COMMENTS

Error Biases in Spoken Word Planning and Monitoring by Aphasic and Nonaphasic Speakers: Comment on Rapp and Goldrick (2000)

Ardi Roelofs
Max Planck Institute for Psycholinguistics and F. C. Donders Centre for Cognitive Neuroimaging

B. Rapp and M. Goldrick (2000) claimed that the lexical and mixed error biases in picture naming by aphasic and nonaphasic speakers argue against models that assume a feedforward-only relationship between lexical items and their sounds in spoken word production. The author contests this claim by showing that a feedforward-only model like WEAVER++ (W. J. M. Levelt, A. Roelofs, & A. S. Meyer, 1999b) exhibits the error biases in word planning and self-monitoring. Furthermore, it is argued that extant feedback accounts of the error biases and relevant chronometric effects are incompatible. WEAVER++ simulations with self-monitoring revealed that this model accounts for the chronometric data, the error biases, and the influence of the impairment locus in aphasic speakers.

An important issue for cognitive psychology generally and for psycholinguistics specifically is the extent to which cognitive systems consist of autonomous components (e.g., Donders, 1868/1969). To what degree is a large process split up as a hierarchy of component processes that are as independent of one another as the overall task allows? Recently, Rapp and Goldrick (2000) evaluated a number of theoretical positions—differing along the autonomy–interactivity dimension—on spoken word production, one of the most highly practiced skills in humans. Conceptually driven spoken word production obviously requires top-down feedforward activation from concepts to lexical items and from lexical items to their sounds. The question is whether activation cascades from level to level and, most importantly, whether there exists bottom-up production-internal feedback of activation. Recently, Norris, McQueen, and Cutler (2000) argued that spoken word comprehension proceeds in a strictly bottom-up feedforward fashion from sounds to lexical items: "Feedback is never necessary" (p. 299). Can production also do without feedback, as maintained by Levelt and colleagues (Levelt, Roelofs, & Meyer, 1999a, 1999b; Levelt et al., 1991a, 1999b; Roelofs, 1997, 2003; Roelofs, Meyer, & Levelt, 1996)?[1]

Rapp and Goldrick (2000) assessed five different theoretical positions on spoken word production by examining their ability to explain four sets of empirical findings concerning speech errors in picture naming by aphasic and nonaphasic speakers. The findings are (a) a lexical error bias and (b) a mixed error bias, both observed with normal speakers; (c) a pattern of only semantic errors in the absence of form errors arising from either a "conceptual" (patient K.E.) or a "postconceptual" (patients P.W. and R.G.B.) locus of damage in aphasia; and (d) differential effects of form relatedness according to the impairment locus—the presence of the mixed error bias with postconceptual damage and its absence with conceptual damage. To help evaluate the theoretical positions, Rapp and Goldrick studied the effects of manipulating interaction parameters in simulations of five specific connectionist models that were supposed to represent the full theoretical spectrum. Rapp and Goldrick concluded that the error data from picture naming (i.e., findings a, b, and d, above) and their simulation results refute models that assume a feedforward-only relationship between lexical items and their sounds and that their analyses favor a restricted-interaction account (RIA). This account is exemplified by the RIA model of Rapp and Goldrick, which is illustrated in Figure 1.

The RIA model assumes that spoken word planning is achieved through spreading activation and selection from a network with four layers of nodes representing lexical concepts, semantic features, lexical items, and speech segments. The nodes are bidirectionally connected except for the semantic feature nodes and the lexical nodes, which makes planning in the model "restricted" interactive. At each selection point (except for segment selection), the most active node is selected by introducing external activation.

[1] Rapp and Goldrick (2000, p. 463) incorrectly listed the models proposed by Roelofs (1992, 1997) as "discrete feedforward" models holding that only selected lexical items (lemmas) activate their speech segments. Rather, Roelofs (1992) proposed an interactive model for lexical selection only, and Roelofs (1997) proposed a feedforward cascading model for word-form encoding only. In these articles, no claim was made concerning a discrete or cascading relation between lexical selection and word-form encoding—this was done in Levelt et al. (1999b).
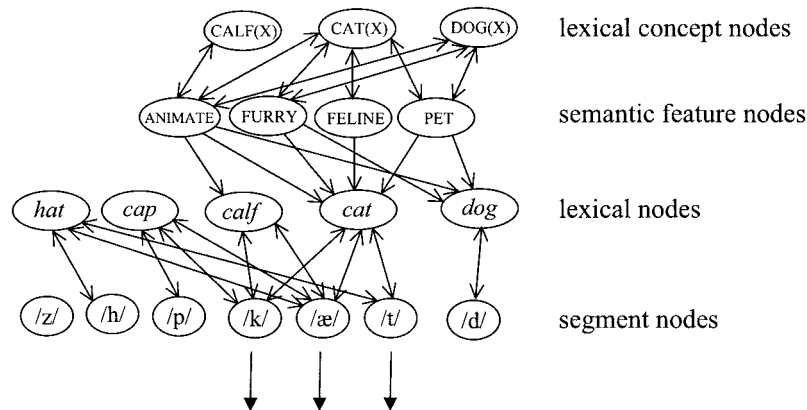
Figure 1. Illustration of the speech production network of the restricted-interaction account model (Rapp & Goldrick, 2000). The network has four layers of nodes, which are bidirectionally connected except for the semantic feature nodes and the lexical nodes.

Word planning begins by giving each of the semantic feature nodes corresponding to a target concept a jolt of activation. After a fixed number of time steps, the most highly activated concept node is selected by jolting its semantic feature nodes. After another fixed time interval, the most highly activated lexical node is selected by giving it a jolt of activation. Finally, the most highly activated segment nodes are selected after a set number of time steps.

The aim of this commentary is to show that feedforward-only models should not be dismissed on the basis of Rapp and Goldrick's (2000) analyses. I argue that the specific feedforward models that Rapp and Goldrick evaluated do not represent the class of feedforward models as a whole. In the next two sections (Lexical Bias and Mixed Error Bias), it is argued that existing models within this class such as WEAVER++ (Levelt et al., 1999a, 1999b; Roelofs, 1992, 1996, 1997, 2003) exhibit the error biases both production internally and in feedback that is speech comprehension based and that supports a speaker's self-monitoring of speech planning and production. Feedback through the comprehension system has been part of feedforward production models since the early 1980s (Levelt, 1983). Furthermore, although Rapp and Goldrick did not use production latency data in evaluating the models, they referred to chronometric evidence for production-internal feedback reported elsewhere, in particular, the "mixed distractor" latency effect observed by Damian and Martin (1999) and Starreveld and La Heij (1995, 1996). In the Latency Effect of Mixed Distractors section, it is argued, however, that Rapp and Goldrick's production-internal feedback account of the error biases is incompatible with a production-internal feedback account of the chronometric findings. In the Accounting for Both the Error and Latency Findings section, I report results from WEAVER++ simulations that included self-monitoring, demonstrating that this model accounts not only for the chronometric data but also for the error biases and the influence of impairment locus in brain-damaged individuals. In the Cohort Effects on the Speech Error Biases section, it is shown that WEAVER++'s predictions concerning certain monitoring effects on speech errors are empirically supported.

During conversation, speakers not only talk but they also listen to their interlocutors' speech and monitor their own speech for errors (and often repair them). Although the interplay among speaking words, listening, and self-monitoring stands at the heart of spoken conversation, it is often neglected in models of language use. The interplay is also ignored by Rapp and Goldrick (2000). I conclude that it is important for theoretical analyses of spoken word production to take the relationship among speech production, comprehension, and self-monitoring into account—even for the naming of single pictured objects.

