Published in *Psychological Review*, 2000, v. *107*, 709-734.

Superadditive effects of multiple lesions in a connectionist architecture: Implications for the neuropsychology of optic aphasia

Mark Sitton*
Institute of Cognitive Science
University of Colorado
Boulder, CO 80309–0344

Michael C. Mozer
Department of Computer Science
and Institute of Cognitive Science
University of Colorado
Boulder, CO 80309–0430

Martha J. Farah
Department of Psychology
University of Pennsylvania
Philadelphia, PA 19104–6196

Abstract

Traditional accounts of neuropsychological disorders have been based on the assumption that brain damage results in a single focal lesion of the cognitive architecture. Although this strategy has been productive for neuropsychology, some syndromes defy explanation in this way, resulting in an unparsimonious architecture with many specialized components or connections. Optic aphasia is one such syndrome. Patients with optic aphasia have difficulty naming visually presented objects. However, the deficit is not in visual recognition per se because patients can pantomime the appropriate use of objects, and the deficit is not in naming per se because they can name objects from auditory cues. To explain optic aphasia, past accounts have hypothesized multiple semantic systems or multiple functional pathways to visual naming. Farah (1990) instead sketched an account based on lesions to multiple pathways—one that maps visual input to semantics, and the other that maps semantics to naming responses. This account supposes that the effects of the two lesions are *superadditive,* meaning that a task requiring both damaged pathways (e.g., naming a visually presented object) would manifest a much higher error rate than expected based on the sum of error rates on two tasks requiring just one pathway or the other (e.g., gesturing the appropriate use of a visually presented object, and naming from auditory cues). We have explored this hypothesis by modeling superadditive effects of damage in a connectionist architecture. The resulting model explains other phenomena associated with optic aphasia, including the tendencies of patients to: produce a large number of naming errors that are semantically related to the target but few visually related errors, show response perseveration from one trial to the next, "home in" on the correct response over time, and make fewer errors on naming from a verbal description than on gesturing the use of an object from a visual presentation. More broadly, superadditive effects of damage provide a novel class of explanations for neuropsychological deficits that might previously have seemed to imply the existence of highly specialized processing components.

* Please address correspondence to the second author. E-mail: mozer@cs.colorado.edu

Introduction

The field of cognitive neuropsychology has been defined in terms of two related goals: To develop and test theories of normal cognition using behavioral data from brain-damaged patients and reciprocally, to understand and remediate the cognitive disorders of brain-damaged patients using the theories and methods of cognitive psychology (Coltheart, 1984). Both goals require bridging the realms of abstract, information-processing theories of the cognitive architecture and the behavioral abilities of persons with neurological damage. In the case of the first goal, the required bridge is *inferential*, in that we must infer the identity and organization of information-processing components from the behavior of patients in particular tasks. Such inferences depend, perhaps not surprisingly, on a number of assumptions about the cognitive architecture and its response to damage (see Shallice, 1988, and Farah, 1994, for further discussion).

Cognitive neuropsychologists typically make the assumption that a selective impairment in patient behavior can be attributed to the loss of a single component of the cognitive architecture, and that the normal function of this component can be understood quite directly in terms of the scope of the patient's impairment. In other words, a selective impairment in ability X implies the existence of a component of the cognitive architecture dedicated to X. One important and influential example of this form of inference comes from research on memory. A selective impairment in explicit, declarative memory, with preserved implicit and nondeclarative memory, has been used to infer a distinct cognitive module for explicit declarative memory (Squire, 1992). Similarly, one of us has argued that a selective impairment in face recognition, with relative preservation of nonface object recognition, can be used to infer a component of visual recognition that is relatively specialized for faces (e.g., Farah, 1996).

This form of inference is straightforward, widely used, and often successful. Nonetheless, it sometimes yields conclusions that seem suspect. In some of these cases, simpler and more sensible conclusions regarding the cognitive architecture result from the application of less straightforward inferences between behavioral impairment and cognitive architecture. Increasingly, neuropsychologists are questioning the assumption that a focal behavioral impairment emerges directly from a focal cognitive impairment, and instead are considering the possibility that the behavioral impairment is the result of potentially complex interactions between the impaired cognitive component and intact components. Connectionist modeling has played an important role in conceiving and testing these alternative inferences based on interactivity (e.g., Hinton & Shallice, 1991; Mozer & Behrmann, 1990; see Farah, 1994, for a review and discussion).

