pictures as well as a large set of filler items. This section describes how we selected the critical targets.

Experiment 1: Naming

The objectives of this experiment were to select "unanimous" items, meaning pictures given the same name by most subjects, and to determine the naming latencies for these pictures.

Method

Three hundred pictures of objects were collected from different sources and mounted on 2 × 2-in. (4.08 × 4.08 cm) slides. Each of 20 subjects (students paid for their services) was presented with 4 warm-up slides, followed by all 300 experimental slides. The instruction was to name each picture as quickly as possible. Presentations were paced 4.3 s apart; each slide was visible for 2 s. Presentation order was systematically varied among subjects. Responses were recorded on one channel of an audiotape. The other channel contained trigger pulses relating to slide onset. Naming latencies were determined from these audio recordings by means of a voice key.

Results

There was full agreement among subjects in the labeling of 78 slides; 19 subjects agreed on a different set of 43 slides, and 18 subjects on another set of 31 slides. Setting our cutoff point at an agreement of 18 out of 20 subjects, we were left with 152 slides for which the naming latencies were measured from the audiotape recordings. The mean naming latencies for these slides ranged from 649 to 1,330 ms. These slides were further explored in Experiment 2.

Experiment 2: Recognition

A subject's naming latency obviously involves various components (Glaser & Glaser, 1989; Seymour, 1979; Theios & Amrhein, 1989). There is, first, the pictorial encoding of the presented object. The speaker must recognize the object to be named. Second, there are lexical access, lexical selection, and phonological encoding, which may or may not overlap in time. Finally, there is the initiation of articulatory execution (cf. Levelt, 1989). For the main experiment, we needed a set of test pictures for which the recognition times were homogeneous and long. Homogeneity was a requirement for the following reason: To trace the time course of lexical access, it is important to minimize the variability of the onset of lexical access for the test probes used. Because, by hypothesis, lexical access immediately follows object recognition, recognition times for the objects used should be maximally similar.

They should also be long for the critical items in the main experiment. The presentation and the processing of an acoustic lexical decision probe is itself a process extended over time. If lexical decision latencies are to be affected by Stage 1 of lexical access, that is, by the speaker's semantic search for the target, the acoustic test word must be recognized before the speaker enters Stage 2 of lexical access, the phonological encoding of the target name. Ideally, recognizing the acoustic test probe should, for short SOAs, be completed during the phase of object recognition. To allow sufficient time for this to occur, the pictures used should have long recognition latencies. The experiment was designed to select pictures with homogeneous and long recognition latencies and is a slightly modified version of

Wingfield's (1968) procedure of measuring object recognition latencies.

Method

The starting point for the selection of materials of this experiment was the set of 300 slides used in Experiment 1. They consisted of two sets. There were the 152 (almost) unanimous items of Experiment 1, and there were the remaining 148 nonunanimous items. Two of the latter set had to be discarded for independent pictorial reasons, leaving us with 146 nonunanimous items. A subject saw a randomized series of these 298 slides.

In the experiment the subject was presented with 298 word–picture pairs. The word was acoustically presented (from tape), and the picture appeared 1 s after word offset. The subject's task was to decide whether word and object agreed, that is, whether the object in the picture was the one mentioned. To make this decision, the subject had to recognize the object in the picture. Yes and no responses were given by means of pushbuttons. Half the subjects gave yes responses with the right index finger, the other half with the left one. The subjects' task was to decide as quickly as possible, and reaction times were measured.

The pairs were arranged in such a way that for all 152 unanimous items, the response had to be yes. For the 146 nonunanimous items, the response had to be no; these pairs were constructed by assigning the 146 picture names to the wrong pictures.

Onsets of word–picture pairs were paced 5 s apart. Each slide was visible for 2 s. There were 20 paid subjects in this experiment (students of Nijmegen University), who had not participated in Experiment 1.

Results

Average recognition latencies for the 152 unanimous items ranged from 417 to 684 ms. The recognition latencies for these items, together with their naming latencies measured in Experiment 1, were used to select the 16 experimental target pictures for the main experiment. The 16 items had to have (a) long but homogeneous recognition latencies and (b) long but homogeneous naming latencies. The final list of the 16 selected target pictures, their names, and naming and recognition latencies are presented in Table 1.

Selection of Lexical Decision Items

In the main experiment, each of the 16 target pictures was followed by an acoustic test probe for lexical decision. These lexical decision items, or test probes, could be of four kinds. They could be semantically related to the target word (e.g., the picture shows a sheep, hence the target word is *sheep;* the test probe is a close associate, *wool*); they could be phonologically related (e.g., the target word is *sheep,* and the test probe is *sheet*); they could be identical to the target name (e.g., the target word is *sheep,* and the test probe is *sheep*); and, finally, they could be unrelated to the target (e.g., target *sheep,* test probe *knife*).

The semantically related test probes were selected as follows. The names of the 16 target pictures were, in written form, presented to 49 subjects (students of Nijmegen University not involved in the earlier experiments). They were asked to write down their first associate word to each of the 16 items. The most frequent associates were selected as semantic test probes.

The phonologically related test probes were recruited from the dictionary. For each target name, we took a test word that maximally shared the target word's initial sequence of phonemes. This common stretch was minimally word-initial vowel–consonant or consonant–vowel, but in most cases more (e.g., consonant–consonant–vowel–consonant in target *krokodil* and test probe *kroket*).

The unrelated test probes were all monosyllabic words. An unrelated test item had no semantic relation to the target word, nor did it share a word-initial segment sequence with the target word. The semantic, phonological, and unrelated test words for the 16 target pictures are also listed in Table 1.

MAIN EXPERIMENTS

Experiment 3: Lexical Decision and SOA

In this experiment, a subject named a series of pictures, among them the 16 target pictures. The target pictures were sooner or later followed by an acoustic test probe. The test probe could come early, shortly after presentation of the picture; it could come late, shortly before the naming response was initiated; or it could come somewhere in between these two extremes. Hence, in all three SOA conditions (short, medium, and long), the acoustic test stimulus followed the target picture. For each target picture, there were four acoustic test probes, a semantically related one, a phonologically related one, an item identical to the target word, and an item unrelated to the target. Hence, there were 64 critical acoustic probes. A subject would receive only 16 of these 64 items in total, 1 for each target picture. We measured lexical decision latencies for these four kinds of probe. As remarked earlier, these lexical decision latencies were measured to trace the evolution of the naming response. This makes the present paradigm rather different from the standard naming paradigm, where a distracter stimulus is presented at different SOAs (as in Glaser & Glaser, 1989, or Schriefers et al., 1990). In the latter paradigm, the naming latency is the dependent measure, and the distracter stimulus can precede or follow the picture (negative and positive SOAs, respectively). Technically speaking, our paradigm uses negative SOAs only: The lexical decision item to which the subject responds is always *preceded* by the picture. However, we will ignore the negative sign and express SOAs in absolute values. Before describing the method of this experiment in detail, we must introduce the presession in which each subject participated.

Presession

We expected that the 64 critical test probes would vary greatly in their lexical decision latencies. They differ in duration, in phonological structure, in frequency of usage, and in meaning. To control for these sources of variation, each test probe was made its own control by means of a preexperiment. Each subject participated in the preexperiment about 1 week before the main experiment. In the preexperiment, subjects were presented with the 16 acoustic probes they would receive in the main experiment as well as with an additional 10 so-called control words (defined later). These 26 words were mixed with 26 nonwords. The resulting 52 items were presented in random order, and a subject's task was to make lexical decisions

Table 1
Pictured Objects, Latencies, and Test Probes

	Reaction time (ms)		Test probe			
Object	Naming	Recognition	Identical	Semantic	Phonological	Unrelated
1. DESK	936	576	bureau	stoel (chair)	buurman (neighbor)	muts (cap)
2. CACTUS	1,077	601	cactus	stekel (sting)	kakkerlak (cockroach)	tas (bag)
3. INFLATOR	964	560	fietspomp	band (tire)	file (queue)	wip (seesaw)
4. RIFLE	979	568	geweer	oorlog (war)	gewei (antlers)	koets (coach)
5. RAKE	926	556	hark	tuin (garden)	harp (harp)	bel (bell)
6. ICE CREAM	949	561	ijsje	zomer (summer)	ijzer (iron)	brief (letter)
7. CLOCK	1,021	551	klok	tijd (time)	klos (spool)	film (movie)
8. BUTTON	960	561	knoop	jas (coat)	knook (bone)	zwaard (sword)
9. CROCODILE	976	592	krokodil	leer (leather)	kroket (croquette)	muur (wall)
10. RADIO	984	586	radio	muziek (music)	radar (radar)	kerk (church)
11. SCREW	1,060	607	schroef	moer (nut)	schroot (scrap)	taart (tart)
12. CIGAR	990	593	sigaar	rook (smoke)	cycloon (cyclone)	poes (cat)
13. THERMOMETER	970	552	thermometer	koorts (fever)	termijn (term)	noot (nut)
14. FEATHER	933	576	veer	kip (chicken)	veen (peat)	slot (lock)
15. FINGER	1,061	578	vinger	ring (ring)	vink (finch)	kwast (brush)
16. BAG	984	547	zak	doek (cloth)	zang (song)	koe (cow)

Note. Probes are presented in Dutch, with English translations in parentheses.

for each acoustically presented probe. The list of 52 items was preceded by 20 warm-up trials, 10 words and 10 nonwords. Items were spaced apart by 2.5 s plus the subject's lexical decision reaction time, with a maximum of 4 s. In this way, we obtained a baseline lexical decision latency for each test probe and subject. The equipment for the acoustic presentation of the words and nonwords was the same as in the main session of the experiment (discussed later). Moreover, the same acoustic tokens were used in both the presession and the main session of the experiment.

