Planning levels in naming and reading complex numerals

MARJOLEIN MEEUWISSEN
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ARDI ROELOFS
*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
and F. C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands*

and

WILLEM J. M. LEVELT
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

On the basis of evidence from studies of the naming and reading of numerals, Ferrand (1999) argued that the naming of objects is slower than reading their names, due to a greater response uncertainty in naming than in reading, rather than to an obligatory conceptual preparation for naming, but not for reading. We manipulated the need for conceptual preparation, while keeping response uncertainty constant in the naming and reading of complex numerals. In Experiment 1, participants named three-digit Arabic numerals either as house numbers or clock times. House number naming latencies were determined mostly by morphophonological factors, such as morpheme frequency and the number of phonemes, whereas clock time naming latencies revealed an additional conceptual involvement. In Experiment 2, the numerals were presented in alphabetic format and had to be read aloud. Reading latencies were determined mostly by morphophonological factors in both modes. These results suggest that conceptual preparation, rather than response uncertainty, is responsible for the difference between naming and reading latencies.

The naming of objects usually takes longer than reading their names. Cattell (1886) was the first to measure this difference in latency between naming and reading, which he explained in terms of greater practice for reading than for naming. However, a study by Brown (1915) showed that even after 10 days of practice in naming, reading was, on average, 131 msec faster than naming. Ligon (1932) investigated the effect of practice by examining the development of the difference between reading and naming in a large group of school children whose ages varied from 6 to 16 years. Naming and reading times decreased with age, but the difference between naming and reading remained constant (286 msec, on average). Given that the difference was present from the outset, it seems that a differential amount of practice between naming and reading is not the crucial factor. Stroop

(1935), however, contested Ligon's conclusion. On the basis of the observation that practicing color naming yields interference from incongruent colors in word reading, he argued that, as a consequence of different amounts of training, "the word stimulus has been associated with the specific response 'to read', while the color stimulus has been associated with various responses: 'to admire', 'to name', 'to reach for', 'to avoid', etc." (p. 660).

Fraisse (1967, 1969) showed in a series of experiments that more time is required to name than to read one and the same stimulus. He used the sign O and had participants either read it as "oh" or name it as "circle." He found a latency difference of about 100 msec between naming and reading. Following Stroop (1935), he assumed that the difference reflected a higher degree of uncertainty in terms of number of alternatives in naming, as compared with reading. More recently, Theios and Amrhein (1989) reviewed the literature concerning the time difference between naming and reading, and they estimated the difference in time to be 160 msec, on average. Theios and Amrhein argued that the difference between naming and reading is due to both the number of possible response alternatives (Fraisse, 1967, 1969; Stroop, 1935) and a difference in functional architecture. In particular, they argued that reading a word aloud entails a fast process of transcoding a stimulus into the required articulatory response by means of grapheme-to-phoneme transforma-

tions, without necessarily accessing the mental lexicon. By contrast, naming a picture requires additional planning—namely, selecting or computing the meaning of the depicted object (*conceptual preparation*) and selecting the correct name from other plausible alternatives.

Ferrand (1999) empirically contrasted response uncertainty and conceptual preparation by using Arabic numerals, ranging from 1 to 20, for naming and reading. A crucial attribute of Arabic numerals, which they share with their written names, is that there is no response uncertainty. That is, a given Arabic numeral represents only one precise magnitude and corresponding response. Given the same degree of uncertainty in the case of Arabic numerals and the corresponding written number words, reading and naming times should be equal, which was indeed what Ferrand observed. This suggests that response uncertainty, rather than obligatory conceptual preparation, is the crucial factor (see Stroop, 1935).

However, Ferrand's (1999) argument presumes that the naming of Arabic numerals requires conceptual preparation, just like object naming (see Brysbaert, 1995; McCloskey, 1992), but that does not need to be the case (Cipolotti & Butterworth, 1995; Dehaene, 1992). It may be that Arabic numerals and written number words are processed in a similar way, different from that for pictures (Roelofs, 2003a). In the theory for word production advanced by Levelt, Roelofs, and Meyer (1999; Roelofs, 1992, 1997, 2003b)—henceforth, LRM—the reading of words can be achieved without conceptual preparation, whereas picture naming obligatorily engages the conceptual level. Roelofs (2003b) showed that an implemented model for this theory, WEAVER++, accounts for a wide range of findings on the naming of colors/objects and reading their names.

The LRM theory holds that the conceptually driven planning of words is a process with stages, leading from conceptual preparation to the initiation of articulation. Three planning levels can be distinguished (see Figure 1): conceptual preparation, lemma retrieval, and form encoding. Conceptual preparation is the activation and selection of lexical concepts for the utterance. Lemma retrieval involves the retrieval of memory representations of the syntactic properties of the words to be expressed. After a lemma has been selected, the form of the utterance is encoded, which includes access to morphological, phonological, and phonetic information about the word.

Empirical evidence in support of these levels of representation in numeral production comes from chronometric evidence from our laboratory (Roelofs, 2003a) and from speech errors. For example, Sokol and McCloskey (1988) have reported a single case study on patient J.S., who had no problem understanding numerals but made errors in both speaking and writing them. J.S.'s impairment was characterized by numeral-syntax and morpheme errors. The numeral-syntax errors made by J.S. were quite comparable in amount and nature for spoken and written numeral production. For example, when presented with the complex numeral 146,359, he erroneously responded with "one hundred thousand forty-six three hundred fifty-nine" in both output modalities. In contrast, J.S.'s morpheme substitution errors were in the spoken modality, rather than in written numeral production. For example, he responded "*six* hundred nine" to the visually presented numeral 309 but correctly wrote "three hundred nine." These findings are in line with the LRM theory's notion of a modality-independent process of lemma retrieval and construction and a modality-specific process of form encoding.
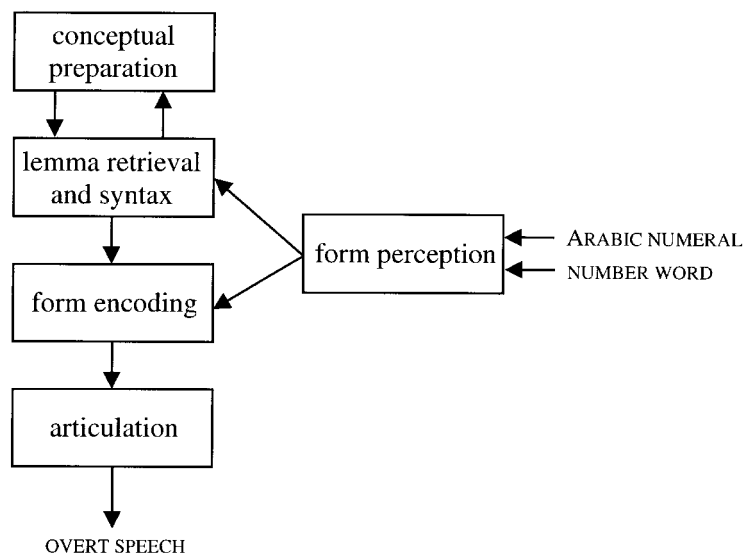