In this article, we extend the interactive paradigm of neuropsychological inference further, showing how a focal behavioral impairment can be caused by interactions between *multiple* impaired cognitive components. This new form of inference expands the set of neuropsychological impairments that can be accounted for with simple cognitive architectures. We focus on the syndrome of optic aphasia, which has long puzzled neuropsychologists, and show that the new type of inference explains not only the major defining characteristics of the syndrome, but a number of associated features as well. We then conclude with a review of other syndromes that have heretofore seemed problematic for cognitive neuropsychology, and sketch out ways in which interactions between multiple lesions provide explanations in terms of simple theories of the cognitive architecture.

Optic aphasia

Optic aphasia is a neuropsychological disorder in which the naming of visually presented stimuli is impaired, in the absence of a general visual agnosia (visual recognition impairment) or a general anomia (naming impairment).

Optic aphasia can be contrasted with agnosia in several respects. First, unlike agnosic patients, optic aphasics can often demonstrate recognition of visually presented objects nonverbally. These nonverbal demonstrations include the ability of patients to gesture or pantomime the appropriate use of an object (e.g., Gil et al., 1985; Lhermitte & Beauvois, 1973; Riddoch & Humphreys, 1987b) and to sort visual items into their proper superordinate categories (e.g., Assal & Regli, 1980; Coslett & Saffran, 1989; Riddoch & Humphreys, 1987). A second contrast to agnosic patients is that many of the visual naming errors made by optic aphasics are *semantically* related to the target (e.g., *snake* → "frog") or are repetitions of previous responses (*response perseverations*), whereas the visual naming errors made by agnosics are typically *visually* related to the target (e.g., *snake* → "rope"). Third, optic aphasics appear to be relatively insensitive to the quality of visual stimuli, showing roughly equal naming performance when presented with line drawings, color pictures, or three-dimensional objects. Agnosics, on the other hand, appear to be extremely sensitive to the quality of the stimulus, showing better performance as the visual quality increases (Davidoff & de

Blesser, 1993; Farah, 1990). Finally, optic aphasics are usually described as being unimpaired in everyday life, whereas agnosics are often noticeably handicapped by their inability to recognize objects, people, and locales.

Optic aphasia can also be contrasted with anomia. Unlike most anomic patients, optic aphasics can name objects presented in the tactile modality (Assal & Regli, 1980; Coslett & Saffran, 1989; Larrabee et al., 1985; Poeck, 1984; Riddoch & Humphreys, 1987b; Spreen, Benton, & Van Allen, 1966), by the sounds they make (Assal & Regli, 1980; Gil et al., 1985; Spreen, Benton, & Van Allen, 1966), as well as by spoken definition (all cases).

Other characteristic symptoms shared by optic aphasics include alexia and the ability to *home in* on the correct name of a visually presented object, if given sufficient time to respond. This homing-in process is a kind of verbal bouncing around as the patient converges on the correct name of an object. As an example, in one of the most thoroughly documented cases of optic aphasia, the patient Jules F. produced the following response when presented with a picture of a bus (Lhermitte & Beauvois, 1973, p. 707): "a wagon … public transport since there is a back door … a stage coach … it would be … no … a city cab … not a cab but a city bus."

The neuropathology of optic aphasia shows a fair degree of uniformity. All cases appear to have unilateral left posterior lesions. In cases where sufficient localizing evidence exists, the damage seems to include the occipital and temporal cortex and the splenium of the corpus collosum (Schnider, Benson, & Scharre, 1994).[1]

Specific data to be explained

The defining feature of optic aphasia, as well as its most remarkable characteristic, is the disproportionately large error rate when naming visually presented stimuli (which we will often abbreviate as $V \Rightarrow N$), relative to the error rate when naming objects from auditory or other nonvisual cues ($A \Rightarrow N$), or when gesturing the appropriate use of an object or sorting objects by semantic category to demonstrate recognition of visually presented stimuli ($V \Rightarrow G$). Table 1 summarizes the experimental literature on these three tasks. Any model of optic aphasia must account for this pattern of data.