Main Session

Method

Procedure. Each of the three SOA conditions (short, medium, and long asynchrony between presentation of the target picture and of the acoustic test probe) was run on a different set of 64 subjects. The critical acoustic test probe could be semantic (S), phonological (P), identical (I), or unrelated (U) (see earlier descriptions); these are the four *test-probe conditions.* A target picture was shown only once to a subject. For 4 of the 16 target pictures, a subject would receive the identical test word, for another 4, the semantic test word, and so on for the phonological and the unrelated test words. The 64 subjects in an SOA condition were divided into four groups of 16 subjects; in other words, there were four *group conditions.* All 16 subjects in a group received the same pairings of target pictures and test words. However, the pairings were rotated among the four groups: The 4 target pictures that were paired with I test probes for the first group of 16 subjects were paired with S probes for the second group of 16 subjects, with P probes for the third group, and with U probes for the fourth group, and so on.

The three *SOA conditions*—short, medium, and long—were realized as follows. Although the recognition times for the selected slides were reasonably homogeneous (ranging from 547 to 607 ms in Experiment 2), we decided to reduce the effect of variable recognition times even further by making the SOA for each target picture dependent on its recognition time. In the short SOA condition, the acoustic test probe for a picture was initiated 500 ms before the picture's mean recognition latency as measured in Experiment 2 (Table 1). Hence, the shortest short SOA was 47 ms (for Picture 16), and the longest short SOA was 107 ms (for Picture 11); the average SOA was 73 ms. In the medium SOA condition, the onset of the test probe began 200 ms before a picture's measured recognition latency. The average medium SOA was, therefore, 373 ms. In the long SOA, the acoustic test probe began 100 ms after a picture's recognition latency. Hence, the average late SOA was 673 ms.

For each SOA, the same procedure was used, although for a different set of 64 subjects. Each subject was presented with a sequence of 190 pictures of objects to be named. The instruction stressed the importance of naming each picture quickly and accurately. In addition, the subject was told that, occasionally, an acoustic test probe would be presented via the headphones during the naming task. In that case, the subject should push the yes button if the probe was a Dutch word and the no button if it was not a Dutch word; also, the naming response was to be given.

The set of 190 trials was composed as follows. There were, first, 30 practice trials. In 20 of these, only a picture was presented to be named. The other 10 practice trials involved an acoustic test probe, the SOA of which was the average SOA for the condition (short, medium, and long). Five of these probes were real words that were unrelated to the corresponding pictures; 5 were nonwords. The 30 practice trials were presented in random order. The remaining 160 trials consisted of (a) the 16 critical trials in which the 16 target pictures of Table 1 were combined with test words, four of each kind (S, P, I, and U); (b) 10 so-called control items—these were a set of 10 pairs of pictures and unrelated test words that were given to all subjects (these items, which invited a positive lexical decision response, made it possible for us to compare subjects or groups of subjects in terms of their base rate in lexical decision); (c) 26 trials, in which pictures were combined with a nonword test probe; and (d) 108 trials in which there were just pictures to be named. Of the 190 trials, therefore, only 62—less than one third—required a lexical decision response. For the trials in Conditions b and c, we used 100, 400, and 600 ms for the short, medium, and long SOA conditions, respectively. These values are within the range of the SOAs for the critical test probes.

The 160 trials were presented in quasi-random order: There were never more than three subsequent lexical decision items; 2 critical trials (of Type a) were always separated by at least 4 other items and were always of a different kind (in terms of S, P, I, and U); the first 3 items of the series were not positive lexical decision items. There was, moreover, another restriction on order of presentation. The 160 pictures were on slides, in two trays of 80. Each tray could be run forward or backward, and Tray 1 could precede or follow Tray 2, which allowed for eight different orders. In each of the four groups of subjects in an SOA condition, 2 subjects were assigned to each of these eight orders.

A trial consisted of a 1-s presentation of the picture, followed by a varying period in which the naming response should be made. For the lexical decision trials, we maintained a 2-s time-out from the beginning of the test probe. That is, if no lexical decision was initiated during that period, the response was registered as incorrect. For all lexical decision trials, the next picture appeared 5.5 s after the pushbutton reaction, leaving the subject plenty of time to complete the naming response. When the trial did not involve a lexical decision, the next trial was initiated 4.5 s after picture onset.

Stimuli. All visual stimuli were taken from the initial set of 298 (see Preparation of Materials section), and the 16 critical test stimuli were the ones listed in Table 1. The acoustic test probes consisted of (a) 64 critical word probes—the S, P, I, and U probes for each of the 16 critical slides, which are listed in Table 1; (b) 10 unrelated word probes for the control items (5 monosyllabic and 5 bisyllabic words); (c) 5 unrelated word probes for the practice trials; and (d) 31 mono-, bi-, or trisyllabic phonotactically legal nonwords, 5 of which were used in the practice trials. These nonwords were all different from the nonwords used in the presession. Words and nonwords had been spoken by a Dutch woman, tape-recorded, digitized at a sampling rate of 20 kHz, and stored on the disk of a PDP 11/55 computer.

Apparatus. The subject was seated in a dimly lit soundproof booth, facing a translucent screen on which the pictures were projected from a Kodak carousel projector outside the booth. The acoustic test probes were presented to the subject via open Sennheiser headphones. They were generated from the digitized files via a digital-to-analog converter under the control of a PDP 11/55. The same computer collected the subject's pushbutton lexical decision reaction times, measured from the onset of the acoustic test probe to the subject's pushbutton reaction. Half the subjects gave the yes responses with the right index finger and the no responses with the left one; this was reversed for the other half. The subjects' naming responses were registered via a Sennheiser microphone and recorded on one channel of a Revox tape recorder. On the other channel, timing pulses were set that corresponded to shutter openings of the slide projector. The shutter openings were controlled from the PDP 11/55.

Subjects. There were, in final analysis, 64 subjects for each of the three SOA conditions, 192 in total. Each subject participated in both the presession and the main session of the experiment. Quite a few more subjects were actually tested, but their data had to be discarded because they did not meet our strict performance criteria (8 more subjects for the short SOA, 8 more for the medium SOA, and 16 more for the long SOA). These criteria were as follows: For each condition (S, P, I, and U) there should be at least two test probes (out of four) to which the subject gave a correct lexical decision reaction in both the presession and the main session and to which he or she gave a correct naming response in the main session. Every subject not meeting this criterion was replaced. The subjects were taken from the Max Planck Institute subject pool. Most of them were undergraduate students of Nijmegen University. They were paid Dfl. 17 for their participation in the two sessions of the experiment.

Results

The main results are shown in Table 2, which displays mean lexical decision latencies for the critical test probes (S, P, I, and U) in the presession and main sessions of the three SOA conditions. Although there was obviously no SOA variable in the presession, the table presents different presession data for the three SOAs. This is so because different subjects participated in the different SOA conditions. The table presents the relevant presession data for the subjects in the corresponding SOA condition. It also presents the average values that were used in the model simulations to be discussed later.

The statistical analyses to be reported are based on what we called *differential scores.* They were obtained by subtracting each subject's decision latency for an item in the presession from that subject's decision latency for the same item in the main session. The means of these differential scores are also presented in the table. In computing differential scores, we defined missing values pairwise. That is, whenever there was a missing value in a subject's matrix of either the presession or the main experiment, the corresponding cell in the subject's matrix of differential scores was also treated as a missing value. Finally, if one data point of a presession or main session pair was missing, the other point was also removed from the data set. The presession and main session mean values in Table 2 are based on the remaining data. The reason for this strict procedure was to make sure that any obtained result could be traced back to within-subject and within-item data.

A missing value in the presession arose when no lexical decision response or an incorrect one was given for an item. In the main experiment, a missing value arose when any of the following conditions held: (a) There was no lexical decision response or an incorrect one, (b) there was an incorrect naming response, or (c) the naming response was initiated before the onset of the acoustic lexical decision probe. The latter case only occurred for the long SOA condition (in 9% of the critical trials); this explains the higher overall percentage of missing values in that

Table 2

Lexical Decision Latencies and Differential Scores (in Milliseconds) and Percentage of Missing Values for Test Probes Presented at Three Stimulus-Onset Asynchronies (SOAs) During Picture Naming (Main Session) and Without Picture Naming (Presession)

	Short SOA				Medium SOA				Long SOA			
	I	S	P	U	I	S	P	U	I	S	P	U
Presession	775	770	857	732	755	746	821	704	764	769	820	699
Main session	1,080	1,061	1,186	973	1,006	1,002	1,139	945	918	964	1,109	910
Differential	306	291	329	241	251	256	318	241	153	194	289	211
% missing	19	13	24	6	9	16	22	6	19	23	29	14

Note. Test probes: I = identical, S = semantic, P = phonological; U = unrelated. The average results for the simulation were I = 765, S = 762, P = 833, and U = 712.

SOA condition. Both the presession and main session data were corrected for outliers, which were values that exceeded 2 *SD*s from either the subject's mean or the item's mean. The missing values were replaced by Winer's (1971) procedure.

Using differential scores, as defined, involves the assumption that the presession baseline and the main session experimental effect are additive. It should, in particular, not be the case that test probes with a long (presession) lexical decision latency are more sensitive to the experimental manipulations than items with a short baseline latency. Although the results of this experiment may cast doubt on the correctness of this assumption (the phonological test probes generally produce the strongest effects; they also have, on average, the highest presession latencies; see Table 2), the subsequent experiments seem to back up the assumption (see Tables 3, 4, and 5).