Figure 1. Planning levels in spoken numeral production, following Levelt, Roelofs, and Meyer (1999; Roelofs, 2003a).

The LRM theory holds that complex numerals are constructed and represented as composite number concepts and lemmas. In the conceptually driven generation of complex spoken numerals, a composite number concept is used to retrieve the corresponding lemmas, which are then syntactically ordered. Next, the appropriate form is constructed, which includes retrieving morphemes and segments, constructing the appropriate phonological syllables, and retrieving motor programs for these syllables. For example, telling the time from an analogue clock in Dutch minimally involves conceptually determining the reference point (full hour, half past the hour, or next full hour) and the distance from the reference point in minutes. Next, the corresponding lemmas need to be retrieved (e.g., *kwart*, English *quarter*; *voor*, English *to*; and *drie*, English *three*) and syntactically ordered. Finally, the corresponding morphemes (e.g., <kwart>) and phonemes (/k/, /w/, etc.) need to be retrieved, the phonemes should be syllabified, and the corresponding motor programs need to be recovered.

In contrast, for canonically naming a complex Arabic numeral such as 245 (e.g., as a house number) in Dutch, it suffices to determine the hundreds, tens, and ones from the input, retrieving and ordering the corresponding lemmas (in Dutch, the ones are mentioned before the tens in the utterance), recovering the morphemes <twee>, <honderd>, <vijf>, <en>, and <veertig>, retrieving their phonemes, syllabifying them, and recovering the motor programs. Similarly, reading aloud the alphabetically written numeral TWEEHONDERDVIJFENVEERTIG may be achieved by retrieving the corresponding morphemes <twee>, <honderd>, <vijf>, <en>, and <veertig> respecting the order in the input, retrieving their phonemes, syllabifying them, and recovering the motor programs. Alternatively, the phonemes may be retrieved through application of grapheme–phoneme correspondence rules, after which the phonemes are syllabified and the corresponding motor programs are recovered. Whether one or two nonconceptual routes (i.e., a lexical-form route and a grapheme-to-phoneme route) exist for reading words is hotly debated (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996). The discussions of this issue have concentrated on reading morphologically simple, monosyllabic words. Not much is known about reading complex numerals, which are polysyllabic, morphologically complex words. The issue of planning routes in reading complex numerals is addressed in the present article.

Our theory holds that, unlike canonically naming complex Arabic numerals as house numbers, naming Arabic numerals as clock times demands conceptual preparation. Naming digital clock times (e.g., 2:45) involves determining the hours and minutes from the input, conceptually determining the reference point (for 2:45, this is the next full hour) and the distance from the reference point in minutes (i.e., 15), retrieving the corresponding lemmas (the lemma *kwart* for 15 and the lemmas of *voor* and *drie*) and syntactically ordering them, retrieving the corresponding morphemes and phonemes, syllabifying the phonemes, and recovering the corresponding motor programs.

To test between the response uncertainty and the conceptual preparation explanations of the difference in reading and naming times, we compared the naming of complex Arabic numerals and the reading of the corresponding names. In Experiment 1, participants named all three-digit Arabic numerals between 200 and 955 ending with a 0 or a 5, either canonically as house numbers or as clock times—henceforth referred to as *mode*. For example, the participants produced the Dutch equivalent of "two hundred forty-five" in response to 245 and "quarter to three" to 2:45. In Experiment 2, the participants simply read aloud the numerals presented in alphabetic format, again as house numbers and as clock times. For example, they said (in Dutch) "two hundred forty-five" in response to TWO HUNDRED FORTY-FIVE and "quarter to three" in response to QUARTER TO THREE.

With respect to house number naming in Experiment 1, the LRM theory predicts that lemma and word-form level factors, such as the morphophonological length of the utterance, will mostly determine naming latencies, because a nonconceptual route suffices for naming the Arabic numerals. In contrast, in the case of digital clock time naming, the theory predicts that conceptual level factors will additionally determine naming latencies. In Experiment 2, in which the numerals had to be read aloud, we would expect no conceptual involvement whatsoever for both modes of response. Therefore, we would predict that factors operating at the lemma and word-form encoding level will determine naming latencies in both response modes. In contrast, if response uncertainty determines the difference in time between reading and naming (Ferrand, 1999), the naming latencies should not differ between the modes and the input formats, because uncertainty would be held constant among these conditions.

EXPERIMENT 1

In the first experiment, we tested our theory's prediction that the latencies for naming house numbers should be determined mainly by morphophonological factors, whereas the latencies for naming clock times should exhibit a strong conceptual involvement. In contrast, under the uncertainty account (Ferrand, 1999), there should be no difference between the response modes.

Method

Participants. Thirty speakers participated in the experiment. They were undergraduate students at Nijmegen University and native speakers of Dutch and had normal or corrected-to-normal vision. They were paid for their participation.

