trol operations into the model. We must address questions that require psycholinguists to become attention theorists, attention theorists to become psycholinguists, and so forth. In what follows, I want simply to say a few words about the varieties of attention that might ultimately need to be implanted in order to fix WEAVER'S attention deficit disorder.

To a student of attention, the universe of tasks examined by Levelt et al. is like Disneyland to an 8-year-old. Things begin simply enough. Meeting Levelt et al.'s task demands requires selective attention to one stimulus rather than the other. The picture and the word both provide visual information, but only one of these sources of information will support a correct response. How can this selection be accomplished?

It could be spatial. Picture and word do not occupy exactly the same points in space, so implanting a "posterior attention system" (Posner & Raichle 1994; Rafal & Henik 1994) that selects inputs from particular spatial locations might do the job. However, simply picking a location and increasing the gain on all information coining from that location will not be sufficient, especially when it is the picture that must be selected. Because the word is superimposed on the picture, selecting the region of space in which the picture can be found will get the word, too - just what the participant needs to avoid.

Perhaps the problem could be solved by *not* being spatial. Under some circumstances, the visual system appears to select *stimuli* rather than points or regions of space. This is called "object-based attention" (e.g., Duncan 1984; Kanwisher & Driver 1992; Valdes-Sosa et al. 1998), and there is debate about the conditions under which visual selection will be object-based rather than spatial. However, much of the evidence for object-based attention involves facilitated responding to secondary stimuli that happen to appear inside the boundaries of the stimulus that has been selected - as if selecting a stimulus fills its boundaries with spatial attention, facilitating *all* information from the resulting spatial region. Thus this type of selection will also fail to meet the task demands imposed by Levelt et al., at least when the stimulus that must govern responding is the picture. If selecting the picture fills it with spatial attention, then the word will be selected, too.

Therefore we must consider attentional processes that select on something more like the ontological status of the visual entities available to perception - what *kinds* of objects are they? - rather than their spatial locations or physical boundaries. Though we know that people can do this, we don't know much about how they do it. Neuroimaging evidence on the color-word Stroop task suggests that midline frontal regions centered in the anterior cingulate cortex might play an important role in this type of selection, perhaps interacting with various regions in lateral prefrontal cortex, each of which "knows" how to interact with more specialized posterior perceptual and memorial structures that handle a particular class of information. This network of interacting structures centered on cingulate cortex has been called the "anterior attention system" (Posner & Raichle 1994) and has been hypothesized to be the "executive control" component of "working memory" (e.g., Carr 1992; Carr & Posner 1995).

In addition to spatial, object-based, and ontological considerations, selecting the right stimulus for Levelt et al.'s tasks also has a temporal aspect. A large literature on sequential interference effects such as the "psychological refractory period" and the "attentional blink" indicates that when two stimuli appear in close temporal succession, as in Levelt et al.'s tasks, one or the other or both suffer interference and are processed more slowly (and are more likely to be missed entirely, if presentation is brief) than if they had been separated by a longer interval. Such interference is quite common if the first stimulus requires an overt response (this is the "Psychological Refractory Period" or PRP; see, e.g., Pashler 1994), and it can also occur if the first stimulus requires attention and decision or memory storage even if it does not require an immediate response (this is the "attentional blink" or AB; see, e.g., Arnell & Jolicoeur 1997; Chun & Potter 1995; Shapiro & Raymond 1994). But so what? If the interference simply adds main-effect increments to all target response latencies, then no harm is done to Levelt et al.'s arguments about priming effects, which are differences among these latencies that would remain unchanged. However, there is evidence that the AB interacts with priming effects, at least when stimulus presentation is brief and correct responding is the dependent measure. Related items suffer less interference than unrelated items (Maki et al. 1997). Attention-based interference between stimuli occurring in close temporal succession may be something Levelt et al. can't ignore.

Finally, meeting Levelt et al.'s task demands requires yet another type of attention. Once the right stream of information processing has been selected - the one representing the picture or the one representing the word - participants must select the right code or level of processing within that stream from which to construct their response. Sometimes this is the basic level, when the task is object naming, and sometimes it is the superordinate level, when the task is categorization. We are now clearly dealing with executive control and memory retrieval operations commonly attributed to "working memory" and the "anterior attention system," rather than to input selection operations of the "posterior attention system" (Rafal & Henik 1994). Added complications arise when the word to which these representations correspond is newly learned and hence relatively unfamiliar and weakly established in conceptual and lexical memory. It appears diat a special class of inhibitory retrieval operations are brought to bear when such a weak code must be retrieved to meet task demands, and the consequences of its operation for related codes in conceptual and lexical memory is inhibitory. Dagenbach and I (1994) have conceived of this inhibition as arising from a center-surround attentional mechanism that centers on the weak code and suppresses stronger competitors in nearby semantic space. This contrasts with the facilitative effects of semantic relatedness observed with better-learned materials that gets translated into automatic spreading activation in WEAVER. Thus taking account of differences in degree of learning among the lexical items being processed may add a final layer of complexity to the task of fixing WEAVER'S attention deficit.

Sharpening Ockham's razor

Anne Cutler[a] and Dennis Norris[b]

[a]*Max-Planck-Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands,* **anne.cutler@ mpi.nl** [b]*MRC Cognition and Brain Sciences Unit, Cambridge CB22EF, United Kingdom. dennis.norris@mrc-cbu.cam.ac.uk*

Abstract: Language production and comprehension are intimately interrelated; and models of production and comprehension should, we argue, be constrained by common architectural guidelines. Levelt et al.'s target article adopts as guiding principle Ockham's razor: the best model of production is the simplest one. We recommend adoption of the same principle in comprehension, with consequent simplification of some well-known types of models.