Figure 3 presents the differential scores for each of the four test-probe conditions (S, P, I, and U) over the three SOAs. Subject and item analyses of variance (ANOVAs) were performed on the differential scores with SOAs and groups as between-subjects variables and test probes as a within-subject variable. The SOA variable turned out to be significant, $F_1(2, 180) = 4.2$, $p < .05$ ($MS_e = 401{,}341$); $F_2(2, 120) = 40.2$, $p < .0005$ ($MS_e = 42{,}188$). The average differential scores for short, medium, and long SOAs were 292, 266, and 212 ms, respectively, suggesting decreasing interference between naming task and lexical decision task over SOAs. This interpretation was confirmed by an independent ANOVA of the control items, the 10 items that all subjects had in common. For these items, picture and acoustic test probe were always unrelated, and the invited lexical decision responses were always positive. Here, also, a significant SOA effect was found, $F_1(2, 180) = 3.37$, $p < .05$ ($MS_e = 171{,}988$), and $F_2(2, 180) = 9.14$, $p < .005$ ($MS_e = 63{,}501$). For the control items, the mean differential scores were 224, 205, and 165 ms for the short, medium, and long SOAs, respectively.

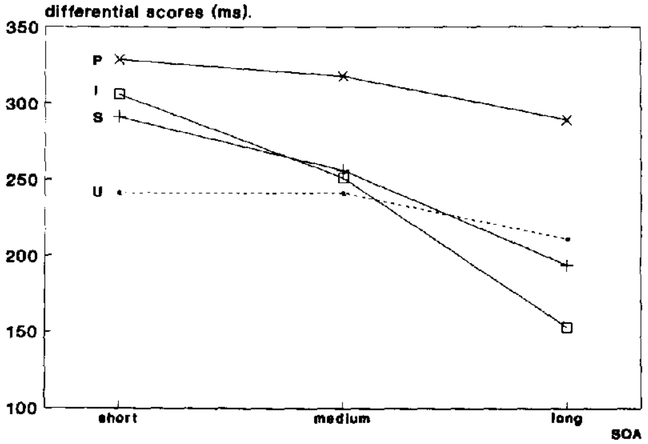

Figure 3. Lexical decision latencies in Experiment 3. (Mean differential scores for phonological [P], identical [I], semantic [S], and unrelated [U] acoustic test probes at three stimulus-onset asynchronies.)

However, the sloping effect was not uniform for the different kinds of test probe. Test-probe conditions (S, P, I, and U) showed a significant overall effect, $F_1(3, 540) = 15.5$, $p < .0005$ ($MS_e = 70{,}153$); $F_2(3, 60) = 7.2$, $p < .001$ ($MS_e = 151{,}040$). In addition, there was a significant interaction of SOAs and test probes, $F_1(6, 540) = 3.0$, $p < .01$ ($MS_e = 70{,}153$); $F_2(6, 120) = 5.0$, $p < .0005$ ($MS_e = 42{,}188$). This means that SOA curves for the four test-probe conditions had significantly different shapes, to which we return shortly. Groups within SOAs was not a significant variable, $F_1(3, 180) < 1$ ($MS_e = 401{,}341$), nor was the interaction between SOAs and groups: $F_1(6, 180) < 1$ ($MS_e = 401{,}341$). However, there was a significant Groups × Test Probe interaction, $F_1(9, 540) = 3.6$, $p < .0005$ ($MS_e = 70{,}153$). But the triple interaction Group × Test Probe × SOA was not significant, $F_1(18, 540) < 1$ ($MS_e = 70{,}153$). This means that the course of test-probe conditions over SOAs is independent of groups. And that is how it should be.

We now further examine the differences between test-probe conditions at different SOAs. We analyzed simple effects on the basis of the F value for the interaction of the test probe and SOA factors. For each SOA, there was a significant effect of test-probe conditions for both the subject and item analyses. Most relevant for the interpretation of these effects are the differences among, on the one hand, the S, P, and I conditions for a particular SOA and, on the other hand, the U condition for that SOA. The unrelated-word condition can be considered a baseline for the evaluation of the S, P, and I probes. The issue is whether S, P, and I probes show lexical decision effects that are different from the U probes for the same pictures. To find out, we applied Newman-Keuls paired-comparisons tests (with $p < .05$) within each of the three SOA conditions.

For the short SOA, S, P, and I conditions all differed significantly from the U condition for both the subject and item analyses. Lexical decisions were, on average, 50 ms slower to S probes than to U probes. P-probe decision latencies were 88 ms slower than U-probe latencies, and I probes were 65 ms slower. There were, however, no significant differences between S, P, and I conditions.

Also, for the medium SOA condition there was full agreement between subject and item analyses: Only the P-probe latencies were significantly slower than the U-probe latencies (by 77 ms); they were also significantly slower than the S- and I-probe latencies. There were no significant differences between S, I, and U.

For the long SOA condition, the P-probe latencies were significantly slower than the U-probe latencies, by 78 ms (and also significantly slower than both S and I). The I-probe latencies were significantly faster than the U probes, by 58 ms. However, S and U did not differ significantly. On these points, there was full agreement between the subject and item analyses. There was one more significant difference in the item analysis only, namely between S- and I-probe latencies, but this is irrelevant for our argument.

In summary, we found significant S, P, and I effects in the short SOA condition, a significant P effect in the medium SOA condition, and significant but opposite P and I effects in the long SOA condition.

Discussion

What do these data tell us about the two-stage model and the two activation-spreading models? All three models agree in predicting semantic activation shortly after picture presentation, and that is what was found: a significant effect for the semantic probes at the short SOA. Furthermore, all three models predict phonological activation at long SOAs, and that also was found: Lexical decisions to P probes are significantly slower than are those to U probes at long SOAs. However, other findings seem to be problematic for the two-stage and the backward-spreading connectionist models. Contrary to the prediction of the two-stage model (see Figure 1b), there is evidence for early phonological activation. And contrary to the backward-spreading connectionist model (see Figure 2c), there is no evidence for late semantic activation. Hence, for the time being, the forward-only activation-spreading model seems to be the only one to survive without damage.

Still, we argue that the evidence is insufficient for drawing this conclusion. Although each individual naming token proceeds through discrete stages, perhaps the distribution over tokens does not. What is needed is a statistical model of how the naming task interferes with the lexical decision task, a model that predicts the mean lexical decision latencies from assumptions about stages in naming and in lexical decision, and their phonological or semantic interference over SOA. Such a model is presented later in this article. It shows that the present data are compatible with the two-stage view. Furthermore, the model accounts for the striking I-probe results: At short SOAs, lexical decisions to identical probes are relatively slow, but at long SOAs, they are relatively fast (compared with U probes). Later, we discuss these results in connection with the model to be presented. In the General Discussion section, we reconsider the relation between these data and the feedback model, in particular the significance of the identical probe data.

Before turning to the issue of phonological activation of semantic alternatives, we report on one control experiment. It concerned the semantic effect obtained in the short SOA condition.

Experiment 4: Recognition Versus Lexical Access

The interpretation of the semantic interference results obtained in the previous experiment needs further scrutiny. The idea of the two-stage model is that the semantic, selectional stage sets in after recognition of the object. There is, however, the possibility that these processes are not distinct. The recognition of an object may not really be distinguishable from the semantic activation of a corresponding lexical item. Alternatively, the obtained semantic interference effect in the short SOA may in fact be due to the picture recognition process alone; it may have nothing to do with lexical access (see Levelt, 1989, for further discussion).

In order to determine whether the semantic interference effect can be attributed to picture recognition, rather than to lexical access, we conducted a control experiment in which the subject's task was one of recognition memory. The subject's task was to decide whether he or she had seen the picture before; no naming response was required. This task requires that the subject recognize the picture. If the same semantic interference effect were to arise as in lexical decision, its source should be picture recognition, not lexical access. If, however, the semantic interference effect were to disappear, it apparently would be dependent on the preparation of a naming response; that is, it would be a real lexical effect. In the latter case we would, moreover, have evidence that recognition and lexical selection should be distinguished in the process of picture naming.

Method

Stimuli. The same visual and acoustic stimuli are used as in Experiment 3.

Procedure. The experiment differed from the previous one in two major respects. First, the subject's task was not to name the pictures presented, but to recognize them. To make this task possible, the main session of the experiment consisted of two phases. During the first phase, a set of 92 slides were presented, one by one. Each slide was shown for 1.5 s, followed by a 2-s pause. The subject was instructed to inspect each object carefully because he or she would be asked to recognize these objects in the subsequent phase of the experiment. During the second phase, the 190 slides of Experiment 3 were presented to the subjects, and these included the 92 slides that had appeared during the first phase. The subject was asked to say *oud* ["old"] when shown a previously presented object and *nieuw* ["new"] in the other cases. Also, as in Experiment 3, the subject was told that occasionally an acoustic stimulus would be presented to which a pushbutton lexical decision response had to be given.

Exactly the same combinations of pictures and words were used as in the previous experiment. This implies that we had, again, four groups of subjects in a Latin square design. We made sure that the 16 critical target pictures (the same ones used in Experiment 3) should be given a "new" response. The 92 "old" items consisted of 15 of the 30 practice trials, 13 of the 26 items that were combined with a nonword lexical

Table 3

Lexical Decision Latencies, Differential Scores, and Percentage of Missing Values for Test Probes Presented During a Picture Recognition Task

	Test probe			
	Identical	Semantic	Phonological	Unrelated
Presession	750	748	801	692
Main session	1,113	1,027	1,098	947
Differential	364	279	297	256
% missing	10	6	18	6

Note. The average stimulus-onset asynchrony was 73 ms.

decision probe, 54 of the 108 slides that were not combined with an acoustic test probe, and the 10 control items of the previous experiment. The 92 slides in Phase 1 were presented in the same random order to all subjects. In Phase 2, the randomization was the same as in Experiment 3.

About 1 week before the main session, the subjects participated in a presession in which they were asked to do a lexical decision task on all acoustic test probes that would appear in the main session. The presession was identical to the one of Experiment 3.

The second difference with Experiment 3 was that only one SOA was used, namely the "short" condition (with an average SOA of 73 ms). This modification was made because our main objective in this experiment was to establish whether the semantic effect of Experiment 3 would reappear under a recognition task. That effect had appeared under the short SOA condition only.