Materials. The stimuli consisted of three-digit Arabic numerals ranging from 200 to 955 and ending with either a 0 or a 5. Collapsed across the hundreds and the hours, this yielded 12 different number

types: (00) 200, 300, . . . (05) 205, 305, . . . (10) 210, 310, . . . (15) 215, 315, . . . (20) 220, 320, . . . (25) 225, 325, . . . (30) 230, 330, . . . (35) 235, 335, . . . (40) 240, 340, . . . (45) 245, . . . (50) 250, . . . and (55) 255. . . . For each type, there were eight instances (e.g., for type [00], the instances were 200, 300, 400, 500, 600, 700, 800, and 900). Note that numerals ranging from 100 to 155 were excluded from the stimulus set. For numerals in the interval of 100–199, no explicit information is encoded in Dutch utterances about which particular hundred is involved. For instance, in Dutch, the numeral 105 is pronounced as "honderdvijf " (hundred five), whereas the numeral 205 is pronounced as "tweehonderdvijf " (*two* hundred five). Furthermore, we controlled for possible voice key artifacts by instructing the participants to start each response with the same word—*op* (*at*) for the house numbers and *om* (*at*) for the clock times. In this way, the responses in all the conditions started with the same phoneme. In Dutch, there are three utterance referents for clock times (full hour, half past the full hour, and next full hour), rather than the two in English, yielding utterances in Dutch like "vijf voor half vier" (literally, "five to half four"), rather than utterances like "three twenty five" in English.

With respect to the instruction to name clock times in the relative way, only the number types (20) and (40) have two naming options. For example, 2:40 can be named as "tien over half drie" (literally, "ten past half three") and as "twintig voor drie" (literally, "twenty before three"), the latter being infrequent. To make sure that there was no response uncertainty, given the instructions, we asked the participants before each experiment to name several prototypes of the clock times in a relative way (the standard way for telling time in Dutch, in contrast to the absolute way, which is standard in American English). All the participants responded correctly for all clock types 100% of the time. Crucially, they all named the number types (20) and (40) in the most common way—for example, 2:20 as "tien voor half drie" (literally, "ten before half three") and 2:40 as "tien over half drie" (literally, "ten past half three"). So it is safe to assume that in our experiment, there was no response uncertainty for the clock times.

Design. Each mode (house number or clock time) was tested in a separate block of trials. Each experimental block randomly presented the 96 stimuli (12 types × 8 instances), which were repeated three times, yielding 288 trials per mode. A practice block of 12 trials, containing items of each stimulus type, preceded a block of trials. The order in which each response mode was tested was counterbalanced across participants.

Procedure. The participants were tested individually. They were seated in a dimly lit, soundproof cabin, in front of a computer monitor (NEC Multisync) and a Sennheiser microphone. The distance between the participant and the screen was approximately 50 cm. The experiment was run with the NESU experimental software developed at the Max Planck Institute. Naming latencies were measured using a voice key apparatus. Preceding each experimental block, the participants were provided with a written instruction stating how the three-digit numerals had to be named (either as house numbers or as clock times). Furthermore, they were asked to respond in a fluent manner.

The structure of a trial was as follows. First, the participant saw a warning signal (an asterisk) for 200 msec, directly followed by the display of a three-digit Arabic numeral for 100 msec. The stimuli were presented in black on a white background. Before the next trial started, the screen went blank for another 1.4 sec. The total duration of the trial was 1.7 sec. Naming latencies were collected for each participant individually. Each trial was recorded using a DAT recorder. An experimental session lasted about 20 min.

Analyses. A trial was considered invalid when it included a speech error, a voice key error or a time-out or when a wrong oral response was given. Invalid trials were excluded from the statistical analyses. Furthermore, voice key errors and time-outs were excluded from the error analyses. Voice key errors and time-outs occurred, respectively, in 1.3% and 0.1% of the trials in the house number mode and in 0.8% and 0.2% of the trials in the clock time mode.

Results and Discussion

Table 1 gives the mean naming latencies, the standard deviations, and the percentages of errors for the house number and clock time modes. Figure 2 shows the mean naming latencies for the 12 number types in the house number mode (upper panel) and the clock time mode (lower panel).

Errors. The participants made fewer errors in the house number mode than in the clock time mode [$F_1(1,29) = 38.57, p < .0001; F_2(1,84) = 66.53, p < .0001$]. Errors varied depending on the number type [$F_1(11,319) = 5.29, p < .0001; F_2(11,84) = 4.01, p < .0001$]. Furthermore, the effect of number type depended on the mode [$F_1(11,319) = 3.75, p < .0001; F_2(11,84) = 3.89, p < .0001$]. As is indicated by Table 1, most errors were made in the slowest conditions, so there is no evidence for a speed–accuracy tradeoff.

Naming latencies. Naming latencies were, on average, 169 msec shorter in the house number mode than in the clock time mode [$F_1(1,29) = 134.48, p < .0001; F_2(1,84) = 8,967.61, p < .0001$]. Moreover, the latencies depended on type [$F_1(11,319) = 50.50, p < .0001; F_2(11,84) = 86.90, p < .0001$]. Furthermore, the effect of type varied with mode [$F_1(11,319) = 33.20, p < .0001; F_2(11,84) = 47.23, p < .0001$].

To gain further insight into the factors underlying the interaction between type and mode, we ran multiple regression analyses over the total data set. The following groups of predictor variables were entered into the analyses: (1) magnitude information (absolute magnitude of the whole number displayed and absolute magnitude of the last two digits and its logarithm [suggested by Wim Fias, personal communication, November 12, 2002]), (2) numeral length (number of morphemes, number of syllables, and number of phonemes), and (3) frequency (logarithm of the whole-form frequency and logarithm of the morpheme frequency). We fitted a multilevel mul-

Table 1
Mean Naming Latencies (in Milliseconds), Standard Deviations, and Percentages of Errors (E%) per Mode and Type for Experiment 1, With Naming of House Numbers and Clock Times Presented in Arabic Format