Lexical Bias

Lexical bias is the finding that form errors create real words rather than nonwords with a frequency that is higher than would be expected by chance (e.g., Dell & Reich, 1981). Most form errors are nonword errors, but word outcomes tend to be statistically overrepresented. For example, in planning to say "cat," the error "hat" (a word in English) is more likely than the error "zat" (not a word in English). A lexical bias in speech errors is not always observed. Whereas Dell and Reich (1981) and Nooteboom (in press) found a strong lexical bias in their corpora of errors in spontaneous speech, Garrett (1976) and Del Viso, Igoa, and Garcia-Albea (1991) found no such effect and Stemberger (1985) found only a weak effect. In an analysis of the errors in picture naming of 15 aphasic speakers, Nickels and Howard (1995) found no evidence for lexical bias.

On the feedback account of lexical bias that Rapp and Goldrick (2000) favored (see Figure 1), form-related errors arise either because an incorrect lexical item was selected (e.g., *hat* instead of *cat*) and then properly encoded (yielding the malapropism "hat") or because the correct lexical item was selected (*cat*) but then the segments were misselected (e.g., /h/ for /k/, also yielding "hat"). The error bias is assumed to be due to activation spreading from shared segment nodes to lexical nodes (e.g., from the /æ/ node activated by the target *cat* back to *cat* and the competitors *calf, cap,* and *hat*) and from these lexical nodes to other segment nodes (i.e., from *hat* to /h/). This does not happen for nonwords because there are no lexical nodes for such items in the network (i.e., there is no node *zat* to activate /z/). Because of the activation feedback in planning to say "cat," *hat* has a higher probability of becoming erroneously selected than a word that is unrelated in form. In

addition, because there is no lexical node for *zat,* there is no chance that this item is selected. Furthermore, because of the lexical support for indirectly activated segments like /h/ in planning to say "cat," it is more likely that /h/ is erroneously selected (yielding the error "hat") than that /z/ is selected (yielding the error "zat"). By contrast, a feedforward-only speech production model has no backward links between segment nodes and lexical nodes that could give rise to such segmental influences on lexical selection and lexical influences on segment selection.

Contrary to what the feedback account implies, however, the lexical error bias is not an automatic effect, as already suggested by the seminal study of Baars, Motley, and MacKay (1975). In RIA, activation automatically spreads back from segment nodes to lexical nodes. However, when all the target and filler items in an error-elicitation experiment are nonwords, word slips do not exceed chance. Only when some words are included in the experiment as filler items does the lexical error bias appear (Baars et al., 1975). Similarly, lexical influences on segment processing in speech comprehension depend on the exact experimental context (see Norris et al., 2000, for a review). The effect of filler context should not occur with automatic feedback of activation. Therefore, Levelt (1983, 1989) and Levelt et al. (1999b), among others, have argued that lexical bias in speech production is not due to production-internal activation feedback but that the bias is at least partly due to self-monitoring of speech planning by speakers. When an experimental task exclusively deals with nonwords, speakers do not bother to attend to the lexical status of their speech plan (as they normally often do, apparently), and lexical bias does not arise. Similarly, Norris et al. (2000) assumed that when the use of lexical knowledge by listeners is not encouraged, lexical effects on segment processing in speech comprehension do not arise, accounting for the observation that lexical effects are dependent on the exact experimental situation rather than mandatory.

Rapp and Goldrick (2000) rejected a self-monitoring account of speech error biases by arguing that "not only does it require seemingly needless reduplication of information, but because the specific nature of the mechanism has remained unclear, the proposal is overly powerful" (p. 468). This critique may hold for certain monitoring proposals in the literature (see Postma, 2000, for a recent review) but certainly not for all. In particular, Levelt (1983, 1989) proposed that self-monitoring of speech planning and production is achieved through the speaker's speech comprehension system. Of course, self-monitoring is not equivalent to speech comprehension; it requires some additional cognitive operations. For example, lexical selection errors may be detected by verifying whether the "lemma" of the word recognized in inner speech corresponds to the lexical concept prepared for production, which is an operation specific to self-monitoring. Nevertheless, properties of the speech comprehension system, as assessed by comprehension research, independently constrain what the monitor can do. Self-monitoring via the speech comprehension system does away with the critique of "needless reduplication of information."

The assumption of comprehension-based self-monitoring has also been adopted for WEAVER++ (Levelt et al., 1999b). The idea of self-monitoring through the speech comprehension system is in agreement with the results from functional brain imaging studies, which suggest that self-monitoring and speech comprehension are served by the same neural structures (e.g., McGuire, Silbersweig, & Frith, 1996). Recently, Hartsuiker and

Kolk (2001) extended and formalized Levelt's (1989) comprehension-based theory of self-monitoring and repair as a computational model. Simulations showed that there was a good fit between the monitoring–repair model and empirical data concerning the distribution of time intervals between making an error and interrupting articulation, the distribution of intervals between interruption and making the repair, and the effect of speech rate on these intervals.

Planning and Self-Monitoring in WEAVER++

WEAVER++ (Word Encoding by Activation and VERification; Roelofs, 1992, 1997; see Levelt et al., 1999b, for a recent review) is a computationally implemented model of spoken word production. It implements word planning as a staged process, moving from conceptual preparation (including the conceptual identification of a pictured object in picture naming) via lemma retrieval (recovering the word as a syntactic entity, including its syntactic properties, crucial for the use of the word in phrases and sentences) to word-form encoding, as illustrated in Figure 2. Unlike RIA, WEAVER++ assumes two different lexical levels, namely levels of lemmas and morphemes (the latter representations are involved in word-form encoding), but this is not crucial for present purposes (see Levelt et al., 1999b, and Roelofs, Meyer, & Levelt, 1998, for a theoretical and empirical motivation of the distinction). Comprehending spoken words traverses from word-form perception to lemma retrieval and conceptual identification. In the model, concepts and lemmas are shared between production and comprehension, whereas there are separate input and output representations of word forms. Consequently, the flow of information between the conceptual and the lemma stratum is bidirec-
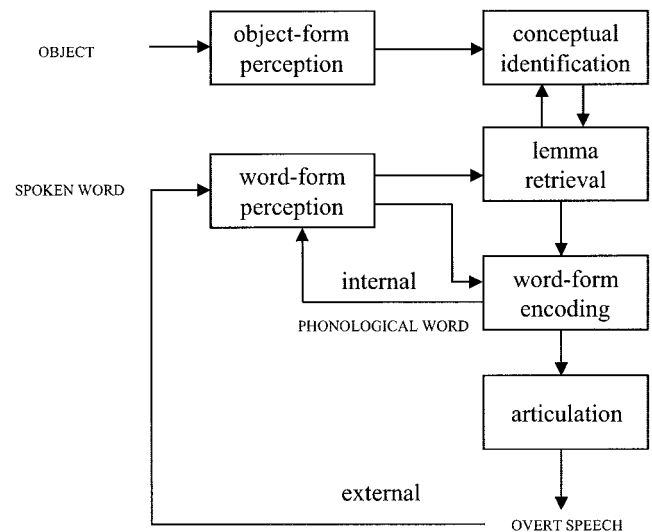
Figure 2. The flow of information in the WEAVER++ model during object naming. After lemma retrieval, spoken word planning happens in a strictly feedforward fashion, with feedback occurring only via the speech comprehension system. Internal monitoring includes feeding the rightward incrementally constructed phonological word back into the speech comprehension system, whereas external monitoring involves listening to the pronunciation of the word.

tional (Roelofs, 1992),[2] whereas it is unidirectional between lemmas and forms as well as within the form strata themselves (top-down for production and bottom-up for comprehension). After lemma retrieval in production, spoken word planning is a strictly feedforward process (Roelofs, 1997).