In addition to this defining characteristic of optic aphasia, there are a number of other associated phenomena. Although it is possible that these phenomena are functionally independent and co-occur only because their critical lesion sites are located near one another, a more unified explanation is also possible. In the present paper, we attempt such a unified explanation. The associated phenomena in question include the tendency for errors to be either perseverative or semantically related to the target response, but not visually related (though some errors show a combination of visual and semantic similarity to the target), and the "homing in" process whereby optic aphasics give a sequence of responses to a visual object and may eventually converge on the correct name. In addition, there are two

TABLE 1. Performance of optic aphasia patients on three tasks. Where table cell contains "low", no quantitative data was available, but error rate was presumably low or else patient would not have been diagnosed as having optic aphasia.

	V⇒N (visual stimulus, naming response)	V⇒G (visual stimulus, nonverbal response)	A⇒N (auditory stimulus, naming response)
Lhermitte & Beauvois (1973)	27%	0%	4%
Gil et al. (1985)	36%	0%	low
Riddoch & Humphreys (1987)	54%	25%	low
Manning & Campbell (1991)	58%	25%	0%
Larrabee et al. (1985)	70%	low	low
Hillis & Caramazza (1995)	75%	65%	low
Ferro & Santos (1984)	77%	low	10%
Coslett & Saffran (1992)	79%	0%	32%
Assal & Regli (1980)	97%	75%	low
Poeck (1984)	100%	25%	10%
Coslett & Saffran (1989)	100%	50%	low

1. It should be noted that the structural lesion visible on CT or MRI scan may underestimate the extent of damaged and dysfunctional tissue.

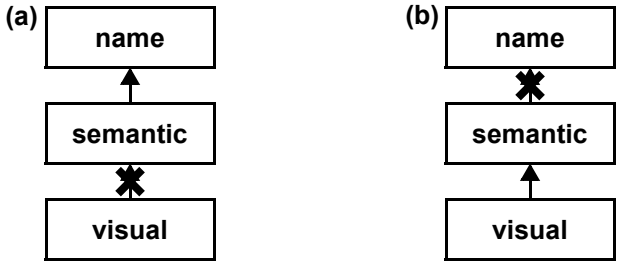

FIGURE 1. Possible disconnections in a simple functional architecture for visual naming. In this and subsequent figures, the boxes denote levels of representation and the arrows denote functional pathways mapping from one level of representation to another. The "X" denotes hypothesized damage to a pathway.

weak trends observable across cases, which may or may not hold up as new cases are reported: gesturing to visual objects is generally worse than naming to auditory cues; and the worse the naming of visual objects, the worse the gesturing to visual objects.

Before presenting our model and testing its ability to account for all of the foregoing characteristics of optic aphasia, we review previous attempts to explain the defining characteristic of this disorder: a relatively selective impairment in the naming of visual objects.

Models of optic aphasia

The highly isolated nature of the visual naming deficit displayed by optic aphasics seems to invite explanation in terms of a disconnection between intact visual centers and intact naming centers. Upon closer examination, however, this explanation does not work. A commonly accepted model of the functional architecture underlying visual naming, which we will refer to as the *canonical* model, consists of a visual processing system which feeds its output into a semantic system, which in turn feeds its output into a naming system (Figure 1). The basic reasoning behind this model is that one cannot name a visually presented object until one first knows what the object is. If optic aphasia is a simple disconnection syndrome, then the disconnection should occur somewhere in this architecture. The difficulty arises in that every possible locus of disconnection conflicts with at least one essential characteristic of optic aphasia. One cannot place the lesion in vision, semantics, or the pathway connecting them (Figure 1a) because patients can demonstrate their recognition of visually presented objects nonverbally. Nor can one place the lesion in naming or the pathway between semantics and naming (Figure 1b) because patients are unimpaired in their ability to name objects presented in the tactile or auditory modalities, which also presumably feed their output into the semantic system. Five different models of optic aphasia have been proposed to resolve this paradox.

A direct visual naming pathway

To account for the characteristic naming deficit of optic aphasia, Ratcliff and Newcombe (1982) proposed an alternative model of visual naming that posits a direct, uninterrupted pathway between vision and naming (Figure 2). In this architecture, specific visual percepts can evoke their corresponding name directly. The information emerging from the visual naming pathway that is mediated by semantics, combines with the information emerging from this direct pathway to yield reliable naming of objects. Optic aphasia results when the direct visual naming pathway becomes disconnected. This forces the system to rely solely on the semantics-mediated pathway, which is not entirely reliable on its own. While Ratcliff and Newcombe admit that this direct visual naming pathway hypothesis is somewhat ad hoc, they point out that there is some evidence from studies of dyslexia that appear to support its existence. For example, Schwartz, Saffran, and Marin (1980) describe a dyslexic patient who was able to pronounce written words, even irregular ones, despite appearing to be unaware of what they meant. It should be pointed out, however, that there have not been any documented cases of individuals who can name visual objects without any knowledge of what the objects are.