Levelt, Roelofs & Meyer propose an account of lexical access, which forms part of the overall enterprise of Levelt and his colleagues - beginning with Levelt (1989) - to model the process of speaking. As clearly stated in the present paper and elsewhere, their enterprise is guided by the principle of Ockham's razor: Choose the simplest model that will explain the relevant data. Thus their theory currently proposes that flow of information between processing levels is unidirectional, that activation may be facilitatory but not inhibitory, and so on (sect. 3.2.5).

As comprehension researchers, we laud Levelt et al.'s methodological stringency, which we consider all too unusual, especially within a connectionist modelling framework. In this commentary we recommend that the same razor could be used to trim excess growths in the modelling of comprehension processes as well.

In research on language comprehension, connectionist models are now not at all scarce; McClelland and Rumelhart (1981; Rumelhart & McClelland 1982) changed the way of life for reading theoreticians, and McClelland and Elman (1986), with TRACE, changed it for researchers in spoken-language understanding. The models McClelland and his colleagues proposed were interactive activation networks with topdown connections between levels of processing, and this feature of the architecture was invoked to explain important experimental findings. In each case, however, it turned out that the topdown connections were *not* necessary to explain the observed effects. We will describe one case from each area: the word superiority effect in reading, and the apparent lexical mediation of compensation for coarticulation in spoken-language understanding.

The "word superiority effect" is the finding that letters can be more easily identified in words (e.g., BRAT) than in nonwords (e.g., TRAB). This was easily simulated in the interactive activation model of McClelland and Rumelhart (1981); feedback connections from the word level to the letter level in the model allowed activation to flow topdown and increase the availability of the target letter nodes. Norris (1990), however, built a simple network with no feedback connections and with separate, unconnected, output nodes for words and letters; these two sets of outputs were separately trained and, after training, the model identified letters in words better than letters in nonwords. Thus the availability of feedback connections was not necessary for the appearance of the word superiority effect.

Compensation for coarticulation is a shift in the category boundary for a particular phoneme distinction as a function of the preceding phonetic context. Elman and McClelland (1988) apparently induced such compensation from lexical information; the preceding phonetic context supplied in their experiment was in fact a constant token ambiguous between [s] and [S], but it occurred at the end of Christma* versus fooli*. Listeners' responses to the phoneme following this constant token were shifted in the same direction as would have been found with the really different phonemes at the end of Christmas and foolish. Subsequently, it has been shown that this result is dependent on listeners' knowledge of transitional probabilities rather than on individual lexical items (Pitt & McQueen 1998). However, Elman and McClelland simulated their result in TRACE and attributed it to TRACE'S feedback connections between the lexical and the phoneme level. Norris (1993), though, managed to simulate these experimental findings in a network that did not make use of feedback connections during recognition. The simulation used a recurrent network with interconnected hidden units. In the critical case, the network was only ever trained to identify phonemes and did not learn to identify words. The network developed a bias towards identifying ambiguous phonetic tokens consistently with the words it had been trained on, and this bias exercised an effect on identification of following phonemes in just the way that the network had been trained to achieve with unambiguous tokens. Importantly, the availability of feedback connections was again not necessary for the appearance of the lexically sensitive compensation for coarticulation effect.

Thus the incorporation of topdown connections in these influential models of comprehension was in no way motivated by a need to explain empirical observations. The models' architects simply chose a more complex structure than was necessary to explain the data.

Further, as Norris et al. (submitted) have argued, topdown feedback in a model of comprehension is not even necessarily able to deliver on the promises, concerning a general improvement in recognition performance, in which its proponents seem to have trusted. Consider feedback from the lexical level to prelexical processing stages. The best word-recognition performance is achieved by selection of the best lexical match(es) to whatever prelexical representation has been computed. Adding feedback from the lexical level to the prelexical level does not improve the lexical level's performance: either the match selected is the best one or it is not. Feedback can certainly result in changed performance at prelexical levels. For instance, if the output of prelexical processing is a string of phonetic representations corresponding to "feed*ack" where the * represents some unclear portion, topdown activation from the lexicon might change the prelexical decision from uncertainty to certainty that there had been a [b]. But suppose that there had not in fact been a [b]? Suppose that the speaker had actually made a slip of the tongue and said feedpack, or feedfack? In that case, the topdown information flow would, strictly speaking, have led to poorer performance by the prelexical processor, since it would have caused a wrong decision to be made about the phonetic structure of the input. (In many listening situations this would of course matter very little, but it might be disastrous for a language production researcher who was interested in collecting slips of the tongue!)

Thus topdown connections can clear up ambiguity in prelexical processing, but they do so at a potential cost, more importantly, they do not result in an improvement of word recognition accuracy. Simulations with TRACE, for instance, have shown that the overall accuracy of the model is neither better nor worse if the topdown connections that the model normally contains are removed (Frauenfelder & Peeters 1998).

Ockham's razor clearly applies: models without topdown connections can explain the currently available comprehension data as well as the models with topdown connections, so in the absence of any superiority of performance to force a choice between the models, selection must be motivated on architectural grounds alone. We propose, with Levelt et al., that the simpler model should always be the first choice.

Levelt and his colleagues at present confine their modelling to the process of speaking. But in principle one would surely want to see models of comprehension and of production that were interdependent, as well as, of course, architecturally constrained in a similar fashion. The processes themselves are interdependent, after all - especially the ability to speak depends on the ability to perceive. Therefore it cannot in the long run be the case that the model of speaking is unconstrained by perceptual processes. The Grand Unified Theory of language perception and production is some time in the future. But we would argue that the success so far of the Levelt et al. enterprise, with its highly constrained and pared-down theory, suggests that Ockham's razor should be kept honed for general psycholinguistic use.