Following Levelt (1983, 1989), the model assumes two self-monitoring routes, an internal and an external one, both operating via the speech comprehension system. The external route involves listening to self-produced overt speech, whereas the internal route (which is assumed to be responsible for the error biases) includes monitoring the speech plan by feeding a rightward incrementally generated phonological word back into the speech comprehension system (Levelt et al., 1999b). A phonological word representation specifies the syllables and, for polysyllabic words, the stress pattern across syllables. So, even in extant feedforward speech production models, there exists feedback of activation from sounds to lexical items (see Levelt et al., 1999a, 1999b, for an extensive discussion of this point), except that the feedback engages the speech comprehension system rather than the production system itself.

On the account of monitoring via the speech comprehension system, there is no "needless reduplication of information" (Rapp & Goldrick, 2000, p. 468). The form representations in speech production differ from those in speech comprehension, but this is not needless because it serves speech production and comprehension functions. Furthermore, the reduplication is supported by empirical evidence. For example, it explains dissociations between production and comprehension capabilities in aphasia (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Nickels & Howard, 1995). Under the assumption that speech production and comprehension are accomplished via the same word-form network, network models—including RIA—exhibit a strong correlation between speech production and comprehension accuracy in aphasia (as assessed in computer simulations by Dell et al., 1997, and Nickels & Howard, 1995, for an RIA-type model), which is empirically not supported by form errors. Thus, Rapp and Goldrick (2000) also need to make the assumption of separate word-form representations serving speech production and speech comprehension for their model. In addition, because RIA has no backward links between lexical items and their meaning (i.e., semantic feature nodes), the model needs to duplicate lexical nodes for production and comprehension and to include backward links between one set of lexical nodes and the semantic feature nodes to achieve speech comprehension. In contrast, WEAVER++ does not reduplicate lemmas for production and comprehension purposes.

WEAVER++ implements a number of specific claims about how the spoken word production and comprehension networks are related, as shown in Figure 3. To account for interference and facilitation effects from auditorily presented distractor words on picture-naming latencies, Roelofs (1992, 1997; Roelofs et al., 1996) assumed that information activated in a speech comprehension network activates compatible segment, morpheme, and lemma representations in the production network (see also Levelt et al., 1999b). For example, lemmas activated by spoken distractors compete with lemmas activated by a pictured object for selection in picture naming, leading to interference. For convenience, Figure 3 shows segment and lexical form nodes in the comprehension network (following McClelland & Elman, 1986;

Norris, 1994), but this is not critical (see Lahiri & Marslen-Wilson, 1991, for a model of speech comprehension that has no such segments). Covert self-monitoring involves feeding over time the rightward incrementally constructed phonological word representation from speech production into the speech comprehension system (Levelt et al., 1999b).

One of the key observations from the research on spoken word recognition is that as speech unfolds, multiple word candidates become partially activated and compete for selection. For example, the speech fragment /kæ/ activates a cohort of words including *cat, calf, captain,* and *captive* (e.g., Marslen-Wilson & Welsh, 1978; Zwitserlood, 1989). The multiple activation concerns not only the forms but also the syntactic properties and meanings of the words. Computationally implemented models of spoken word recognition, such as Shortlist (Norris, 1994) and TRACE (McClelland & Elman, 1986), all instantiate this insight in one form or another. WEAVER++ assumes that both external and internal speech activate such "cohorts" of word candidates.

Lexical Error Bias in WEAVER++

A lexical error bias arises within WEAVER++ in three ways. First, the bias occurs when speakers use lexicality as one of their monitoring criteria, as suggested by Levelt and colleagues (Levelt, 1989; Levelt et al., 1999b); clearly, lexicality is only a superficial criterion. According to models of spoken word recognition (e.g., McClelland & Elman, 1986; Norris, 1994), word candidates become active and compete for recognition as their onset segments are activated. When mismatch occurs, words become deactivated, until a single candidate stands out from the rest and is thereby recognized. The latter will happen for words but not for nonwords. Therefore, word errors meet the lexicality criterion and pass, whereas nonword errors will be detected on this basis.[3]

On this view of activation and competition in spoken word recognition and production, the speed and accuracy of confirming the lexical status of a perceived spoken word during planning to speak a word should be lower when the production and recognition targets are similar, compared with when they are not. This predic-

---

[2] WEAVER++ has backward links from lexical items (i.e., lemmas) to their meaning, implied by the assumption that lemmas and lexical concepts are shared between speech production and comprehension. Furthermore, activation cascades from meanings to lexical items in the model. In contrast, Rapp and Goldrick (2000) applied the discreteness and cascading assumptions to all lexical levels of the speech production system in their simulations.

[3] Nooteboom (in press) argued against this account on the basis of the observation that although spontaneous speech errors showed a strong lexical bias in his error corpus, there was no evidence that upon actually making errors, nonword errors were more often corrected than real-word errors. However, using a lexicality criterion in internal monitoring does not imply that overt nonword errors are more often repaired than real-word errors (cf. Levelt, 1989). First, self-monitoring is attention demanding. Therefore, it is unlikely that speakers can attend simultaneously to all aspects of their speech and at the same time equally well to their internal and external speech. Instead, monitoring internal speech (yielding the error bias) may have as a result that external speech can be monitored less well (causing the absence of a repair bias). Second, even if speakers would detect all errors in their external speech, they may not bother to correct each and every error.
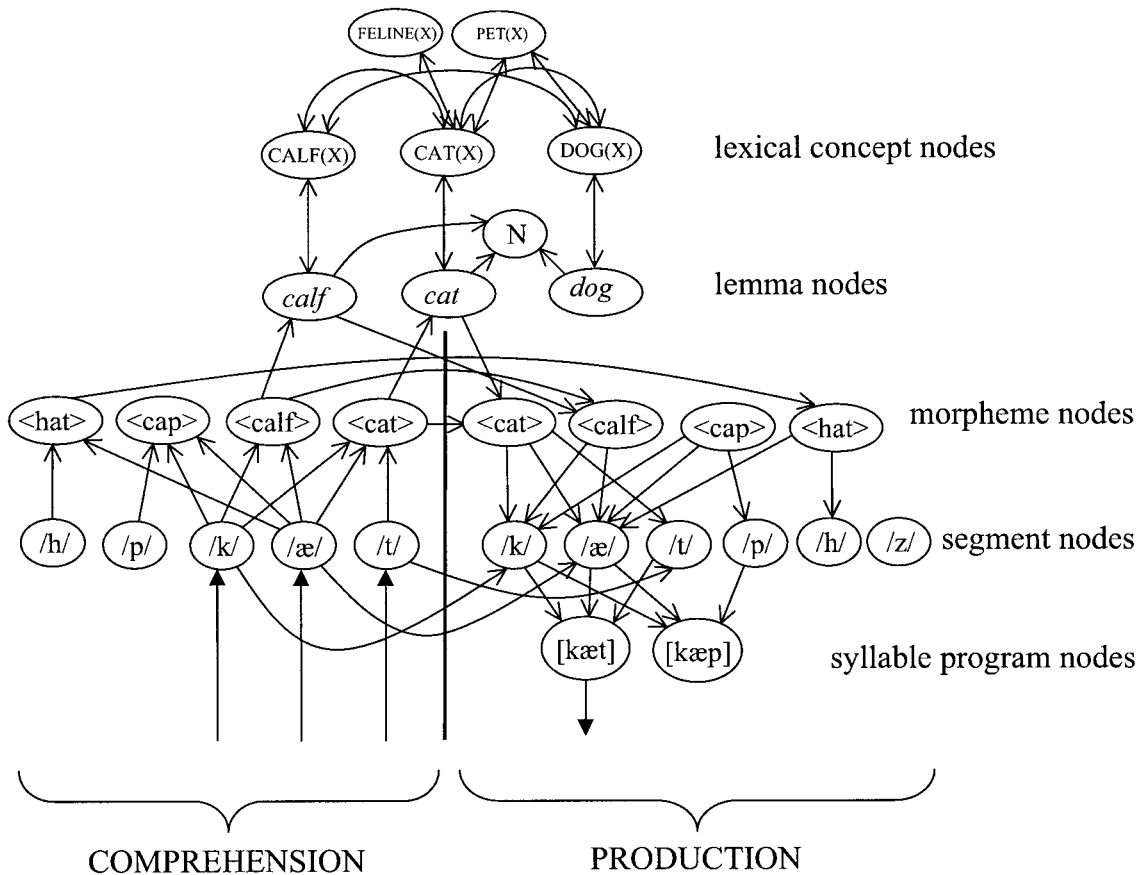
Figure 3. Illustration of the speech production and comprehension networks of the WEAVER++ model. Lexical concept nodes (e.g., CAT(X)) make up a conceptual stratum; lemma nodes (e.g., *cat*) and syntactic property nodes (e.g., N = noun) make up a syntactic stratum; and morpheme (e.g., <cat>), segment (e.g., /k/), and syllable program nodes (e.g., [kæt]) make up a form stratum. The form stratum differentiates between speech production and comprehension, whereas the syntactic and conceptual strata are shared.