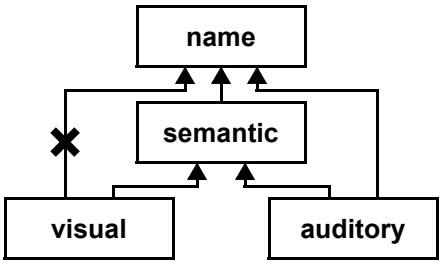

FIGURE 2. A schematic depiction of the Ratcliffe and Newcombe (1982) account of optic aphasia.

Modality-specific semantic systems

Beauvois (1982) presents a model of optic aphasia in which semantics is a nonunitary entity. In this model, each modality has a corresponding semantic system to which it, alone, is directly connected (Figure 3). Visual semantics is composed of imagery-like visual information about objects, while verbal semantics consists of verbal associations and abstract properties of objects. Optic aphasia arises when there is a disconnection between verbal semantics and visual semantics. Since the naming system in this model is assumed to be connected only to verbal semantics, the proposed disconnection will result in impaired visual naming but intact auditory naming. Evidence supporting the model comes from a case described by Beauvois and Sailant (1985) in which the patient, MP, was described as having "optic aphasia for colors." While being unable to perform tasks which required matching color names to color percepts (i.e., visual-verbal tasks), MP was able to perform tasks which required either verbal-verbal or visual-visual associations of colors. Although the model does a good job of explaining this case, it does not appear to account for the ability of optic aphasics to sort visually dissimilar items into the same superordinate category. Furthermore, the idea that each input and output modality has a separate semantic representation of all our knowledge seems ad hoc, not to mention unparsimonious.

McGuire and Plaut (1996) have recently described an interesting variation of the modality-specific semantic systems account, in which connectionist learning mechanisms give rise to a semantic system whose representational elements are *softly tuned* to a specific input and output modality: the elements respond most strongly to these input and output modalities, but they yield partial responses to others. In this model, damage to visual semantics spares gesturing to a greater degree because of systematicity in the semantics-to-gesture mapping. Although this account is intriguing, the systematicity assumption and its empirical consequences must be carefully examined, and the account does not offer an explanation of the preserved nonverbal sorting performance of optic aphasics.

Impaired access to semantics from vision

Riddoch and Humphreys (1987b) and Hillis and Caramazza (1995) have attempted to explain optic aphasia with a unified semantic system, by hypothesizing an impairment in accessing that semantic system from vision. The preservation of gesturing is explained by these authors in slightly different ways. Riddoch and Humphreys proposed a direct pathway from visual representation to gesture, as shown in Figure 4. This pathway embodies the use of *affor-*

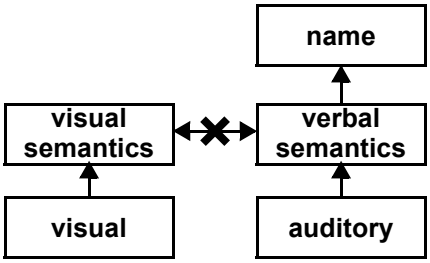

FIGURE 3. A schematic depiction of the Beauvois (1982) account of optic aphasia.

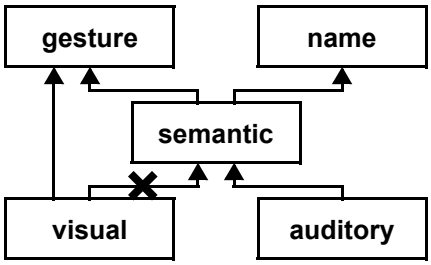

FIGURE 4. A schematic depiction of the Riddoch and Humphreys (1987b) account of optic aphasia.

dances (Gibson, 1979) in the appearance of an object that constrain the ways in which it can be used. According to their account, optic aphasics retain the ability to gesture appropriately to visually presented objects because of a combination of object affordances and the generally less precise and specific nature of most gestures compared to verbal labels. Hillis and Caramazza make a related point about the relation between visual information and semantic information, specifically about the uses of objects. The shape of a chair (a visual characteristic) and the possibility of sitting in it (the type of semantic characteristic tested in gesturing tasks) have a special relation, one that may survive damage that disrupts the activation of a semantic representation from visual input. Whereas nonverbal responses may be initiated by activation of isolated semantic features from isolated visual features, Hillis and Caramazza argue that naming requires access to a complete semantic representation.