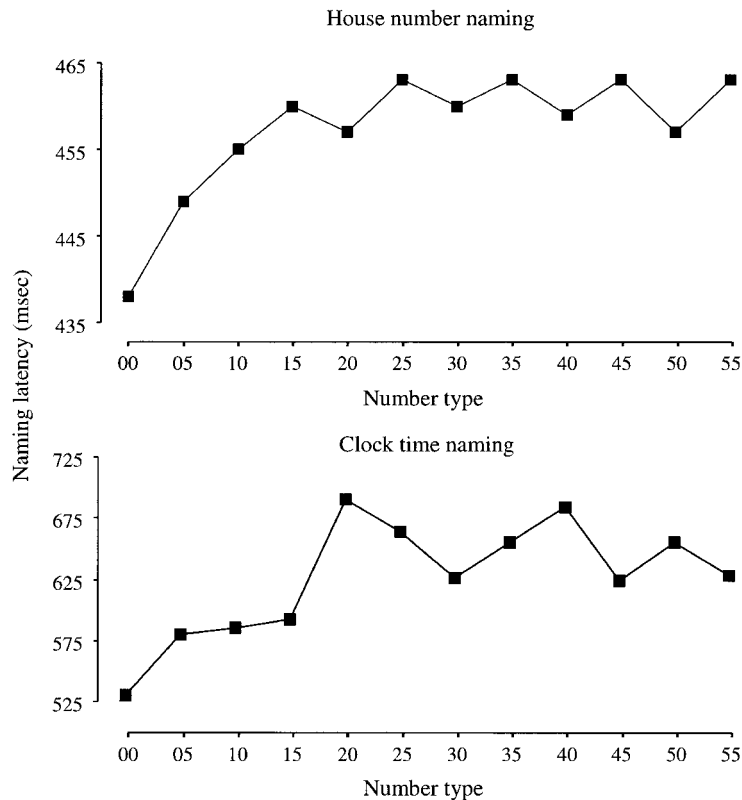|	Mode					
	House Number			Clock Time		
Type	M	SD	E%	M	SD	E%
00	438	100	0.6	530	86	0.1
05	449	99	1.7	580	111	2.2
10	455	105	1.9	585	113	1.9
15	460	108	1.1	592	114	2.1
20	457	108	1.7	690	156	5.3
25	463	105	1.7	664	151	4.6
30	460	111	1.4	626	127	2.8
35	463	106	1.1	655	141	4.0
40	459	104	0.8	684	159	6.3
45	463	109	2.1	624	127	3.3
50	457	105	0.8	655	132	5.3
55	463	110	2.5	628	124	5.4
Total	457	106	1.4	626	128	3.6

House number naming

Clock time naming

Figure 2. Latencies, in milliseconds, for the naming of house numbers and clock times in Arabic format in Experiment 1.

tiple regression model (Pinheiro & Bates, 2000; see also Lorch & Myers, 1990) to the data, with logarithm of the naming latencies as the dependent variable and participant as the error stratum. In all the analyses reported in this article, we first entered the total set of variables as predictors and assessed which made a major contribution. From there, we constructed the best-fitting model.

The best-fitting regression model for the house number naming latencies included three predictor variables—namely, the logarithm of the whole-form frequency, the logarithm of the morpheme frequency, and the numeral length (i.e., the number of phonemes). We observed significant effects of all three predictors. A greater log morpheme frequency led to longer response latencies [$t(7,766) = 2.72, p < .01$]. Longer numerals likewise elicited longer response latencies [$t(7,766) = 3.79, p < .001$]. In contrast, a greater log whole-form frequency led to shorter latencies [$t(7,766) = -2.38, p < .02$]. All effects remained significant in sequential analyses of variance—that is, after partialing out the variance contributed by the other two variables ($p < .02$ for all analyses).

Furthermore, we observed three random effects: a main effect of participant (estimated standard deviation = 0.1311), an interaction of participant and trial number, indicating that the participants responded more quickly over trials (estimated standard deviation = 0.0004, log-

likelihood ratio = $233.86, p < .0001$), and an interaction of participant and number of phonemes (estimated standard deviation = 0.0024, log-likelihood ratio = 8.57, $p < .05$). In addition, the random effects were pairwise correlated: participant and trial ($-.46$; i.e., the decrease of response latency in the course of the experiment was greater for slow participants than for fast participants), participant and number of phonemes (.43; i.e., slow participants were especially slow when responding to longer numerals), and trial and number of phonemes ($-.76$; i.e., the participants who became faster during the experiment responded relatively slowly to longer numerals). The standard deviation of the residual error in the model was 0.154. The correlation between the observed and the predicted naming latencies was .64, indicating a multiple R^2 of 42%. Adding an extra group of predictor variables (i.e., the absolute magnitude of the whole number displayed and its logarithm or the absolute magnitude of the last two digits and its logarithm) did not significantly increase the amount of variance accounted for.

These multiple regression results suggest that for the production of a house number, the numeral length and frequency of the utterance are the major determinants for speech production onset. As Figure 2 shows, a "zigzag" relation can be observed between utterance length and naming latencies for the number types from (10) to (55).

The production of an additional number component (e.g., "three hundred twenty-*five*" in response to 325, as compared with "three hundred twenty" in response to 320) is accompanied by an increase in naming latencies.

The same predictor variables as those used for the house number mode were also entered in a multiple regression analysis of the clock time naming latencies. In addition, we included another group of predictors pertaining to the conceptual preparation needed for clock time naming: (4) utterance referent (full hour, half past the hour, or next full hour) and distance from referent (quarter/zero, five, ten minutes), with quarter/zero indicating a conceptually prominent distance. We again fitted a multilevel multiple regression model to the data, with logarithm of the naming latencies as the dependent variable and participant as the error stratum.

The best-fitting regression model included the utterance referent, the distance from the referent, the logarithm of the morpheme frequency, and the logarithm of the whole-form frequency as predictor variables. We observed significant effects for all four predictors. Naming latencies differed depending on the utterance referent $[t(7,863) = 31.27, p < .001]$. Furthermore, naming latencies varied depending on the distance from the referent $[t(7,863) = 4.14, p < .001]$. A greater log morpheme frequency led to longer response latencies $[t(7,863) = 6.92, p < .001]$. In contrast, a greater log whole-form frequency led to shorter latencies $[t(7,863) = -3.57, p < .001]$. All effects remained significant after partialing out the variance contributed by the other three variables ($p < .01$ for all analyses).

The best-fitting model included two random effects: a main effect of participant (estimated standard deviation = 0.1559) and an interaction of participant and trial number, indicating that during the experiment the participants became faster (estimated standard deviation = 0.0004, log-likelihood ratio = 395.51, $p < .0001$). In addition, the random effects (participant and trial) were pairwise correlated ($-.63$; i.e., the decrease of response latency in the course of the experiment was greater for the slow participants than for the fast participants). The standard deviation of the residual error in the model was 0.139. The correlation between the observed and the predicted naming latencies was .67, indicating a multiple R^2 of 44%. Adding an extra group of factors did not significantly increase the amount of variance accounted for.