tion has been confirmed in a study by Levelt et al. (1991b), who combined picture naming with auditory lexical decision. Participants were asked to name pictured objects, and on some critical trials, they had to make a lexical decision by means of a keypress to an auditory probe presented shortly after picture onset. Thus, the speakers had to monitor for the lexical status of spoken probes while preparing to speak the name of the object. Compared with unrelated word probes, monitoring responses were slower and more error prone for word probes that were semantically related, phonologically related, or even identical to the picture name. Levelt et al. (1991b) showed that there was a good fit between a mathematical model implementing the claims about the effect of similarity on competition in spoken word recognition and production and the empirically observed lexical decision latencies.

Second, lexical bias arises during self-monitoring in WEAVER++ when some of the form-related errors are due to morpheme or lemma selection failures (i.e., if they are malapropisms) rather than segment selection failures. RIA also allows for lexical bias in segment selection and for malapropisms, except that the errors are due to production-internal feedback rather than to self-monitoring, as is the case in WEAVER++. As Levelt et al. (1999a) put it,

A malapropism may occur when a speaker can generate only an incomplete form representation of the intended word (as in a TOT [tip-of-the-tongue state]). This incomplete form is fed back to the conceptual system via the comprehension system, which leads to the activation of several lemmas that are phonologically related to the target. These lemmas typically will be semantically unrelated to the target. If one of the form-related lemmas of the appropriate grammatical category is selected, a malapropism will occur. (p. 64)

On this view, interactivity is a result of trouble shooting rather than the source of the error.

Third, lexical bias occurs in WEAVER++ during the access of motor programs for syllables (i.e., syllable program nodes in the WEAVER++ network). Comprehension-based feedback for self-monitoring activates compatible morpheme representations in the production network, which activate corresponding segment and syllable program nodes (in WEAVER++, activation cascades through the form network; see also Footnote 1). Consequently, the loop favors the selection of syllable program nodes that correspond to words. The production system retrieves the syllable program nodes at the same time the segments are going through the internal loop, and the spreading of activation from production nodes to

comprehension nodes back to production nodes takes time. However, program node retrieval latencies are variable and vary with speaking rate. As a result, in case of long retrieval latencies or slow speaking rate, it is possible that activation may have cycled through the comprehension system before a syllable program node is selected, which gives room for the lexical influence. Thus, one expects that lexical bias occurs only with slow speaking rate, when there is sufficient time for activation to cycle through the comprehension system, which has been confirmed empirically (Dell, 1986). Note that in a context that does not encourage self-monitoring (i.e., feeding the phonological word back into the comprehension system, which is presumed to be under a speaker's control) and the use of lexical knowledge, such as the all-nonwords condition of Baars et al. (1975), a lexical error bias should not occur, in agreement with the empirical findings.

In conclusion, in a model without backward links from segments to lexical items such as WEAVER++, there are several factors that give rise to a tendency to produce word over nonword errors at a higher rate than would be predicted by chance.

Mixed Error Bias

When the word *cat* is intended by a speaker, the substitution of *calf* for *cat* is more likely than the substitution of *dog* for *cat* if error opportunities are taken into account (Dell & Reich, 1981; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996). The error bias is often called the *phonological facilitation of semantic substitutions* or the *mixed error effect* because the erroneous words (e.g., *calf* for *cat*) share both semantic and phonological properties with the target word. On the feedback account (Dell & Reich, 1981), implemented in RIA, the mixed error bias happens because of production-internal activation feedback from segment nodes to lexical nodes. Semantic substitution errors may arise as failures in lexical node selection (Dell & Reich, 1981). The word *calf* shares segments with the target *cat.* Therefore, in planning to say "cat," the lexical node of *calf* receives feedback from these shared segments (i.e., /k/ and /æ/), whereas the lexical node of *dog* does not (see Figure 1). Consequently, the lexical node of *calf* has a higher level of activation than the lexical node of *dog* (assuming that calves and dogs are equally catlike, conceptually), and *calf* is more likely than *dog* to be involved in a lexical node selection error resulting in a word substitution.

The mixed error effect is not strong evidence for feedback. Rapp and Goldrick (2000) showed that the effect occurs at the segment rather than at the lexical level in a feedforward-only cascading version of their model. The mixed error effect is then a semantic effect on phonological errors rather than a phonological effect on semantic substitution errors. Similarly, Levelt et al. (1999b) argued that the mixed error effect occurs in WEAVER++ when erroneously both a target and a coactivated intruder are selected in lemma retrieval. In a cascading model, activation automatically spreads from one level to the other, whereas in a discrete multiple-output model, the word-form activation is restricted to the selected items. The mixed error bias occurs in WEAVER++ because the form of a target like *cat* primes the form of an intruder like *calf* but not of an intruder like *dog*. Of course, *calf* will also prime *cat,* but when only a target gets "signaling activation" (cf. Dell, 1986), *cat* will prime *calf* more than vice versa (in RIA, nodes are selected by giving them a jolt of activation, whereas in WEAVER++, selec-

tion and providing signaling activation to a target are different operations). Therefore, *calf* has a higher chance than *dog* to be produced instead of *cat.* The assumption of multiple output underlying certain speech errors is independently motivated by the occurrence of blends, like a speaker's merging of the near synonyms *close* and *near* into the error "clear." Levelt et al. (1999b; cf. Roelofs, 1992) argued that blends reflect a speaker's indecisiveness: Two lemmas are selected as targets because they are equally appropriate in the context of use. The two forms may combine, yielding a blend (no error occurs when one of the forms is encoded slower than the other or is filtered out through self-monitoring). Dell and Reich (1981) observed the mixed error bias also for blends. Thus, contrary to what Rapp and Goldrick claimed, the mixed error effect in normal speakers can be accounted for by models such as WEAVER++.