Subjects. Seventy-one subjects from the Max Planck Institute subject pool participated in this experiment; they were mostly undergraduate students of Nijmegen University. All subjects participated in both the presession and the main session of the experiment and received Dfl. 17 for their participation. Sixty-four subjects were included in the final analysis, following the same criteria as used in Experiment 3, except that the correct naming criterion of that experiment was not applicable here. It was replaced by a correct recognition criterion.

Results

The mean lexical decision reaction times for the four categories of critical test probes (S, P, I, and U probes) are presented in Table 3, for both the presession (lexical decision without naming) and the main session of the experiment. They were derived from the differential scores in the same way as for Experiment 3. That is, the means were based on just those presession and main session data for which there were no missing values in the differential scores. The differential scores themselves are also presented. Whereas in Experiment 3 an incorrect naming response resulted in a missing value, here an incorrect recognition response was treated as a missing value. (Occasionally a subject erroneously gave an "old" response to a critical picture even though these items had not been included in the Phase 1 presentation of 92 slides.)

The subject ANOVA showed no significant differences among the four groups of 16 subjects, $F_1(3, 60) = 1.2$, $p = .32$ ($MS_e = 321{,}881$). There was a significant effect of test probes, $F_1(3, 180) = 8.8$, $p < .0001$ ($MS_e = 62{,}282$); $F_2(3, 60) = 6.92$, $p < .001$ ($MS_e = 79{,}303$). There was no significant Group × Test Probe interaction, $F_1(9, 180) = 1.6$, $p = .12$.

In regard to the significant effect of test-probe conditions, a Newman-Keuls test (with $p < .05$) revealed the same pattern for both the subject and the item analyses. There were three significant differences, namely those between the I condition and the three other kinds of probe (S, P, and U). The difference between I and U probes means that we obtained a significant effect of picture recognition on the decision latency for an acoustic probe that was the name of the presented object. The effect is one of interference and amounted to 108 ms.

However, the difference between the S and U conditions was only 23 ms and not significant. In other words, there was no significant effect of picture recognition on decision latencies for S probes. The nonsignificant 41-ms difference between P and U probes was almost solely due to relatively long decision latencies for P test probes in one of the four groups. Because the Groups × Test Probe interaction was also insignificant, however, we do not speculate on the cause of this outlier.

Discussion

The main purpose of Experiment 4 was to investigate whether the 50-ms semantic effect obtained in Experiment 3 might be due to object identification instead of to lexical selection. If object identification were the cause, a similar semantic effect should have been obtained in a recognition task where no naming response is required. Such an effect was not obtained, however, and we conclude that a condition for the semantic effect to arise is that a naming response be given; in other words, the semantic effect has to do with accessing the lexical item, not with identifying the object.

A final remark should be made on the strong (and unexpected) interference effect obtained for the I probes. A preliminary point is that this result shows our recognition task to be sensitive enough to produce results. If no effect had been obtained in this experiment, one could have dismissed the negative semantic result as being due to insensitivity of the procedure. That argument can now be put aside.

How can this interference effect for the I probes be explained? The I condition is special in that the picture matches the word's meaning (it is known that there is semantic access in lexical decision). However, the subject should ignore this match and instead judge the potential match between the picture and the items in the recognition set. This ignoring of a given match in favor of another apparently requires additional processing resources. One could argue that a similar state of affairs should arise for the semantic probes: In those cases, there is also a semantic match that should be ignored. But here we only found a nonsignificant 23-ms effect. That is, however, what one should expect if our conjecture is correct. The semantic relation in this experiment is one of word–word association, not one of picture–word association. If the picture's name is not (or rarely) retrieved during the recognition task, the word–word match will not become apparent when an S probe is presented. However, the picture–word match in the I condition will become apparent even when the picture's name is not retrieved. This reasoning presupposes that word associations are indeed associations between words, rather than between concepts. Levelt

(1989) has argued for this claim on the basis of speech errors. Word associates appear in word substitutions but hardly ever in blends, which are typically due to the activation of closely related concepts. If word associations are due to conceptual priming, one would expect to find blends of associated words.

SEMANTIC ACTIVATION WITHOUT PHONOLOGICAL ACTIVATION

In the introduction, the issue is raised of whether phonological encoding is restricted to selected items only or whether any activated item will, to some extent, become phonologically active. This is, in fact, a main empirical distinction between the two-stage theory and the activation-spreading theories. The former predicts no phonological activation of nonselected items (see Figure 1b), whereas the latter do predict such activation (see Figure 2b, c). In this section, two experiments are reported that were designed to decide this issue. Experiment 5 was conducted to check whether semantic associates of the target name are phonologically active. So, if the target name is *sheep,* is there evidence for the phonological activation of *wool*? Experiment 6 was designed to test whether semantic alternatives to the target name show phonological activation. If the target is *sheep,* will there be phonological activation of *goat*?

Experiment 5: Phonological Activation of Semantic Associates

In this experiment, we tested whether, during preparation of a naming response, not only the target but also close semantic associates are phonologically activated. It is known from speech-error research that associates can substitute for target items, such as in *don't burn your toes* for *don't burn your fingers* (see Levelt, 1989, for a review). Apparently, associates can become coactivated with the target. The substantial lexical decision effect for associate probes in Experiment 3 may have been caused by this coactivation. However, will these coactivated nontarget items also become phonologically active? The activation-spreading theories predict that this will be the case, whereas the two-stage theory predicts that it will not.

Method

The experiment was quite similar to the short SOA condition of Experiment 3. There was one difference: The phonological lexical decision probes were replaced by probes that were phonologically related not to the target names but to their semantic associates. For instance, if the target is *sheep* and the semantic associate *wool,* the phonological test probe would be *wood.* If *wool* is both semantically and phonologically activated when *sheep* is the target, there should be a lexical decision effect for both the *wool* and the *wood* probes. That the former effect appears we already know from Experiment 3; occurrence of the latter effect was at issue in this experiment. We chose the short SOA condition for two reasons. First, Experiment 3 had shown a significant effect for associate test probes (the S condition) for that SOA condition only. Second, of all three SOA conditions, the short one had produced the strongest phonological effect.

Stimuli. The visual stimuli were the same as in Experiment 3. The acoustic lexical decision items were also the same (see Table 1), except that the phonologically similar test probes were replaced by probes that were phonologically similar to the semantic test probes (i.e., *wood* instead of *sheet* when the target name was *sheep*). The 16 replacing phonological items are given in the Appendix. They were carefully selected to be phonologically close to the semantic test probes but were semantically unrelated to the latter as well as to the target names. Although all other acoustic probes were the same as in Experiment 3, they were all newly recorded (with the same female speaker as in Experiment 3). This was to prevent the new phonological items from, for accidental reasons, sounding different from the other items.

Table 4

Lexical Decision Latencies, Differential Scores, and Percentage of Missing Values for Test Probes Presented During Picture Naming

	Test probe			
	Identical	Semantic	Phonological	Unrelated
Presession	760	744	808	732
Main session	1,068	1,006	1,026	937
Differential	307	262	218	206
% missing	8	11	14	7

Note. The phonological test words are phonologically related to the semantic test words, which are associates. Average stimulus-onset asynchrony was 73 ms.

Procedure. The procedure was in all respects identical to Experiment 3, with presession and main session short SOA.

Subjects. Eighty subjects participated in this experiment. Sixteen subjects who did not meet the strict selection criteria of Experiment 3 were excluded from the final analysis, leaving us with 64 subjects, four groups of 16. Subjects were taken from the Max Planck Institute subject pool and paid Dfl. 17 for their participation in the two sessions of the experiment.

Results

The results of the presession (where there was no naming task, but only lexical decision) and the main session, as well as the mean differential scores, are shown in Table 4. The treatment of missing data was as in the previous two experiments. The subject and item analyses of variance were run on these differential scores, with test probes as a within-subject variable and groups as a between-subjects variable. The subject analysis showed that the four groups of subjects did not differ significantly, $F_1(3, 60) = 2.55$, $p = .06$ ($MS_e = 246{,}043$). There was a significant effect of test probes, $F_1(3, 180) = 11.55$, $p < .0005$ ($MS_e = 46{,}810$), and $F_2(3, 60) = 5.3$, $p < .005$ ($MS_e = 101{,}850$), and a significant interaction between groups and test probes, $F_1(9, 180) = 2.03$, $p < .05$ ($MS_e = 46{,}810$). This latter result needs further scrutiny, and we return to it later.

The test-probes effect was further analyzed by means of a Newman-Keuls test (with $p < .05$). As in Experiment 3, the S and I conditions differed significantly from the U condition (by 56 and 101 ms, respectively), although the S condition differed only on the subject analysis. However, the new P condition only differed by an insignificant 12 ms from the U condition on both analyses. In addition, the I condition was significantly slower than both the S and P conditions, and the latter two also differed significantly, but only on the subject analysis.

What about the interaction between groups and test probes?

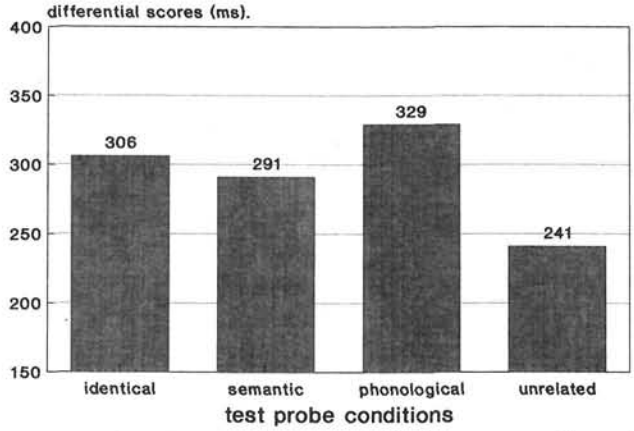

Figure 4. Differential scores for the short stimulus-onset asynchrony of Experiment 3. (Test-probe conditions: identical to target, semantic associate of target, phonologically related to target, and unrelated to target. Average stimulus-onset asynchrony was 73 ms.)