These multiple regression results suggest that conceptual preparation is required for digital clock time naming. The differences in reference points (full hour, half past the hour, next full hour) were reflected in the naming latencies. Utterances referring to the full hour had a mean onset latency of 583 msec, utterances referring to the half hour took 673 msec to begin, and utterances referring to the next full hour took 643 msec to begin $[F_1(2,58) = 69.3, p < .0001; F_2(2,61) = 115.4, p < .0001]$. Furthermore, the differences in minutes were also reflected in the naming latencies. Utterances mentioning a 10-min distance took, on average, 654 msec

to begin, utterances mentioning a 5-min distance took 632 msec to begin, and utterances that mentioned a zero distance or a quarter took 594 msec to begin $[F_1(2,58) = 74.2, p < .0001; F_2(2,93) = 17.9, p < .0001]$.

These findings support the idea of a procedural semantics (see Johnson-Laird, 1983) for clock time naming. In preparing a conceptual representation for a clock time on the basis of digital input, a speaker has to determine the reference hour and minutes. Let the input be *xyz* and assume that there is a procedure that, as a first step, determines the referent. If *yz* is smaller than 20, the reference hour is "*x*"; if it falls between 20 and 40, the reference hour is "half ($x + 1$)"; and if it is larger than 40, the reference hour is "$x + 1$" (cf. Bock, Irwin, Davidson, & Levelt, 2003; Meeuwissen, Roelofs, & Levelt, in press). If the specifications of the hour ("*x*") and the next hour ("$x + 1$") are achieved by separate semantic procedures that take time, utterances referring to the full hour ("*x*") should be produced more quickly than those referring to the next full hour ("$x + 1$") and those referring to half past the hour ["half ($x + 1$)"], as was empirically observed. The data suggest that, overall, specifying a 5-min distance takes less time than specifying a 10-min distance. However, the effect of the minutes (5 vs. 10) depends on the reference hour [full hour, half past the hour, next full hour; $F_1(2,58) = 4.79, p < .012; F_2(2,58) = 3.03, p < .06$]. The latency difference between specifying 5 and 10 min is around 25 msec for half past the hour and the next full hour, but it is only 5 msec for the full hour. This may be explained by the fact that whereas specifying the minutes relative to the half hour and next full hour requires a numerical transformation of the input (e.g., the 20 in 2:20 has to be transformed into "ten," and the 25 in 2:25 into "five"), such a transformation is not required for the full hour (e.g., the 05 in 2:05 remains "five" and the 10 in 2:10 remains "ten"). In summary, the findings for the clock times support a procedural semantics in which hour and minute transformations are carried out.

To conclude, whereas the naming latencies in the house number mode were determined mostly by morphophonological factors, the clock time naming latencies revealed an additional conceptual involvement. Because response uncertainty was held constant, the data suggest that two naming routes exist. Which of the two routes is adopted in the naming of complex numerals appears to depend on the naming mode—house number versus clock time. Importantly, canonically naming Arabic numerals (as house numbers) does not seem to require conceptual preparation, contrary to what Ferrand (1999) assumed.

EXPERIMENT 2

In the second experiment, the numerals of Experiment 1 were presented in an alphabetic format in both response modes. For example, instead of 215 and 2:15, the stimuli were now the Dutch equivalents of TWO HUNDRED FIFTEEN and QUARTER PAST TWO. According to our

theory, conceptual transformations are now no longer necessary for clock time production. Consequently, morphophonological factors should determine the response latencies in both response modes.

Method

Participants. Twenty speakers, who did not participate in Experiment 1, took part in the experiment. They were undergraduate students at Nijmegen University and native speakers of Dutch and had normal or corrected-to-normal vision. They were paid for participation.

Materials. The same set of stimuli was used as that in Experiment 1, except that they were presented in an alphabetic format in both response modes. Again, the participants were instructed to start each response with the same word: *op* (*at*) for the house numbers and *om* (*at*) for the clock times.

Design and Procedure. The design and procedure were identical to those in Experiment 1. Note that, although the speech output was identical to that in Experiment 1, by presenting stimuli in alphabetic format, the perceptual input now differs considerably between the two modes of response. More specifically, in Dutch, numbers are written as a connected string of letters (e.g., TWEEHONDERDVIJFENDERTIG; English, TWO HUNDRED THIRTY-FIVE), whereas clock times are conventionally divided by spaces (e.g., VIJF OVER HALF DRIE; in English, literally FIVE PAST HALF THREE). This difference might lead to a reduced discriminability of house numbers, as compared with clock times. Therefore, we decided to extend the presentation time for each stimulus up to 1,500 msec in order to make sure each stimulus was fully perceived. The following trial structure was used. First, the participant saw a warning signal (an asterisk) for 200 msec, directly followed by the display of a numeral for 1,500 msec. Stimuli were presented on a white background, printed in black. Before the next trial started, the screen went blank for another 500 msec. The total duration of the trial was 2.2 sec. The participants were instructed to read aloud the numerals as soon as they appeared on screen.

Analyses. A trial was considered invalid when it included a speech error, a voice key error, or a time-out or when a wrong oral response was given. Invalid trials were excluded from the statistical analyses. Furthermore, voice key errors and time-outs were excluded from the error analyses. Voice key errors and time-outs occurred in 1% and 0.3% of the trials in the house number mode and in 0.9% and 0.2% of the trials in the clock time mode, respectively.

Results and Discussion

Table 2 gives the mean reading latencies, standard deviations, and percentages of errors for the house number and clock time modes. Figure 3 shows the reading latencies for the 12 number types for the house number mode (upper panel) and the clock time mode (lower panel).

Errors. The participants made no more errors in the house number mode than in the clock time mode [$F_1(1,19) = 1.34, p > .10; F_2(1,84) = 10.90, p < .05$]. Errors varied depending on the number type [$F_1(11,209) = 4.08, p < .0001; F_2(11,84) = 3.83, p < .0001$]. The effect of number type did not vary with mode [$F_1(11,209) = 1.32, p > .10; F_2(11,84) = 2.25, p < .05$].