Levelt et al. (1999b) argued that a mixed error bias is also intrinsic to self-monitoring. Monitoring is attention demanding and not faultless. It has been estimated that speakers miss about 50% of their errors (Levelt, 1989). As indicated, the probability of detecting an error should be greater when target and error differ greatly than when they are similar. In planning to say "cat" and monitoring through the speech comprehension system, the lemma of the target *cat* is in the speech comprehension cohort of an error like *calf* (when its form is fed back through comprehension), whereas the lemma of the target *cat* is not in the speech comprehension cohort of the error *dog*. Furthermore, the mismatch between target (*cat*) and error is apparent at the beginning of a word like *dog* but only later for a cohort member like *calf*. Consequently, if the lemma of *calf* is erroneously selected for the concept CAT(X), there is a higher chance that the error remains undetected during self-monitoring than when the lemma of *dog* is erroneously selected: a mixed error bias in self-monitoring.

In conclusion, in a feedforward model like WEAVER++, there are several factors that give rise to a tendency to produce mixed errors at a higher rate than would be expected by chance. Thus, the claim of Rapp and Goldrick (2000) that the mixed error bias argues against models like WEAVER++ is not warranted.

Latency Effect of Mixed Distractors

Although Rapp and Goldrick (2000) did not use latency data in their evaluation of the five theoretical positions, they referred to a number of chronometric studies that, according to them, "provide support for cascading activation and/or feedback" (p. 469). Most relevant for the issue of feedback are the latency data and theoretical analyses of Damian and Martin (1999) and Starreveld and La Heij (1995, 1996). Their studies showed that presenting mixed items as distractors in picture naming yields a naming latency benefit compared with distractors that are semantically related only, which yield interference compared with semantically unrelated distractors. The phonological reduction of semantic interference has been taken as evidence for production-internal feedback. However, Rapp and Goldrick's feedback account of the error biases appears to be incompatible with a feedback account of the chronometric findings. According to the feedback account of the error biases exemplified by RIA, mixed items yield more competition in lexical selection than items that are semantically related only, because of the production-internal activation feedback from segment nodes to lexical nodes during the planning of a word.

Consequently, mixed items are involved in a larger number of speech errors. However, exactly opposite to this, the production latency measurements have shown a latency benefit from mixed items, which suggests that production-internal activation feedback gives rise to less competition in lexical selection between a target and a mixed item than between a target and an item that is semantically related only.

Starreveld and La Heij (1995, 1996) conducted picture–word interference studies with written distractor words. Participants had to name pictured objects while simultaneously trying to ignore distractor words superimposed on the pictures. Naming latency was the main dependent variable of interest. The distractors yielded semantic and form effects, and together the effects interacted. Damian and Martin (1999) replicated these results using spoken distractor words. For example, the naming of a picture of a cat was interfered with by the semantically related distractor *DOG* compared with the semantically unrelated distractor *DOLL,* and the naming was facilitated by the phonologically related distractor *CAP* relative to the phonologically unrelated distractor *DOLL.* The semantic interference effect was smaller when target and distractor were phonologically related (distractors *CALF* vs. *CAP*) than when they were unrelated in form (distractors *DOG* vs. *DOLL*): a phonological reduction of semantic interference or a *mixed distractor* latency effect.

According to Damian and Martin (1999), the interaction between phonological and semantic relatedness suggests that production-internal activation feedback from sounds to lexical items exists. The claim is that, because of the production-internal feedback of activation, mixed distractors yield less competition and hence shorter latencies than distractors that are semantically related only. However, Rapp and Goldrick's (2000) claim is that, because of the production-internal feedback of activation, mixed items yield more competition and hence a larger number of errors than items that are semantically related only. For example, in RIA, the mixed error effect occurs because feedback of activation from segment nodes to lexical nodes makes *calf* a stronger competitor than *dog* in planning to say "cat," which is exactly opposite to what has been proposed to explain the latency effect of mixed distractors. The latency data suggest that *calf* is a weaker competitor than *dog* in planning to say "cat."

There is yet another problem for the claim that the chronometric data provide support for production-internal feedback of activation from segments to lexical items. In particular, the chronometric data suggest that the reduction of semantic interference (observed with distractor *CALF* in naming a cat) and the form effect (observed with distractor *CAP* in naming a cat) do not inevitably go hand in hand. On the production-internal feedback account, semantic and form relatedness interact because activation of production segments spreads back to the level at which semantic effects arise, namely the level of lexical selection. Therefore, a reduction of semantic interference for mixed distractors should be observed only in the context of facilitation from form-related distractors. However, this is not supported empirically. Damian and Martin (1999) presented the spoken distractors at three stimulus onset asynchronies (SOAs). The onset of the spoken distractor was 150 ms before picture onset (SOA = −150 ms), simultaneously with picture onset, or 150 ms after picture onset. They observed semantic interference at the SOAs of −150 and 0 ms and form facilitation at the SOAs of 0 and 150 ms. The mixed distractors yielded

no effect at SOA = −150 ms and facilitation at the later SOAs, exactly like the distractors that were related in form only. Thus, the reduction of semantic interference for mixed distractors was already observed at an SOA (i.e., SOA = −150 ms) at which there was no pure form facilitation effect.[4] The temporal dissociation between the mixed semantic–phonological effect and the pure phonological effect suggests that the effects arise at different planning levels, namely the lemma and the word-form level, and that the interaction is a *perceptual cohort* effect rather than one that is due to production-internal activation feedback, as argued by Roelofs et al. (1996).

Of course, Rapp and Goldrick (2000) could adopt the perceptual cohort account of the mixed distractor latency effect proposed by Roelofs et al. (1996). However, this would mean that the chronometric findings are no longer taken as support for production-internal feedback, contrary to what Rapp and Goldrick suggested (p. 469). Moreover, it then needs to be shown that the latency effects may arise because of perceptual cohorts (causing *calf* to be a weaker competitor than *dog* in planning to say "cat") in spite of the presence of production-internal feedback (causing *calf* to be a stronger competitor than *dog* in planning to say "cat"). Furthermore, if the perceptual cohort account is adopted to explain the mixed distractor latency effect, the question comes up whether the error biases may arise from perceptual cohorts (generated by internal rather than external speech) too, which would make the assumption of production-internal feedback superfluous. In the next section, I demonstrate by means of WEAVER++ simulations that perceptual cohorts may explain both the error biases and the latency effects.