Is it the case that within certain groups of subjects, the P and U conditions differed significantly? (Remember that each group received a different set of P test probes.) A Newman-Keuls test (with $p < .05$; of course limited to the subject analysis) revealed that for none of the groups was the difference between the P and U test probes significant. The interaction was entirely due to group differences for the I test probes: Two of the groups showed large and significant differences between I and U test probes, whereas the other two groups showed smaller, insignificant I-probe effects. In other words, the obtained phonological null effect was homogeneous.

Discussion

The experiment replicated the semantic and identical findings of Experiment 3. This is easily seen by comparing Figures 4 and 5. Figure 4 presents the mean differential scores obtained for SOA = 73 ms in Experiment 3. Figure 5 presents the corresponding data for the present experiment, in which the same SOA was used. The I, S, and U bars display the same pattern in the two figures. However, the strong phonological effect that was found in Experiment 3 (i.e., for cases such as target name *sheep* and test probe *sheet*) could not be measured for semantic associates in Experiment 5 (e.g., target name *sheep,* semantic associate *wool,* and P test probe *wood*).

However, one could argue that semantic associates are often not really semantic alternatives to the target names (although speech errors of this kind are occasionally made). Figure 2a depicts the situation where the concept to be expressed activates a semantic cohort, again, a set of items that are sensitive to semantic aspects of the concept. So, if the concept is *sheep,* the semantic cohort might consist of other domestic farm animals, such as *goat* or *cow.* These are semantic alternatives to *sheep,* but not necessarily semantic associates. Hence, to find out whether the situation depicted in Figure 2a, namely the phonological activation of all items in the semantic cohort, is realistic, the phonological activation of semantic alternatives should be studied. This was done in the next experiment.

Experiment 6: Phonological Activation of Same-Category Items

The main purpose of this experiment was to check whether same-category semantic alternatives to the target (e.g., *goat* when the target is *sheep*) would become phonologically activated. Also, we wanted to check whether such items were at all semantically activated upon presentation of the picture. Failure to obtain a phonological effect might after all be adduced to semantic nonactivation of the items.

Method

The experiment was similar to Experiment 5; it differed only in the choice of S and P probes.

Stimuli. The visual stimuli were all the same as in Experiment 5; the 16 critical target pictures were the ones listed in Table 1. There were again four acoustic test-probe conditions: I, S, P, and U. Conditions I and U were the same as before, but S and P differed. The S condition consisted of test probes that were semantic alternatives to the target word, that is, to the picture's name. So, if the picture was one of a sheep, the corresponding S item was *goat* (it was *wool* in the previous experiment). The P condition was made up of phonological probes for these semantic alternatives. So, if the picture displayed a sheep, and *goat* was an S probe, then *goal* would be a P probe. The S and P probes for the 16 target pictures are presented in the Appendix. All acoustic stimuli in this experiment were newly recorded by the same female speaker who had contributed to the previous experiments.

Procedure. The procedure was in all respects identical to the procedure of Experiment 5. There was a presession (lexical decision only) and about 1 week later a main session with short SOA (i.e., with an average SOA of 73 ms). As in the previous experiment, we used the short SOA condition because that is where we found the strongest phonological effect in Experiment 3. We also wondered whether test probes that are semantic alternatives would show an effect for the same SOA condition as used in Experiment 3, where test probes that are associates produced an effect.

Subjects. Sixty-four subjects took part in this experiment. Sixteen subjects who did not meet the strict selection criteria of Experiment 3 were replaced by new subjects. Subjects were selected from the Max

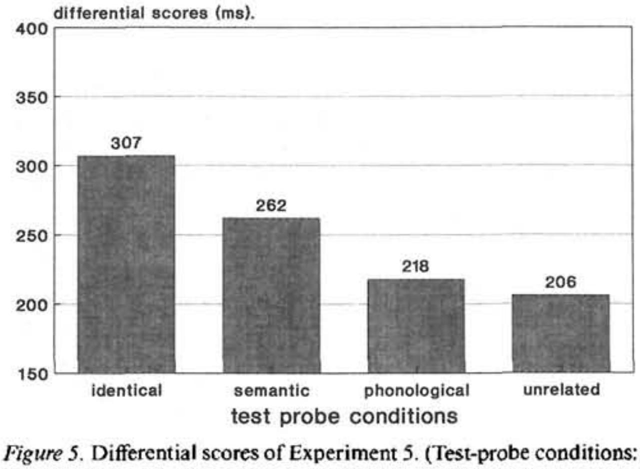

Figure 5. Differential scores of Experiment 5. (Test-probe conditions: identical to target, semantic associate of target, phonologically related to associate, and unrelated to target. Average stimulus-onset asynchrony was 73 ms.)

Table 5
Lexical Decision Latencies, Differential Scores, and Percentage of Missing Values for Test Probes Presented During Picture Naming

	Test probe			
	Identical	Semantic	Phonological	Unrelated
Presession	721	769	785	714
Main session	1,080	1,131	1,039	969
Differential	360	362	254	256
% missing	9	20	21	9

Note. The semantic test words are semantic alternatives to the identical words. The phonological test words are phonologically related to the semantic words. Average stimulus-onset asynchrony was 73 ms.

Planck Institute subject pool and were paid Dfl. 17 for their participation.

Results

The results for the presession and main sessions of the experiment, as well as the differential scores, are presented in Table 5. As in the previous experiments, the presession and main session data are those for which there were differential scores. The differential scores are also presented in Figure 6, which can be compared with Figures 4 and 5. The ANOVA was, as always, based on the differential scores. The subject ANOVA, with test probes as the within-subject variable and groups as the between-subjects variable, showed a significant effect of test probes, $F_1(3, 180) = 14.59$, $p < .0001$ ($MS_e = 65,641$), and $F_2(3, 60) = 5.95$, $p < .01$ ($MS_e = 160,922$). This effect was further analyzed by means of a Newman-Keuls test (with $p < .05$), which gave exactly the same result on subject and item analyses: The U condition differed significantly from the I and S conditions, with no difference between the latter two conditions. Also, the P condition differed significantly from the I and the S conditions. And, most important for this experiment, the U and the P conditions did not differ. In fact, the P-probe latencies were, on average, 2 ms faster than the U-probe latencies.

The subject ANOVA further showed a significant effect of groups, $F_1(3, 60) = 5.75$, $p < .01$ ($MS_e = 290,358$), and a significant interaction between groups and test probes, $F_1(9, 180) = 2.13$, $p < .05$ ($MS_e = 65,641$). A Newman-Keuls analysis of this interaction revealed that it was mainly due to the fact that in one group of subjects the four conditions did not differ significantly from each other. However, none of the four groups showed a significant difference between the U and P conditions; the phonological null effect was homogeneous.[2]

Discussion

The main purpose of this experiment was to test whether a target's semantic alternatives would show phonological activation during preparation of the naming response. We could not find evidence that this is the case. Figure 6 shows that the differential scores for the phonological condition did not even reach the level of the U condition. This should be compared to a strong phonological effect for the target name itself, obtained in Experiment 3 (Figure 4). There, the difference between the P and U results was a highly significant 88 ms (differential scores) for the short SOA.

At the same time, the absence of a phonological effect cannot be attributed to a poor choice of semantic alternatives. Figure 6 and the statistical analysis indicate a strong semantic effect. In fact, Figure 6 shows that the interference effect for the semantic alternatives was of the same size as the interference effect for the test probes that were identical to the target. This finding supports the notion of a semantic cohort.

One might argue that the lack of a phonological effect is due to the choice of a short SOA (average 73 ms) in our experiment. As we said, we made this choice because in Experiment 3 we found both semantic and phonological priming effects for this SOA. However, would mediated priming not take more time? The answer is that the activation-spreading theory that we tested does not involve mediated priming. The concept activates both lexical items simultaneously, and both lexical nodes then spread their activation to the phonological level. Hence, the chosen SOA, for which there is demonstrable phonological activation of the target item, is a fair choice for testing that theory's prediction concerning the phonological activation of semantic alternatives.

The conclusion from Experiments 5 and 6 is that we have not been able to obtain evidence for the phonological activation of items that are semantically related to the target word. Neither close associates nor semantic alternatives appear to become phonologically activated together with the activated target word. This null result supports the two-stage model[3] but is problematic for the activation-spreading theories. It should at least limit the amount of phonological coactivation that one allows in connectionist models of lexical access. We return to this issue in the General Discussion section.

This brings us back to our record. We set out distinguishing three models of lexical access. The prediction from the activation-spreading models that semantically activated items should also become phonologically activated could not be substantiated. In addition, we found (in Experiment 3) no evidence for late semantic activation of the target, predicted by the backward-spreading model. However, the two-stage model also became suspect, because in Experiment 3 we found evidence for

---

[2] The finding of no difference between the U and P conditions was replicated in an additional experiment. There, only these two conditions were realized with a different set of pictures and test probes. One group of 16 subjects saw half of the 16 pictures paired with acoustic test words in the P condition, the other half with test words in the U condition. For the other group of 16 subjects, assignment of pictures to test-probe conditions was reversed. There was neither an effect of test probes nor an interaction between groups and test probes.

[3] McNamara and Healy (1988) found a similar null result in mediated priming. In both lexical decision and reading tasks, there were no mediating effects between words that formed a semantic–phonological chain, such as *queen–(king)–sing*. The authors were careful not to use these findings as strong support for a two-stage theory of production. Also, their phonological relations were always of the rhyming type. There is good evidence, however, that rhymes cause no phonological priming (Meyer, 1990), whereas word-initial similarity does.