Reading latencies. Reading latencies did not differ significantly between the two modes (both F_1 and $F_2 < 1$). However, number type had an effect [$F_1(11,209) = 11.69, p < .0001; F_2(11,84) = 16.71, p < .0001$]. Furthermore, the effect of number type depended on the mode [$F_1(11,209) = 4.42, p < .0001; F_2(11,84) = 3.54, p < .0001$]. To gain further insight into the factors determining the reading latencies, we carried out multiple regression analyses, using the same factors as those in Experiment 1. We fitted multilevel regression models to the data, with logarithm of the reading latencies as the dependent variable and participant as the error stratum.

The best-fitting regression model for the house numbers included the number of morphemes as a predictor variable. We observed significant effects for the predictor in the model. Longer numerals, in terms of the number of morphemes, elicited longer reading latencies [$t(5,312) = 7.25, p < .0001$]. The model included two random effects: a main effect of participant (estimated standard deviation = 0.1526) and an interaction of participant and trial number, indicating that the participants responded more quickly over trials (estimated standard deviation = 0.0004, log-likelihood ratio = 130.49, $p < .0001$). In addition, the random effects (participant and trial) were pairwise correlated ($-.54$; i.e., the slower participants gained more speed during the experiment than did the faster participants). The standard deviation of the residual error in the model was 0.176. The correlation between the observed and the predicted reading latencies was .66, indicating a multiple R^2 of 43%. Adding an extra group of factors did not significantly increase the amount of variance accounted for.

Thus, morphophonological factors determined the onset latencies in the reading of house numbers, as they did in the naming of the house numbers in Experiment 1. However, whereas the number of phonemes affected the naming latencies in Experiment 1, they did not affect the reading latencies in the present experiment. Only the number of morphemes had an effect here.

To gain further insight into the factors determining the reading latencies for the clock time mode, we carried out multiple regression analyses, using the same factors as those in Experiment 1. Again, we fitted a multilevel regression model to the data, with logarithm of the reading latencies as the dependent variable and participant as the error stratum.

Table 2
Mean Reading Latencies (in Milliseconds), Standard Deviations, and Percentages of Errors (E%) per Mode and Type for Experiment 2, With Reading of House Numbers and Clock Times Presented in Alphabetic Format

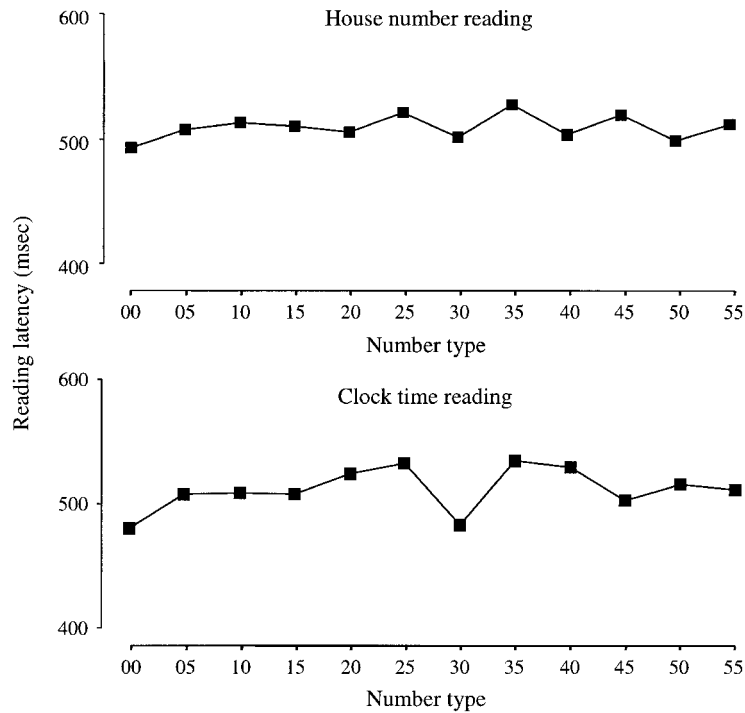| | Mode | | | | | |
| | House Number | | | Clock Time | | |
Type	M	SD	E%	M	SD	E%
00	493	125	1.9	480	115	0.4
05	507	128	0.8	507	137	1.3
10	513	131	3.1	508	132	1.0
15	510	138	2.9	507	133	2.3
20	505	138	2.7	524	130	3.5
25	521	141	3.3	532	137	3.5
30	501	122	1.7	482	113	1.0
35	527	136	3.8	534	141	1.7
40	503	127	2.1	529	140	1.0
45	519	133	5.8	502	124	2.9
50	498	136	1.7	515	129	3.1
55	511	131	5.2	511	135	2.5
Total	509	132	2.9	511	131	2.0

Figure 3. Latencies, in milliseconds, for the reading of house numbers and clock times in alphabetic format in Experiment 2.

The best-fitting regression model included the number of morphemes as a predictor variable. We observed significant effects for the predictor in the model. Longer numerals, in terms of number of morphemes, elicited longer reading latencies [$t(4,424) = 13.15, p < .0001$]. The model included two random effects: a main effect of participant (estimated standard deviation = 0.1136) and an interaction of participant and trial number, indicating that the participants responded more quickly over trials (estimated standard deviation = 0.0004, log-likelihood ratio = 129.24, $p < .0001$). In addition, the random effects (participant and trial) were pairwise correlated ($-.16$; i.e., the slower participants gained more speed during the experiment than did the faster participants). The standard deviation of the residual error in the model was 0.145. The correlation between observed and the predicted reading latencies was .66, indicating a multiple R^2 of 44%. Adding an extra factor did not significantly increase the amount of variance accounted for.

Thus, the manipulation of presentation format has been successful for the clock time mode. As was expected, the conceptual preparation found with an Arabic presentation format in Experiment 1 disappeared with the alphabetical presentation format in the present experiment.

To conclude, the number of morphemes was the sole major determinant of reading latencies in both response modes. This is in agreement with our theory, which presumes that a nonconceptual route suffices for the oral reading of numerals that are presented in an alphabetic format. Whereas the number of phonemes made a significant contribution in Experiment 1, it did not play a role here. We will address this in the General Discussion section.