Accounting for Both the Error and Latency Findings

The mixed error effect can at least partly be attributed to self-monitoring in WEAVER++. If in planning to say "cat," the lemma of *calf* is selected instead of the lemma of *cat* and the corresponding encoded form is fed back through the speech comprehension system, the lemma of *cat* is in the comprehension cohort of the error *calf*. However, if in planning to say "cat," the lemma of *dog* is selected instead of the lemma of *cat,* the lemma of *cat* is not in the cohort of *dog*. Hence, the lemma of *cat* is more active when activation from the word form of the error *calf* is fed back via the comprehension system than when activation from the form of the error *dog* is fed back, and the error *calf* for *cat* is more likely to be overlooked in self-monitoring than the error *dog* for *cat.*

---

[4] Damian and Martin (1999) obtained a by-participants effect of phonological relatedness (i.e., form and mixed versus semantic and unrelated) at SOA = −150 ms. However, this effect appeared to be fully due to the semantically related distractors. Pairwise comparisons revealed that the semantically related distractors increased picture-naming latencies compared with the form, mixed, and unrelated distractors, which did not differ among each other. Compared with the unrelated distractors, the form effect was 5 ms, and the effect of form and mixed distractors combined was 2 ms. The point here is not that form related distractors may not yield facilitation at SOA = −150 ms (e.g., Meyer & Schriefers, 1991, obtained such an early effect, whereas Schriefers et al., 1990, and Damian & Martin, 1999, did not) but that there may be a temporal dissociation between mixed effects and phonological effects.

On this account, the lemma of *calf* is a weaker competitor than the lemma of *dog* in self-monitoring during planning to say "cat" (i.e., *cat* and *dog* are closer in activation than *cat* and *calf*). That *calf* is a weaker competitor than *dog* in planning to say "cat" also accounts for the mixed distractor latency effect in WEAVER++ (cf. Roelofs et al., 1996), except that the effect results from the comprehension of speech of others (i.e., spoken distractor words in selecting *cat* for production) rather than from comprehension-based self-monitoring (after having misselected *calf* or *dog* for production). The mixed distractor *CALF* yields less interference than the distractor *DOG* because the lemma of the target *cat* is primed as a spoken cohort member of the distractor *CALF* but not of the distractor *DOG* (so that *cat* and *dog* are closer in activation than *cat* and *calf*). Thus, WEAVER++ explains why there is both a reduction of semantic interference and a heightened change of misselection for mixed items. In summary, according to WEAVER++ mixed items are weaker competitors rather than stronger competitors compared with semantically related items only, exactly opposite to what RIA maintains (in RIA, production-internal feedback gives rise to the situation in which *cat* and *calf* are closer in activation than *cat* and *dog*). Therefore, selection failures concerning mixed items are more likely to remain unnoticed during error monitoring in WEAVER++, and mixed items have a smaller effect as spoken distractors on latencies, in agreement with the speech error data and the production latency findings.

Rapp and Goldrick (2000) argued that a role of self-monitoring in the error biases cannot be taken seriously "because the specific nature of the mechanism has remained unclear" (p. 468). However, because the effect of spoken distractors has been simulated with WEAVER++ (Roelofs, 1997) and self-monitoring is accomplished via the speech comprehension system, the required mechanism is already partly present in the model. Technically, self-monitoring in WEAVER++ is like comprehending a spoken distractor word presented at a long postexposure SOA, except that the spoken word is self-generated. Below, I report the results of WEAVER++ simulations that included self-monitoring through comprehension-based feedback. The simulation results demonstrate that a single underlying mechanism (i.e., a perceptual cohort mechanism) suffices to explain the mixed distractor latency effect (arising from perceptual cohorts in processing speech produced by others, i.e., spoken distractors), the mixed error effect (arising from perceptual cohorts in the monitoring of self-generated inner speech), and the influence of impairment locus in aphasic speakers (arising from the structure of the lexical network in self-monitoring). The Appendix gives the details of the simulations. Following Rapp and Goldrick, I have attempted to capture the gist of the data rather than to provide a quantitative account (cf. Rapp & Goldrick, 2000, pp. 492–493).

Simulation of the Mixed Distractor Latency Effect

Figure 4 demonstrates that WEAVER++ exhibits the latency effect of mixed distractors in picture naming. The figure plots the effects of semantic, form, and mixed distractors on picture-naming latencies relative to unrelated distractors against SOA. A positive difference score indicates interference, and a negative score indicates facilitation relative to unrelated distractors. The solid lines in Figure 4A show the empirical latency results of Damian and

Martin (1999). At SOA = −150 ms, the semantically related distractor *DOG* yields interference in planning to say "cat," but the mixed distractor *CALF* yields no interference, even though there is no pure form facilitation from *CAP* at this SOA. Figure 4B shows the effect that the auditory distractors have on lemma retrieval in WEAVER++. At SOA = −150 ms, the distractor *DOG* yields semantic interference in planning to say "cat." But, the mixed distractor *CALF* yields no semantic interference, even though there is no pure form facilitation from *CAP* at this SOA. Thus, form relatedness may affect lemma retrieval in case of semantic relatedness even when it yields no pure form facilitation at this level because of a floor effect in speeding up the retrieval process. However, when lemma retrieval is slowed down by a semantically related distractor, form relatedness can have its facilitatory influence (see Levelt et al., 1999a; Roelofs et al., 1996). This accounts for the temporal dissociation between the mixed semantic–phonological and the pure phonological effects observed by Damian and Martin (1999). Figure 4C shows the effects of the distractors after both lemma retrieval and word-form encoding in the model are completed. There are effects of semantic and form relatedness, and together the effects interact, as empirically observed (the correlation between model and data is .93, $N = 9$, $p < .001$). The results of the simulations are identical for conditions with and without comprehension-based feedback. Thus, including self-monitoring in WEAVER++ does not affect earlier published fits of the model (e.g., Roelofs, 1992, 1996, 1997, 2003).

There is no semantic effect at SOA = 0 ms in WEAVER++, but this is not excluded. Whether semantic interference occurs at this SOA depends on the exact temporal parameters of spoken word recognition in the simulations (e.g., the exact materials used, which differs between experiments). Schriefers, Meyer, and Levelt (1990) also observed no semantic interference at SOA = 0 ms (the dashed lines in Figure 4A indicate their results). The simulated patterns correspond to what Schriefers et al. (1990) observed (the correlation between model and data is .98, $N = 6$, $p < .001$). As concerns the pure semantic effect, there exists a small empirical difference between the results of Schriefers et al. (1990) and Damian and Martin (1999), which may be investigated in future research. Important for this article, however, is that the reduction of semantic interference for mixed distractors may be observed at an SOA at which there is no pure phonological facilitation, both empirically and in WEAVER++.