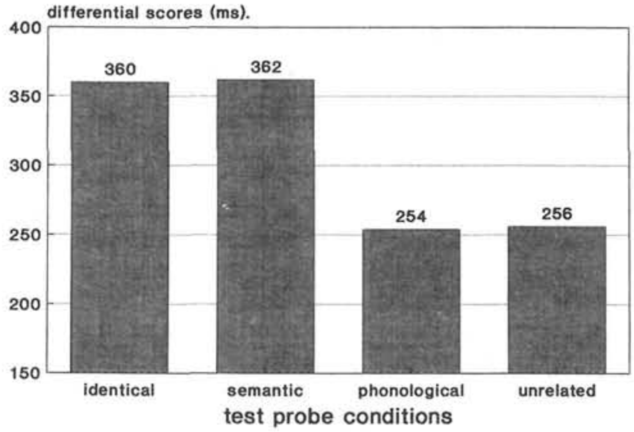

Figure 6. Differential scores of Experiment 6. (Test-probe conditions: identical to target, semantic alternative to target, phonologically related to semantic alternative, and unrelated. Average stimulus-onset asynchrony was 73 ms.)

early phonological activation, coinciding with semantic activation. In the next section, we show that this suspicion is unfounded.

A MODEL OF THE DUAL NAMING–LEXICAL DECISION TASK

In this section, we present a mathematical formulation of the discrete two-stage model and of the dual naming–lexical decision task. The aim is to find out whether the data of Experiment 3 are, under reasonable assumptions about the interaction between naming and lexical decision, compatible with the two-stage model.

The experiments so far have shown that the process of acoustic lexical decision is sensitive to semantic and phonological aspects of the naming response. To model these effects, we assume that to produce a lexical decision response, the subject will proceed through (at least) two stages. The first, phonological stage involves the selection of the lexical item, given the acoustic stimulus information. The second, semantic stage consists of retrieving the word's meaning and coming to a decision. In the model, the assumption is that these two phases are successive. This assumption may be stronger than necessary because work by Zwitserlood (1989) and others has shown that in word recognition, phonological and semantic activation overlap in time. The simplification is, however, innocent, because succession is the limiting case of overlap. If the data can be fitted in the limiting case, they can certainly be fitted in the general case (which has an additional overlap parameter).[4]

The naming process is formalized in accordance with the discrete two-stage model. After perceiving the picture, the subject first proceeds through a semantic stage, leading to selection of the target lemma. Subsequently, a phonological stage is entered, during which the articulatory plan is prepared. We will assume that the ensuing articulatory plan is delivered to an articulatory buffer, from which it is retrieved to execute the naming response (cf. Levelt, 1989). The two stages of lexical access are therefore followed by a buffering stage.

The starting point of the model is, therefore, that there are two relevant stages in lexical decision, P(phonological) → S(semantic), and three relevant stages in naming, S(semantic) → P(phonological) → A(articulatory buffering).

We further assume that the duration of each stage has an exponential probability density function, $f_i(t) = r_i \cdot e^{-r_i t}$. Here, r_i is a rate parameter for stage i, and t is the time in milliseconds from entering the processing stage. The mean duration of a stage is $1/r_i$ ms. Each of the five stages mentioned earlier has its own characteristic rate; they are free parameters in the model. One might object that the exponential distribution is rather different in shape from experimentally obtained reaction time distributions. However, sums of two or more independent exponentially distributed random variables are gamma distributed; therefore, the model predicts total reaction time to follow the gamma rather than the exponential distribution. The general gamma distribution often provides a good fit to empirical reaction time distributions (McGill & Gibbon, 1965). Our model construction follows Vorberg (1990), to which we refer for mathematical details.

How does the preparation of the naming response affect the preparation of the lexical decision response? The following informal considerations are relevant here. Both the phonological and the semantic analysis of the lexical decision item can be affected by the preparation of the naming response.

Let us first consider the phonological analysis in the lexical decision channel. It will, we assume, be complicated if some partial phonological representation in the naming channel boosts phonological competitors to the lexical decision probe. A partial phonological representation is present during the phase of phonological encoding in the naming channel, and that representation will boost phonological competitors if it is compatible with their phonological representations. This will be the case when the lexical decision probe is phonologically similar to the picture name, that is, in case there is a P probe. Partial representation will also be present when the probe is identical to the picture name. In the latter case, a partial phonological representation of the target name can still be compatible with a set of phonological competitors to the lexical decision probe word. If, for instance, the word-initial phonemes of the picture name *sheep*, /ʃi/, have become available, and the lexical decision item is *sheep*, then not only will *sheep* be boosted in activation but also the competitors *sheet*, *sheath*, and so on. The situation is quite different, however, after completion of phonological encoding in the naming channel. In that case, the completed phonological representation will boost

---

[4] There is a theoretical issue, though: Why would there be no temporal overlap between semantic and phonological processing in naming, whereas there is some overlap in word recognition? One reason to expect asymmetry here is that, in general, the speech signal has a one-to-many relation to its semantic interpretation. In connected speech, the speech signal (as a rule) underdetermines the identity of the word; semantic–syntactic context is often essential for the accurate segmentation of the speech signal and the correct identification of the word. In production, however, a word is typically selected on semantic grounds, and these are sufficient. The further phonological processing can proceed without reference to semantic context.

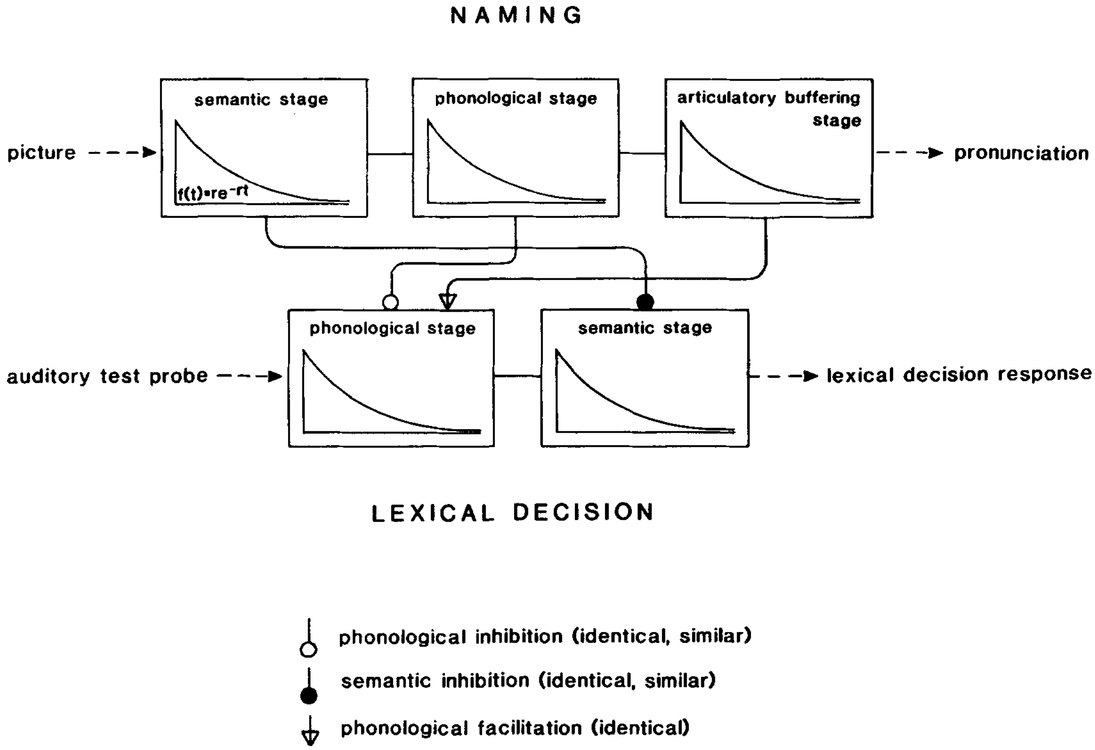

Figure 7. A mathematical model of the interaction between naming and lexical decision, based on the two-stage theory.

only the I probe word in the lexical decision channel, but not its competitors (with which it is not compatible). In other words, as long as the target name is in the articulatory buffer, it will facilitate the lexical decision response to the identical probe.

Turning now to the semantic analysis in the lexical decision channel, we assume that it can be affected by the presence of an active semantically related item in the naming channel. More specifically, we assume that a semantically active item in the naming channel will interfere with the lexical decision for a meaning-related probe word. The interference is of a Stroop-like character: There will be a tendency in the subject to react to the naming target instead of to the lexical decision probe, and that tendency has to be inhibited. The size of that tendency will, as in the Stroop situations, be a function of the *semantic gradient,* the semantic closeness of picture name and lexical decision probe (Glaser & Glaser, 1989). Note that we are extrapolating this (well-established) closeness effect to the identical case as well. This is by stipulation, because there is nothing in the empirical literature to either support or contradict this assumption.

In the following paragraphs, these considerations will be formalized. Our assumptions are depicted in Figure 7. The figure shows the three relevant stages of naming and the two relevant stages of lexical decision with their exponential density distributions. Dependent on SOA and on the factual stage durations during an experimental trial, there can be different kinds and degrees of temporal overlap between the stages of naming and the stages of lexical decision. Three cases are pertinent: (a) There is temporal overlap of the semantic stages of naming and lexical decision, (b) there is temporal overlap of the phonological stages of naming and lexical decision, and (c) there is temporal overlap of the articulatory buffering stage of naming and the phonological stage of lexical decision. Notice that there will never be overlap of both the semantic and the phonological stages.

We now assume that the naming process can only affect the lexical decision process if in any of these three cases the materials being processed are similar or identical. More specifically, the following assumptions are made:

Case a. If the target word in naming is semantically similar or identical to the lexical decision item, and the semantic stages overlap, the rate of the semantic stage in lexical decision is reduced for as long as the overlap of stages lasts. Reduction of the rate means a larger mean duration of the semantic stage in lexical decision; that is, there will be semantic interference. The amount of semantic rate reduction is a free parameter in the model.