COMPARISON OF EXPERIMENTS 1 AND 2

Experiments 1 and 2 showed that the naming of house numbers takes less time than does the naming of clock times (457 vs. 626 msec), whereas the reading times for the house numbers and the clock times are comparable (509 vs. 511 msec). To compare the effects for the different modes (house number vs. clock time) between the naming task (Arabic format) and the reading task (alphabetic format), we carried out a joint analysis of variance with mode as a within-subjects factor and task as a between-subjects factor.

We observed a strong interaction between mode and task [$F_1(1,48) = 66.57, p < .0001; F_2(1,168) = 3,721.85, p < .0001$], which corroborated our findings concerning the planning levels engaged by each task. The longer naming latencies for the clock times than for the house numbers in Experiment 1 can be explained by the additional conceptual operations required for the Arabic input in clock time naming. The finding that this latency difference disappeared for reading of the house numbers and clock times in Experiment 2 suggests that conceptual transformations were no longer necessary for successful reading task completion.

We also found a latency difference between the naming and the reading of house numbers [457 msec and 509 msec, respectively; $t(22) = 14.5$, $p < .001$], despite the fact that the same spoken responses were required in the two tasks (e.g., the participants had to respond "tweehonderd" to both 200 and TWEEHONDERD). This result does not comply with Ferrand's (1999) original finding of no latency difference between naming and reading. However, Ferrand used one-digit and two-digit Arabic numerals ranging from 1 to 20, for which the difference in orthographic length (i.e., the number of digits and letters) between the Arabic and the alphabetic formats (e.g., 2 vs. TWO, etc.) was less than that in our study with three-digit Arabic numerals. The orthographic length was much smaller for the Arabic numerals in our Experiment 1 (all the numerals were three digits long; e.g., 235) than for the alphabetic numerals in Experiment 2 (e.g., TWEEHONDERDVIJFENDERTIG). The greater orthographic length for the alphabetic than for the Arabic format may have caused the difference in reading and naming times (i.e., longer times with greater orthographic length). In line with this account, a recent study by Pinel, Dehaene, Riviere, and LeBihan (2001) showed that identification of two-digit numerals is slower for the alphabetic than for the Arabic presentation format.

On first sight, the differential properties of the presentation formats seem unable to explain all the effects, because when one compares the clock time naming latencies obtained in Experiments 1 and 2, the latency difference goes in the opposite direction (511 msec for the alphabetic format, as compared with 626 msec for the Arabic format). However, given the evidence for conceptual preparation in Experiment 1, but not in Experiment 2, two opposing factors seem to play a role with clock times, making the latency difference difficult to interpret. In particular, the conceptual involvement with the Arabic numerals, but not with the alphabetic numerals, would increase the naming times relative to the reading times, whereas the greater orthographic length for the alphabetic numerals than for the Arabic numerals would increase the reading times relative to the naming times. The net result may be comparable to the latency for reading house numbers, which also involved reading alphabetic numerals without conceptual preparation, as was empirically observed (the reading times for the house numbers and the clock times differed by 2 msec).

Importantly, despite the absolute difference in response latencies between naming and reading house numbers, it is striking that the same factors made a contribution, in the multiple regression analyses, for both presentation formats. As Figures 2 and 3 show, the experiments yielded a similar zigzag pattern for both presentation formats, following the overall length of the utterance to be planned (in terms of number of phonemes and morphemes). Such a pattern suggests that comparable planning processes are carried out in the naming and the reading of house numbers.

GENERAL DISCUSSION

On the basis of equal latencies for naming and reading numerals with equal response uncertainty, Ferrand (1999) argued that the naming of objects is slower than reading their names, due to a greater response uncertainty in naming than in reading, rather than to obligatory conceptual preparation for naming, but not for reading. We questioned the assumption that Arabic numerals have the same status as pictured objects. Instead, we advanced the view that Arabic numerals are processed like written words, and not like pictures. This view also is in accord with the absence of a latency difference between the naming of Arabic numerals and reading their names in alphabetic format when the visual complexity of the Arabic and the verbal numerals is matched (which seemed to hold for Ferrand, 1999, but not for our study), but this occurs for different reasons. More specifically, no conceptual preparation is needed for Arabic numerals in their canonical naming mode, such as in the naming of house numbers.

To test between the response uncertainty and the conceptual preparation accounts, we manipulated the need for conceptual preparation while keeping response uncertainty constant in the naming and reading of complex numerals. In Experiment 1, the participants named three-digit Arabic numerals either as house numbers or as clock times. In line with our theory, the house number naming latencies were determined mostly by morphophonological factors, such as the number of phonemes, whereas the clock time naming latencies revealed additionally a strong conceptual involvement. In Experiment 2, the same numerals were presented in alphabetic format and had to be read aloud. The reading latencies for both house numbers and clock times were determined mostly by morphophonological factors. These results suggest that conceptual preparation, rather than response uncertainty, is responsible for the difference between naming and reading, at least in the case of numerals.

In Experiment 1, the number of phonemes and the morpheme and whole-form frequencies mainly determined the naming latencies in the house number mode. In contrast, in Experiment 2, the number of morphemes had a major influence. Note that the influence of morpheme frequency and the number of morphemes in reading suggests that the reading response was achieved lexically by retrieving morphemes, rather than through applying sublexical grapheme–phoneme correspondence rules. The difference between experiments may reflect a more conservative criterion for articulation initiation in naming than in reading (see Ferreira & Swets, 2002; Meyer, Roelofs, & Levelt, 2003). Information about the shape of the utterance (e.g., its length) is readily available from the visual input for reading numerals in alphabetic format, but not for naming them in Arabic format. If the utterance is planned out morphologically

for both reading and naming but more is planned at the phonological level for naming than for reading (e.g., the first phonological word for reading, but more than the first phonological word for naming), the effect of number of morphemes and phonemes may differ between naming and reading. More extensive planning at the phonological level may mean that the influence of the number of phonemes increases relative to the influence of number of morphemes. This may explain why, in the naming of house numbers (Experiment 1), the number of phonemes had an influence but the number of morphemes did not. It also explains why, for reading (Experiment 2), the opposite pattern holds—namely, that the number of morphemes had an influence but the number of phonemes did not. Whatever the exact reason may be for the difference in influence of form-level factors between Experiments 1 and 2, most important is that form-level factors mainly determined the reading response in both modes in Experiment 2 and the naming of house numbers in Experiment 1, in contrast to the strong conceptual involvement observed for the naming of digital clock times in Experiment 1.