Although there is much truth to these observations concerning the relation between shape and gesture, the preserved gesturing of optic aphasics may not be entirely explicable in this way. Many similar-looking objects are associated with different movements—for example, knitting needles and chopsticks, bowls and helmets. Studies of some patients have found accurate performance even when precise and distinctive gestures are required. For example, Lhermitte and Beauvois (1973) report that their patient made no gesturing errors for a large set of stimuli, which were misnamed 27% of the time. However, studies of other patients have found poor performance on difficult nonverbal tasks that involved sorting objects into their proper superordinate categories—tasks that would seem to require complete semantic access (e.g., Hillis & Caramazza, 1995; Gil et al., 1985; Riddoch & Humphreys, 1987b). These findings might be viewed as ambiguous, as language may play a role in mediating sorting performance (Luria, 1961). But even if taken at face value, studies indicating poor performance on difficult nonverbal tasks may simply point to the fact that some patients indeed have a greater semantic deficit than others, apart from their inability to name visually presented stimuli (as suggested by patients in Table 1 who perform poorly on both verbal and nonverbal tasks). A complete account of optic aphasia should be able to accommodate individual differences in the degree of semantic deficit among patients.

Hemisphere-specific semantic systems

Coslett and Saffran (1989) present a model of optic aphasia that is rooted more in functional neuroanatomy than cognitive psychology. Their model consists of independent, functionally unique semantic systems for each hemisphere of the brain (Figure 5). The model also makes us of the fact that the left hemisphere is more proficient at speech. According to this model, optic aphasia arises when there is a disconnection between visual input and left hemisphere semantics. Several independent sources of evidence appear to support this explanation. First, the predominance of

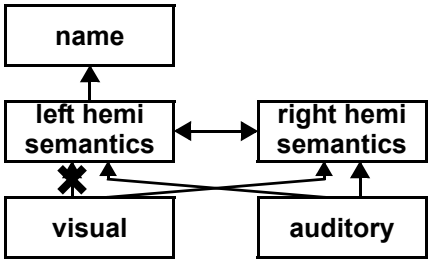

FIGURE 5. A schematic depiction of the Coslett and Saffran (1989) account of optic aphasia.

semantic naming errors made by optic aphasics and their ability to sort items into superordinate categories, as long as the categories are broad enough, is consistent with the hypothesis that right hemisphere semantics are relatively coarse-grained (Zaidel, 1985). Second, Coslett and Saffran present a detailed analysis of their patient's residual reading abilities, and these closely matched the reading abilities of the right hemisphere. Third, Schnider, Benson, and Scharre (1994) present comparative neuroanatomical evidence that optic aphasia is highly correlated with damage to the left occipital cortex and splenium. This seems to imply that optic aphasics process visual stimuli almost exclusively by the right hemisphere, and that the results of this processing have limited access to the left hemisphere and its language output mechanisms. Thus, the hemisphere-specific semantic systems hypothesis accounts well for the defining behavioral characteristics of optic aphasia, and is consistent with the neuropathology. Its weakness lies in the uncertain status of the major assumption behind the hypothesis, namely that the two hemispheres have qualitatively distinct semantic systems. This assumption has been questioned (e.g., Plaut & Shallice, 1993b).

Superadditive impairments in vision and naming

None of the models presented so far provide completely adequate explanations of optic aphasia. In addition, most of the models were constructed primarily to account for the disproportionately large number of visual naming errors made by optic aphasics and have little to say about the associated characteristics accompanying the disorder. Furthermore, each has the property that it includes more pathways or processing systems than the canonical model of visual naming (Figure 1). The impossibility of explaining optic aphasia by a simple disconnection in the canonical model seems naturally to imply that a more complex cognitive architecture is needed. However, Farah (1990) suggested a possible explanation of optic aphasia that requires no additional complexity. Rather than hypothesizing multiple semantic systems or multiple pathways to visual naming, she hypothesized multiple lesions. According to her conjecture, lesions to two pathways in the canonical model—the pathway that maps visual input to semantics, and the pathway that maps semantics to naming responses—might give rise to optic aphasia (Figure 6) if the effect of these lesions were *superadditive,* meaning that a task requiring both pathways (e.g., naming a visually presented object) manifests a much higher error rate than expected based on the sum of error rates on two tasks involving one pathway or the other (e.g., gesturing the appropriate use of a visually presented object, and naming from auditory cues).