Case b. If the target word in naming is phonologically similar or identical to the lexical decision item, and the phonological stages overlap, the rate of the phonological stage in lexical decision is reduced for as long as the overlap of stages lasts. In other words, there will be phonological interference. The amount of phonological rate reduction is a free parameter in the model.

Case c. If the target word in naming is identical to the lexical decision item and the articulatory buffering stage overlaps with the phonological stage of lexical decision, the rate of phonological processing increases. There will be phonological facilitation. The amount of facilitation is a free parameter in the model.

No interactions are assumed in the model other than these three cases.

These are the essential model assumptions. They involve eight free parameters: five for the stage durations (rates) and three for the reduction or facilitation of rates. However, we had to make some additional nonspecific assumptions. First, the four kinds of test items, I, S, P, and U, had different lexical decision latencies in the presession of the experiment. This had to do with the choice of the items. The phonological probes were in particular somewhat slower than the other three kinds of test probe. This was clearly due to the lower word frequency of these probes. (It was not always possible to find a test probe that was phonologically similar to the target and of the same frequency range). These differences are probably irrelevant as far as differential scores are concerned (on which our analyses were all based), but the model had to fit the actual reaction times. Hence, we replaced the one phonological rate parameter for lexical decision by four different parameters, one for each item type.

Furthermore, we had to deal with attentional effects. First, there was the general finding that lexical decisions were several hundred milliseconds slower in the main session of the experiments than in the presession. This is, obviously, an attentional effect. The double-task situation of the main session requires a basic distribution of attention that is different from the concentrated attention in the single-task situation of the presession. This required one additional task-dependent attentional parameter. Second, we assumed that this general parameter was modulated by the factual appearance of a lexical decision item. The attention would then, partially, switch to the lexical decision channel. The amount to which this happened was expressed in a stimulus-dependent attentional parameter. This, then, was the set of 13 parameters used to fit the presession and main session results for the three SOA conditions. These parameters were estimated by minimizing the discrepancy between obtained and predicted reaction time means. As a fit measure, we used the chi-square statistic suggested by the statistic C^2 discussed by Miller and Greeno (1978); the minimization routine followed a modified Davidon-Fletcher-Powell minimization algorithm (see Press, Flannery, Teukolsky, & Vetterling, 1986).

The results of the estimation procedure are depicted in Figure 8. Shown are the lexical decision latencies obtained in the main session of Experiment 3 (data from Table 2), the model estimates of these data, and the presession findings and simulations (note that there was no SOA variable in the presession). It is obvious that the fit is nearly perfect. The chi-square stress measure of fit was as small as 2.20 at 3 *df* (number of data points minus number of parameters). This means that the model predictions are statistically not different from the data.

The parameter estimates are given in Table 6. The inverse rates, $1/r$, are the average stage durations. The interaction parameters i stand for the reduction (if $i < 1$) or incrementation (if $i > 1$) of rates. The task-dependent attentional parameter t differentiates between the single-task and the dual-task situations; if it is 1, it distributes attention equally over the available channels, and there is no room for stimulus-dependent attentional variation; if it is 0, attentional variation is exclusively stimulus dependent. The stimulus-dependent attentional parameter s ranges from 0 (entirely the lexical decision channel) to 1 (entirely the naming channel).[5] Note that none of the parameters take on unrealistic values. So, for instance, the semantic, phonological, and buffering stages in naming during lexical decision were an estimated 115, 270, and 265 ms, respectively. The phonological and semantic–decision stages in lexical decision were estimated at 184 and 583 ms, respectively. These are estimates for the stage durations when attention is 100% and when there is no inhibition or facilitation from the other channel.

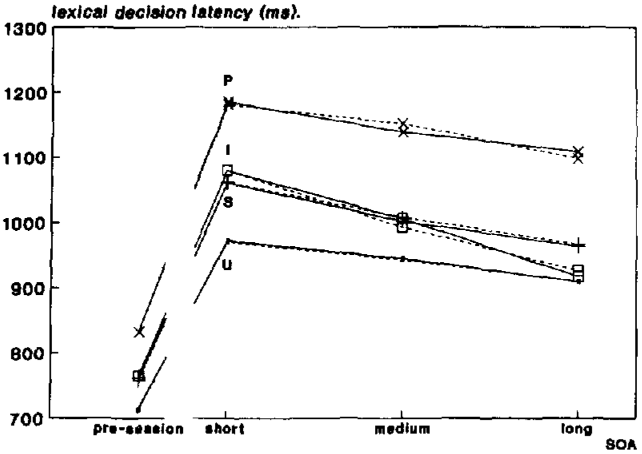

Figure 8. Model predictions of lexical decision latencies (presession and main session) in Experiment 3. (Solid lines indicate actual data; dotted lines indicate results from model. Test probes: P = phonological, I = identical, S= semantic, U = unrelated.)

It is satisfying that the model can handle the data for the I test probes, which are rather special. If the target word is presented as an acoustic test probe at a short SOA, it interferes with lexical decision. This is due to both the phonological boosting of phonological alternatives to the test probe by the picture name's partial phonological representation and the Stroop-like semantic interference discussed earlier. (The tendency to react to the target's name also has to be suppressed, we assumed, when target name and test probe are identical. At this early stage, the target name is not yet fully available, but its meaning is active enough to confuse the subject.) If the identical probe is presented at a long SOA, the lexical decision is faster than for the unrelated item, which indicates facilitation. This is due to the phonological boosting of the identical lexical decision probe by the then complete phonological representation of the identical picture name.

We may conclude that the mathematical implementation of the two-stage model, which makes reasonable assumptions about the interaction between the two tasks, is fully compatible with the results of Experiment 3. It shows that there is no reason to reject the two-stage model in the face of these results.

---

[5] The parameters t and s, together with the number of channels c that are active but not yet complete at any instant, jointly determine rate multiplication factors m_n and m_d, for the naming and the lexical decision channels, respectively, in the following fashion: $m_n = t/c + (1 - t)s$ and $m_d = t/c + (1 - t)(1 - s)$.

Table 6
Parameter Estimates Used for Fitting the Model of the Latencies in Experiment 3

	Estimate
Mean duration (1/*r*) in ms	
Naming	
Semantic stage	115
Phonological stage	270
Articulatory stage	265
Lexical decision	
Phonological stage	
Unrelated probes	131
Semantic probes	175
Phonological probes	249
Identical probes	182
Average	184
Semantic stage	583
Facilitation/inhibition (*i*)	
Semantic–naming → semantic–lexical decision	0.736
Phonological–naming → phonological–lexical decision	0.695
Articulatory–naming → phonological–lexical decision	7.485
Attention (*t*, *s*)	
Task dependent	0.376
Stimulus dependent	0.168

GENERAL DISCUSSION

Let us return to the theoretical alternatives depicted in Figures 1 and 2. The two-stage model (Figure 1) is based on the assumption that lexical selection strictly precedes phonological encoding; only the selected (target) item becomes phonologically encoded. The activation-spreading theories (Figure 2), on the other hand, allow for any semantically activated item to become phonologically active as well. There is, moreover, a temporal overlap of an item's semantic and phonological activation. We distinguished two kinds of activation-spreading accounts. There can be forward-only activation in the lexical network, or there can be backward-spreading activation as well, as is explicitly assumed in Dell's (1986, 1988, 1989) model.

The activation graphs in Figures 1 and 2 are schematic and qualitative only. Any more quantitative predictions are dependent on precise parameter estimations. There exists no quantitative version of the forward-only activation-spreading model. Only Dell's backward-spreading model is explicit enough to allow for quantification. The parameter estimations in Dell (1986), for instance, could simultaneously account for the effects of three independent variables on the relative frequencies of exchanges, anticipations, and perseverations obtained in a single experiment.

We set out to compare these theories by means of two kinds of experiment. The first one (Experiment 3) traced the time course of semantic and phonological activation of the target item in a dual naming–lexical decision task. The second kind (Experiments 5 and 6) checked whether items that are semantically related to the target also become phonologically activated.

We could show that the time-course data were fully compatible with at least one mathematical implementation of the two-stage theory; it yielded an almost perfect fit to the data. The obvious question is whether a similar good fit might be found for the activation-spreading theories.

As far as the phonological and semantic curves are concerned, this should not be a problem for the forward-only type of activation-spreading theory; the obtained curves are compatible with those in Figure 2b. However, we signaled a problem for the theories, such as Dell's (1986), which also involve backward spreading of activation. They predict a late rebound of semantic activation, but this was not supported in the data; there are no semantic test-probe effects for the medium and long SOAs. We return to this issue later.

Activation-spreading accounts, whether forward only or backward spreading, will also have to deal with the findings for I probes. The main result was interference at short SOAs and facilitation at long SOAs. These data are compatible with the two-stage model, given the assumptions outlined in the previous section. Could they also be made compatible with a connectionist picture of the dual task?

Neither of these problems may be unsolvable. As far as the semantic and phonological test-probe data are concerned, the question is whether a parameter estimation can be found that simultaneously satisfies two borderline conditions. The first one is that the feedback from the phonological to the lemma level is weak enough to prevent a measurable semantic reactivation of the target lemma. The second one is that the same feedback is still strong enough to explain the speech error phenomena for which it was proposed to start with. These are, in particular, the phenomena of lexical bias and of mixed errors, as well as the repeated phoneme effect. In other words, there is both an upper and a lower bound on the amount of feedback from the phonological to the lexical level. Can there be simultaneous satisfaction of both boundary conditions, or are they mutually exclusive? Only detailed modeling can answer this question, but the exercise is a complex one. One would need a set of experimental data like that of Experiment 3, but one would need to use materials that could at the same time figure in an experiment such as Dell's (1986), designed to measure the strength of lexical bias and other phenomena for which the backward spreading of activation was built into the model.