Simulation of the Mixed Error Bias and the Effect of Impairment Locus

Self-monitoring via the speech comprehension system requires cognitive operations to detect discrepancies between selections made in production and comprehension. Lexical selection errors may be detected by verifying whether the lemma of the recognized word is linked to the target lexical concept in production. In addition, errors in lexical concept selection may be detected by verifying whether the lexical concept of the recognized word is linked to the conceptual information derived from the to-be-named object. WEAVER++ implements such verification operations by means of condition-action production rules (see Roelofs, 2003, for an extensive discussion of the production rule component of the model and how it plays a role in verbal control). Errors occur when production rules mistakenly fire. The probability of firing by
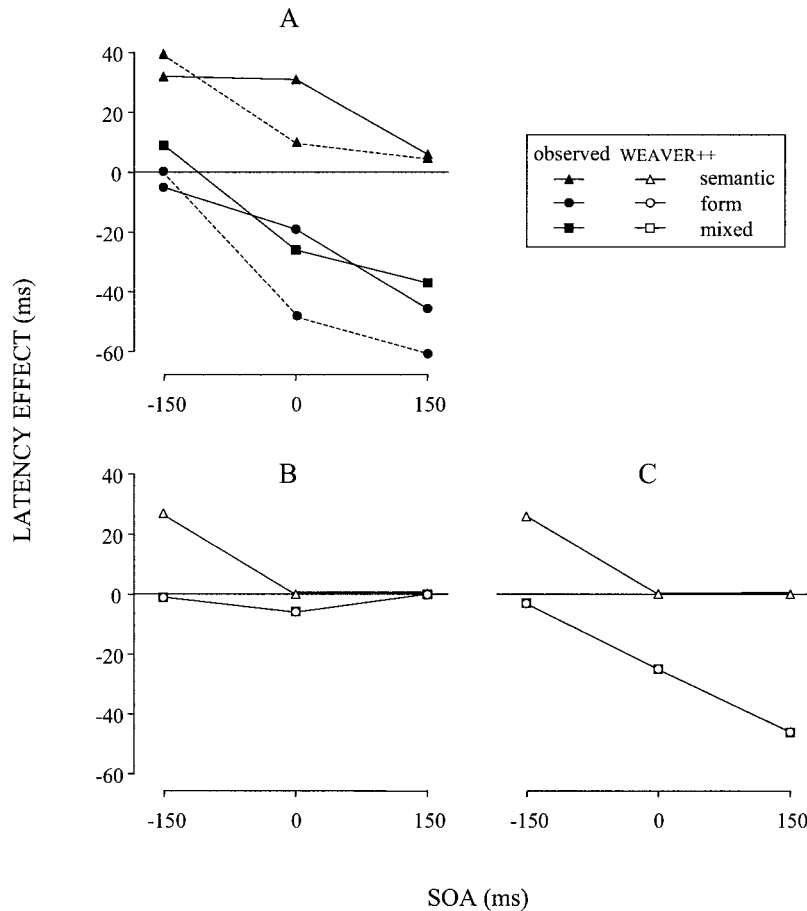
Figure 4. The latency effect (in milliseconds) of semantically related, form related, and mixed spoken distractors relative to unrelated spoken distractors in picture naming as a function of stimulus onset asynchrony (SOA; in milliseconds). A: Solid lines show the empirical findings of Damian and Martin (1999), and the dashed lines show the findings of Schriefers et al. (1990). B: The effects in WEAVER++ after completing lemma retrieval only. C: The effects in WEAVER++ after completing both lemma retrieval and word-form encoding.

mistake is a function of activation differences among nodes (cf. Roelofs, 1992, 1997, 2003).

Figure 5 demonstrates that self-monitoring in WEAVER++ suffices to explain the mixed error effect and its dependence on the functional locus of damage in aphasic speakers. The lower panel shows the influence of comprehension-based feedback on the activation levels of the lemmas of *cat, calf,* and *dog.* When the lemma of *calf* is erroneously selected and monitored via the speech comprehension system (i.e., through the internal monitoring loop), the activation level of the lemma of *cat* is increased because of the form overlap with *calf.* However, when *dog* is fed back, the activation of the lemma of *cat* is not increased. As a result, the difference in activation between the lemmas of *cat* and *calf* is greater than between the lemmas of *cat* and *dog.* Consequently, verification is more likely to fail with the error *calf* than with the error *dog.* The upper panel of Figure 5 shows the influence of comprehension-based feedback on the activation levels of the lexical concepts CAT(X), CALF(X), and DOG(X). The figure shows that the activation of CAT(X) is not much affected by whether a form-related or unrelated item is fed back via the speech

comprehension system. An effect is also not observed earlier or later in time. Thus, the probability of verification failures at the conceptual level should not differ between the error *calf* and the error *dog.* Thus, the mixed error effect in WEAVER++ is likely to arise at the level of lemmas but not at the level of lexical concepts.

Although there are positive connections from lemmas to lexical concepts in the model, the activation differences among the lemmas *cat, calf,* and *dog* are not fully carried over to the concept level for two main reasons. First, only a proportion of the activation of lemmas spreads to the lexical concept level, which reduces absolute activation differences. Second, because lexical concept nodes are part of a conceptual network with bidirectional connections among the concept nodes, small activation differences at this level are rapidly "washed away." For example, the activation difference between the lemma node of *cat* with *calf* fed back and the lemma node of *cat* with *dog* fed back was 0.69 at 325 ms, which is a difference of about 6%. However, the activation difference between the concept node CAT(X) with *calf* fed back and the concept node CAT(X) with *dog* fed back was 0.09 at 325 ms,
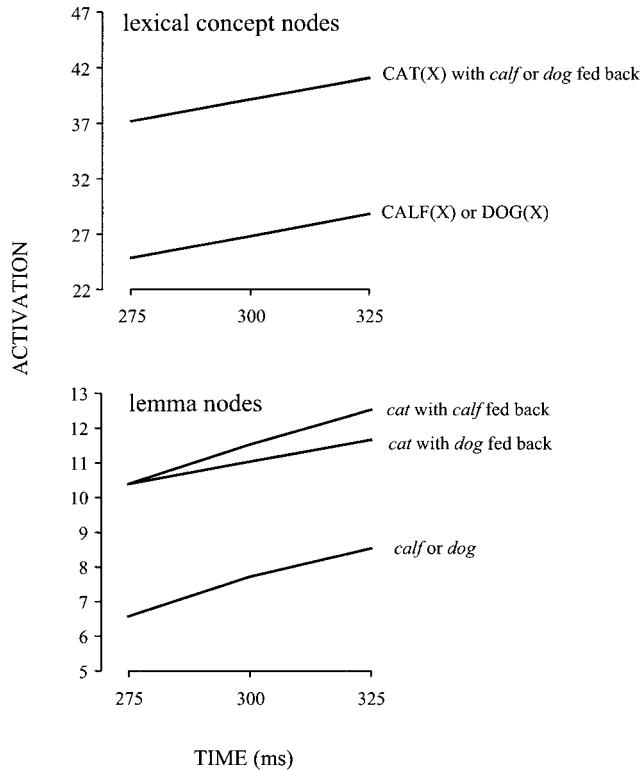
Figure 5. Activation levels of the lexical concept nodes CAT(X), CALF(X), and DOG(X) (top) and the lemma nodes of *cat, calf,* and *dog* (bottom) as a function of comprehension-based feedback of activation in WEAVER++.

which is only a difference of about 0.2%. Thus, although feeding back the forms of *calf* and *dog* yields activation differences at both the lemma and concept levels, the differences at the concept level are much reduced compared with those at the lemma level.

Consequently, erroneous selection of a lexical concept node in naming a picture due to damaged functioning at the conceptual level (yielding the pattern of only semantic errors from a conceptual level lesion, as observed with patient K.E.) should have about an equal chance of being detected when the erroneous concept has a form-related (*calf*) or a form-unrelated (*dog*) name. In contrast, erroneous selection of a lemma for a correctly selected lexical concept due to damaged functioning at the lemma level (yielding the pattern of only semantic errors from a postconceptual or lexical level lesion, as observed with patients P.W. and R.G.B.) should have a greater likelihood of being intercepted when the erroneously selected word has a form-unrelated (*dog*) than a form-related (*calf*) name. Thus, the presence of a mixed error bias in WEAVER++ depends on the locus of the lesion: The bias is present with postconceptual damage but absent with conceptual damage, in agreement with the empirical observations by Rapp and Goldrick (2000).