The absence of evidence for conceptual preparation in the reading of house numbers and clock times in alphabetic format might seem to be at odds with the recent literature on reading, where effects of semantic factors have been reported (e.g., Strain, Patterson, & Seidenberg, 1995). Note, however, that whereas studies of the reading of isolated words have typically investigated the reading of monosyllabic, monomorphemic words, the present experiments examined the reading of polysyllabic, polymorphemic numerals. It cannot be excluded that different types of words vary in the involvement of meaning in reading aloud. Moreover, most reading models (e.g., Coltheart et al., 2001; Plaut et al., 1996) assume both conceptual and nonconceptual routes for reading aloud, supported by studies that suggest that conceptual preparation is not essential for the reading of simple words. For example, the acquired dyslexic patient W.L.P. reported by Schwartz, Saffran, and Marin (1980) could read aloud words (including words with irregular spellings) that she did not understand. Moreover, semantic effects are not observed in all reading tasks. In particular, whereas the effects are obtained in the lexical decision task, they are typically absent in reading aloud. In reviewing the literature, Shibahara, Zorzi, Hill, Wydell, and Butterworth (2003) suggested that, so far, only imageability has been shown to have an effect on reading latencies (e.g., Strain et al., 1995). However, imageability seems low in the case of complex numerals, so an effect of this factor was not to be expected in our experiments. Instead, the semantic factor typically affecting responding to numerals is magnitude.

In our experiments, the magnitude of the whole three-digit numeral and of its last two digits did not make a significant contribution to the latencies for responding to house numbers in either presentation format (Arabic and alphabetic). Although an overall increase in naming latencies can be observed in Experiment 1 (see Figure 2), both the multiple regression results and the specific shape of the latency pattern do not support a magnitude account of the data. If magnitude had played a role in the naming and reading of house numbers, we should have seen a steady increase in latencies from number type (00) to (55) and not the observed zigzag pattern—that is, the increase in latency from numerals ending with 5 (the longer numerals), as compared with those ending with 0 (the shorter ones). It should be noted that whereas magnitude effects have been reported for numerals up to 99 (Brysbaert, 1995), in the present study, more complex, three-digit numerals were investigated. Moreover, magnitude effects are typically observed in tasks in which Arabic numerals have to be manipulated or compared quantitatively (Dehaene, 1992; Fias, 2001), whereas in our experiments the numerals had to be named or read. Our results suggest that the naming and reading of house numbers requires no conceptual preparation, whereas the naming of digital clocks does, which suggests that conceptual preparation depends on the exact task and materials used (see Cipolotti & Butterworth, 1995).

To conclude, our study suggests that naming complex Arabic numerals canonically as house numbers does not obligatorily engage conceptual preparation, whereas naming them as clock times does. This means that the equal latencies for canonically naming Arabic numerals and reading their names (observed by Ferrand, 1999) cannot be taken as evidence for a response uncertainty account of the difference between object naming and reading their names. Instead, the present study suggests that, paralleling the difference between naming Arabic numerals as clock times and house numbers, obligatory conceptual preparation for object naming, but not for reading their names, is a crucial factor.

REFERENCES

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory & Language*, **48**, 653-685.

Brown, W. (1915). Practice in associating color names with colors. *Psychological Review*, **22**, 45-55.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, **124**, 434-452.

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

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, **124**, 375-390.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**, 589-608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, **44**, 1-42.

Ferrand, L. (1999). Why naming takes longer than reading? The special case of Arabic numbers. *Acta Psychologica*, **100**, 253-266.

FERREIRA, F., & SWETS, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory & Language,* **46**, 57-84.

FIAS, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research*, **65**, 250-259.

FRAISSE, P. (1967). Latency of different verbal responses to the same stimulus. *Quarterly Journal of Experimental Psychology*, **19**, 353-355.

FRAISSE, P. (1969). Why is naming longer than reading? *Acta Psychologica*, **30**, 96-103.

JOHNSON-LAIRD, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge: Cambridge University Press.

LEVELT, W. J. M., ROELOFS, A., & MEYER, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, **22**, 1-38.

LIGON, E. M. A. (1932). Genetic study of color naming and word reading. *American Journal of Psychology*, **44**, 103-121.

LORCH, R. F., JR., & MYERS, J. L. (1990). Regression analyses of repeated measures data in cognition research. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 149-157.

MCCLOSKEY, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, **44**, 107-157.

MEEUWISSEN, M., ROELOFS, A., & LEVELT, W. J. M. (in press). Naming analog clocks conceptually facilitates naming digital clocks. *Brain & Language.*

MEYER, A. S., ROELOFS, A., & LEVELT, W. J. M. (2003). Word length effects in picture naming: The role of a response criterion. *Journal of Memory & Language*, **48**, 131-147.

PINEL, P., DEHAENE, S., RIVIERE, D., & LEBIHAN, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage*, **14**, 1013-1026.

PINHEIRO, J. C., & BATES, D. M. (2000). *Mixed-effects models in S and S-PLUS.* New York: Springer-Verlag.

PLAUT, D. C., MCCLELLAND, J. L., SEIDENBERG, M. S., & PATTERSON, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.

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

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

ROELOFS, A. (2003a). *Functional architecture of naming dice, digits, and number words.* Manuscript submitted for publication.

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

SCHWARTZ, M. F., SAFFRAN, E. M., & MARIN, O. S. M. (1980). Fractionating the reading process in dementia: Evidence for word-specific print-to-sound associations. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 259-269). London: Routledge & Kegan Paul.

SHIBAHARA, N., ZORZI, M., HILL, M. P., WYDELL, T., & BUTTERWORTH, B. (2003). Semantic effects in word naming: Evidence from English and Japanese Kanji. *Quarterly Journal of Experimental Psychology*, **56A**, 263-286.

SOKOL, S. M., & MCCLOSKEY, M. (1988). Levels of representation in verbal number production. *Applied Psycholinguistics*, **9**, 267-281.

STRAIN, E., PATTERSON, K., & SEIDENBERG, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1140-1154.

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

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.