As far as the findings for the identical probe are concerned, special assumptions might be made (within the connectionist models) with respect to the interaction of the naming and lexical decision tasks. Such special assumptions are necessary anyhow (also in the two-stage model) to explain the inhibitory effects on lexical decision when target and test probe are semantically related or phonologically related.

So, in regard to the outcome of Experiment 3, only preliminary conclusions are possible. They are that the data are compatible with the two-stage model and, as far as the S and P test probes are concerned, also with the forward activation-spreading model. For the I test-probe data, special assumptions have to be made for both kinds of connectionist model. Whether the backward-spreading model can be reconciled with the data is undecided, the main problem being that the data give no indication of any semantic rebound at longer SOAs.

This brings us to some remarks on the raison d'être of the backward-spreading mechanism. In Dell's (1989) theory it serves to account for the lexical bias effect as well as other familiarity effects, and it can handle the causation of mixed errors. In addition, it provides an explanation for the repeated phoneme effect. Although this is an elegant unification, there is still good reason to explain the former familiarity effects in part or in whole by means of a different mechanism, namely a postlexical editor.

The arguments for assuming the existence of such a monitoring device outside the production apparatus proper are extensively discussed in Levelt (1989). One is that lexical bias is dependent on the speaker's attentional state. Instructions and experimental materials significantly determine what the speaker will be monitoring for, according to the experiments by Baars et al. (1975), Motley, Camden, and Baars (1982), and Bond and Small (1984). There may be a lexical bias, and there may be a bias for letting pass erroneous words of a particular semantic class (as occurs in mixed errors). Another is that the latency at which self-created errors are detected by the speaker is not so much determined by the character of the error as by external (postlexical) attentional factors. On the backward-spreading account, detection of self-produced errors is due to backward spreading from the erroneous node. If, for instance, *red* is the target node but *green* is erroneously selected, the concept node for GREEN will become activated through backward spreading. The speaker then detects that this differs from the intended concept RED (see especially MacKay, 1987, for such an account). This would predict not only that the detection of semantic errors is faster than the detection of phonological errors (the latter involve a larger backward trajectory through the network), but also that error detection will always be quite fast. However, the data support neither prediction. Rather, the speed of detection appears to be determined by where the error appears in the phrase. Phrase-final errors are detected much faster than phrase-internal errors (Levelt, 1983), suggesting a mechanism of selective attention: While constructing a phrase, the speaker's attention is occupied by conceptual planning, but toward the end of a phrase, attention can shift to the self-produced output. There are, in short, good reasons for considering alternative explanations for at least some of the phenomena that Dell's (1986, 1989) backward-spreading mechanism seeks to account for. That is particularly the case for the familiarity effects, such as lexical bias. If the model would (at least in part) be released of providing an explanation for these effects, it might become far easier to reconcile it with our findings in Experiment 3.

We now turn to the issue of the phonological activation of items that are semantically related to the target word. Both connectionist models predict such an effect, but Experiments 4 and 5 did not substantiate the prediction. Neither associates nor same-category items were phonologically active to a measurable extent. This null effect, however, was predicted by the two-stage theory. This finding concerns the most crucial difference between the two-stage and connectionist accounts. Will all semantically activated lexical items become phonologically active (to some extent) as network theories predict, or is it only the selected item that becomes phonologically encoded? Although our negative evidence is clearly supportive of the latter notion, one cannot a priori exclude the possibility that a connectionist account can be reconciled with this finding. One should choose the model's parameters in such a way that the phonological activation of the target becomes substantially stronger than the phonological activation of its semantic associates or competitors. In Dell's theory, this might be handled by boosting the current lexical (or lemma) node so much that its activation is of a different order of magnitude than the activation of semantically related items. This will make the phonological activation of related items negligible in comparison with the phonological activation of the target. However, there are limits to the discrepancy one can impose. A nonnegligible phonological activation of semantic alternatives is, for instance, necessary to handle the occurrence of mixed errors (such as *oyster* for *lobster*). The two-stage model can handle the latter kind of errors by means of the postlexical editing mechanism. Can one simultaneously satisfy both restrictions? To answer this question, experiments are necessary in which for the same materials phonological activation of semantic alternatives and the tendency for combined semantic–phonological errors can be measured. This is certainly not easily realized.

Our general conclusion from this study is this: The more traditional two-stage account of lexical access finds continuing support in the data. The theory says that there is an initial stage of lexical selection, followed by a stage of phonological encoding during which only the selected item becomes phonologically encoded. Further research, both empirical and theoretical, is needed to determine whether activation-spreading accounts can be reconciled with our negative findings on semantic rebounding and on the phonological activation of semantically coactivated lexical items.

References

Baars, B. J., Motley, M. T., & MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14,* 382–391.

Bond, Z. S., & Small, L. H. (1984). Detecting and correcting mispronounciations: A note on methodology. *Journal of Phonetics, 12,* 279–283.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Butterworth, B. (1980). Some constraints on models of language production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 423–459). London: Academic Press.

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 103–135). Cambridge, MA: MIT Press.

Dell, G. S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science, 9,* 3–23.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27,* 124–142.

Dell, G. (1989). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 136–165). Cambridge, MA: MIT Press.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611–629.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47,* 27–52. [Reprinted in Fromkin, V. A. (Ed.). (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.]

Fry, D. (1969). The linguistic evidence of speech errors. *BRNO Studies of English, 8,* 69–74.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam: North-Holland.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 177–220). London: Academic Press.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *The Cambridge survey of linguistics. Vol. 3: Biological and psychological aspects of language* (pp. 69–86). Cambridge, MA: Harvard University Press.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General 118,* 13–42.

Harley, T. A. (1984). A critique of top-down independent levels of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8,* 191–219.

Humphreys, G. W., Riddock, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5,* 67–104.

Jones, H. G. V., & Langford, S. (1987). Phonological blocking in the tip of the tongue state. *Cognition, 26,* 115–122.

Kempen, G. (1977). Conceptualizing and formulating in sentence production. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory* (pp. 259–274). Hillsdale, NJ: Erlbaum.

Kempen, G. (1978). Sentence construction by a psychologically plausible formulator. In R. Campbell & P. Smith (Eds.), *Recent advances in the psychology of language: Vol. 2. Formal and experimental approaches* (pp. 103–123). New York: Plenum Press.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the boundaries in linguistics. Studies presented to Manfred Bierwisch* (pp. 227–252). Dordrecht, The Netherlands: Reidel.

MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8,* 323–350.

MacKay, D. (1987). *The organization of perception and action. A theory for language and other cognitive skills.* New York: Springer.

Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for a model of lexical retrieval. *Journal of Memory and Language, 28,* 462–485.

McGill, W. J., & Gibbon, J. (1965). The general gamma distribution and reaction times. *Journal of Mathematical Psychology, 2,* 1–18.

McNamara, T. P., & Healy, A. F. (1988). Semantic, phonological, and mediated priming in reading and lexical decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 389–409.

Meyer, A. S. (1990). The time course of phonological encoding in language production: Phonological encoding inside the syllable. *Journal of Memory and Language, 29,* 524–545.

Meyer, A. S., & Bock, K. (1990). *Success and failure in lexical retrieval: Effects of semantic and phonological cues.* Unpublished manuscript.

Miller, J., & Greeno, J. G. (1978). Goodness-of-fit tests for models of latency and choice. *Journal of Mathematical Psychology, 17,* 1–13.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Morton, J. (1979). Word recognition. In J. Morton & J. Marshall (Eds.), *Psycholinguistics: Series 2. Structures and processes* (pp. 107–156). London: Elek.

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 21,* 578–594.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical recipes: The art of scientific computing.* Cambridge, England: Cambridge University Press.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in production: Picture–word interference studies. *Journal of Memory and Language, 29,* 86–102.

Seymour, P. H. K. (1979). *Human visual cognition.* New York: St. Martin's Press.

Stemberger, J. P. (1983). Inflectional malapropisms: Form based errors in English morphology. *Linguistics, 21,* 573–602.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). Hillsdale, NJ: Erlbaum.

Theios, J., & Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review, 96,* 5–24.

Vorberg, D. (1990). [Interacting parallel channels: RT predictions for SOA experiments]. Unpublished materials.

Winer, J. B. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology, 81,* 226–234.

Zwitserlood, P. (1989). The locus of the effects of sentential–semantic context in spoken-word processing. *Cognition, 32,* 25–64.

(Appendix follows on next page)

Appendix

Phonological Test Probes Used in Experiment 5

1. *stoep*
2. *steno*
3. *bank*
4. *oorzaak*
5. *thuis*
6. *zoon*
7. *teil*
8. *jacht*
9. *leed*
10. *museum*
11. *moed*
12. *roos*
13. *koord*
14. *kist*
15. *rits*
16. *doel.*

(The numbers correspond to those in Table 1.)

Semantic and Phonological Test Probes Used in Experiment 6

Semantic Test Probes

1. *kast* (cupboard)
2. *vetplant* (thick-leaf)
3. *bandelichter* (tire jack)
4. *pistool* (pistol)
5. *gieter* (watering can)
6. *vla* (custard)
7. *horloge* (watch)
8. *gesp* (buckle)
9. *nijlpaard* (hippo)
10. *telefoon* (telephone)
11. *punaise* (thumbtack)
12. *pijp* (pipe)
13. *windwijzer* (weathercock)
14. *haar* (hair)
15. *teen* (toe)
16. *tas* (pouch).

Phonological Test Probes

1. *kassa*
2. *vetvlek*
3. *bandiet*
4. *pistache*
5. *gitaar*
6. *vlaag*
7. *hormoon*
8. *gems*
9. *nijptang*
10. *telegram*
11. *puree*
12. *pijn*
13. *winstcijfer*
14. *haak*
15. *teef*
16. *tap.*

(The numbers correspond to those in Table 1.)

Received November 21, 1989
Revision received May 9, 1990
Accepted May 14, 1990 ■