Note that WEAVER++ captures the differential effect of form relatedness according to the impairment locus in aphasia despite the presence of backward links from lemmas to lexical concepts in the model. In contrast, a critical claim of Rapp and Goldrick (2000; the feature that makes RIA different from other interactive models,

such as that of Dell et al., 1997) is that the empirically observed effect of impairment locus in aphasia refutes functionally effective backward links from lexical items to their meaning. Thus, the WEAVER++ simulations demonstrate that this claim by Rapp and Goldrick is not warranted.

Cohort Effects on the Speech Error Biases

If the speech error biases arise partly because of self-monitoring that is accomplished through the speech comprehension system, cohort effects on errors are to be expected. In particular, initial segment overlap between target and error should be critical, as suggested by the research on spoken word recognition. Speech comprehension studies have shown that when the first segments of two items differ in more than one or two phonological features (such as voicing), the items are not spoken cohort members of each other (e.g., Connine, Blasko, & Titone, 1993; Marslen-Wilson, Moss, & van Halen, 1996; Marslen-Wilson & Zwitserlood, 1989). In an eye-tracking study, Allopenna, Magnuson, and Tanenhaus (1998) observed that the probability of fixating, for example, a visually presented target dollar was less affected by hearing *COLLAR* (a rhyme competitor of *dollar,* the name of the object) than by hearing *DOLPHIN* (a cohort competitor of *dollar*). This suggests that cohort competitors are more activated than rhyme competitors in spoken word recognition.

Dell and Reich (1981) indeed observed that the mixed error effect for semantic substitutions in their corpus of spontaneous speech errors was strongest for first segment overlap and less strong for second, third, or fourth segment overlap. The same held for blends. Martin et al. (1996) replicated this serial order effect on mixed errors in picture naming, both with normal and aphasic speakers. Is a seriality effect also observed for lexical bias? Most phonological errors concern the initial segments of words, and therefore, Dell and Reich (1981) examined only initial consonant errors in their analyses of lexical bias. So, these data are inconclusive. However, lexical bias may arise partly because some phonological errors are actually lemma or morpheme selection failures resulting in a form-related lexical substitution (malapropism). Thus, on the self-monitoring account of lexical bias, the cohort effect should hold for malapropisms, which was indeed observed by Dell and Reich (1981).

The serial order effect on the error biases does not occur in RIA. Shared segments activate lexical nodes in the network independent of the serial position of the segments within the word. Of course, implementations should not be confused with theory. The RIA model was unconcerned with a possible serial order inherent in the word-form encoding process. Perhaps this shortcoming may be remedied by giving up the assumption that all connection weights are equal and by assuming that onset segments are more strongly connected to lexical nodes than later segments of a word are. Whether this suffices to produce the serial order effect on the errors remains to be shown by computer simulations.

Concluding Remarks

Speakers listen to their own talking, and they correct many of the errors made in the planning and production of speech. To account for this type of output control, researchers have made feedback that serves self-monitoring of speech planning and pro-

duction an integral part of feedforward models of spoken word production for two decades (e.g., Levelt, 1983, 1989; Levelt et al., 1999b). Furthermore, the latency and error data on spoken word production that have been collected over the past few decades are intricate, and therefore not surprisingly, even the simplest models that have been developed to provide a comprehensive account of these data (e.g., WEAVER++) are more complex than the particular models that Rapp and Goldrick (2000) evaluated (see Levelt et al., 1999a, 1999b). In evaluating the various theoretical positions, Rapp and Goldrick have concentrated on speech errors while paying almost no attention to latency data. Moreover, they have evaluated newly constructed feedforward models, even though implemented feedforward models exist in the literature. In particular, Rapp and Goldrick have evaluated the class of feedforward models by studying the effect of manipulating interaction parameters in a specific connectionist model (i.e., a feedforward variant of the RIA model). Not only does this connectionist model not have the explanatory power of existing feedforward models like WEAVER++, but it also leaves out self-monitoring.

Computational examinations such as the study of Rapp and Goldrick (2000) are important but also limited in that not much can be concluded in general from studying specific models. A refutation of a whole class of models requires a mathematical proof or an exploration of the entire parameter space of a fully representative model, neither of which is provided by Rapp and Goldrick. In contrast, the refutation of a general claim about a class of models, like the claim of Rapp and Goldrick that feedforward models cannot account for the error data, requires only one counterexample from that class, which is the logic followed in this commentary. Thus, Rapp and Goldrick's computational exercise can best be construed as an exploration of which variant of a specific model best accounts for the data. Thus construed, the RIA model that Rapp and Goldrick favor accounts for the error data considered, but it cannot be excluded that other extant models can also account for these data. To assess this, one has to analyze these existing models rather than a newly constructed model that is presumed to represent the whole class. In this article, I have shown that a feedforward model like WEAVER++ can account for both the error data and the production latency findings. The challenge for Rapp and Goldrick is to show that restricted-interactive models like RIA can do the same.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

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,* 382–439.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32,* 193–210.

Damian, M. K., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 345–361.

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

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.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104,* 801–838.

Del Viso, S., Igoa, J. S., & Garcia-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20,* 161–185.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30,* 412–431. (Original work published 1868)

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–255). Amsterdam: North-Holland.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42,* 113–157.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38,* 243–294.

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., Roelofs, A., & Meyer, A. S. (1999a). Multiple perspectives on word production. *Behavioral and Brain Sciences, 22,* 61–75.

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

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991a). Normal and deviant lexical processing: A reply to Dell and O'Seaghdha. *Psychological Review, 98,* 615–618.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991b). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1376–1392.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 576–585.

Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S., & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes, 11,* 257–282.

McClelland, J. L., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain, 119,* 101–111.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1146–1160.

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring, and lexicality. *Cortex, 31,* 209–237.

Nooteboom, S. G. (in press). Listening to oneself: Monitoring speech production. In R. Hartsuiker, Y. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech.* Hove, England: Psychology Press.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in

speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299–325.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77,* 97–131.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107,* 460–499.

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

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language, 35,* 854–876.

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

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

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 246–251.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma-lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition, 69,* 219–230.

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

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 686–698.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 896–918.

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

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

Appendix

Details of the WEAVER++ Simulations

Picture naming was simulated with the production of a single consonant–vowel–consonant target represented in a network with approximately the structure of the networks used by Dell et al. (1997), Nickels and Howard (1995), and Starreveld and La Heij (1996). The target was *cat* and the neighbors were *calf* (mixed), *dog* (semantically related), *doll* (form related to a semantic alternative, namely *dog*), and *cap* (form related). WEAVER++ is described in detail in many other places, and I refer to these sources for a detailed description of the model and its simulation (e.g., Levelt et al., 1999b; Roelofs, 1992, 1996, 1997, 2003).

The current simulations used procedures and parameter values that were exactly the same as in earlier simulations of the model. A few small parameter adjustments were made to fine-tune the current fits. The response threshold at the lemma level was set at 1.5, and the correction for the internal SOA at −150 ms. The duration of a speech segment in a perceived spoken distractor was set to 100 ms. The segment duration in perceiving an internally generated phonological word representation was set to 50 ms on the basis of empirical evidence suggesting that perceiving internal speech runs about twice as fast as perceiving external speech (Levelt et al., 1999b), but this is not critical. Because lemma retrieval and word-form encoding have varying rather than fixed latencies in the model, Figure 5 gives only one specific set of activation curves after internal comprehension-based feedback, namely feedback starting 300 ms after picture onset. Earlier or later feedback gives equivalent curves, except that the curves are shifted backward or forward in time.