Clinical case support for this *superadditive-impairment hypothesis* comes from a study of anomic patients conducted by Bisiach (1966). In one experiment, the patients were asked to name pictures of objects or to indicate their identity by some other means (e.g., by circumlocuting or gesturing), if the name could not be produced. The pictures were either line drawings, line drawings with stray marks superimposed, or full-color paintings. The patients' naming performance was poorest for the marked-up drawings, next poorest for the line drawings, and best for the full-color paintings. In contrast, their recognition performance was relatively insensitive to the quality of the stimulus. One way of interpreting these results is that a kind of temporary optic aphasia was induced when these patients were asked to respond to the marked-up drawings. Because general anomia results from an impaired naming system, and the markings on the drawing resulted in impaired visual perception, these findings support the idea that multiple impairments can have superadditive effects on visual naming.

Although the occurrence of two anatomically distinct lesions in optic aphasia would lend further support to the hypothesis, the typical finding of one lesion should not be viewed as disconfirmation; a single lesion can, and generally does, affect multiple functional areas. Furthermore, the left posterior inferior region damaged in optic aphasia is known to house a multiplicity of functionally distinct areas.

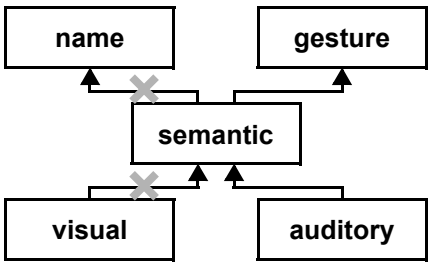

FIGURE 6. A schematic depiction of the superadditive-impairment account of optic aphasia (Farah, 1990). This account involves partial damage to two pathways, one that maps visual input to semantic representations and another that maps semantics to naming responses.

Evidence for the range of distinct perceptual and cognitive abilities that depend on the left temporo-occipital region comes from the variety of impairments other than optic aphasia that may follow damage there in different cases. These include impairments in reading (e.g., Coslett, 1997), visual object recognition (e.g., Feinberg, Schindler, Ochoa, Kwan, & Farah, 1994), visual image generation (e.g., Farah, 1995), memory (e.g., Zola, 1997), semantic knowledge (e.g., Hodges & Patterson, 1995), and lexical retrieval (e.g., Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996). Further support for a posterior temporo-occipital locus for semantic knowledge, in particular, comes from a variety of functional neuroimaging studies (Démonet, Chollet, Ramsay, Cardebat, Nespoulous, Wise, Rascol, & Frackowiak, 1992; Howard, Patterson, Wise, Brown, Friston, Weiller, & Frackowiak, 1992; Klein, Milner, Zatorre, Meyer, & Evans, 1995; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Mummery, Patterson, Hodges, & Wise, 1996; Petersen, Fox, Posner, Mintun, & Raichle, 1988, 1989; Raichle, Fiez, Videen, MacLeod, Pardo, Fox, & Petersen, 1994; Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996; Wise, Chollet, Hadar, Friston, Hoffner, & Frackowiak, 1991). In sum, although the precise mapping of the perceptual and cognitive functions of the left temporo-occipital region has yet to be accomplished, and may possibly vary from individual to individual, there is reason to believe that this region plays an important role in both visual and semantic processes, and certainly no reason to reject this hypothesis. The known functional neuroanatomy is therefore consistent with the superadditive-impairment hypothesis of optic aphasia.

Superadditivity

Before exploring the superadditive-impairment hypothesis in detail, we discuss the notion of superadditivity in general terms. Consider a cognitive architecture that has suffered damage to two pathways, A and B. If a task is to be performed that requires pathway A but not pathway B—call it $task_A$—one would expect poorer performance compared to the undamaged architecture; denote the increased error rate e_A. Similarly, a task $task_B$ would result in error rate e_B. In $task_{AB}$ that involves both damaged pathways, the effects of damage to pathways A and B might contribute independently to performance, in which case the error rate would be e_A+e_B. If, however, the two sources of damage interact, one might obtain superadditivity, i.e., the condition in which

$$e_{AB} > e_A + e_B. \tag{1}$$

The ratio

$$e_{AB} / (e_A + e_B) \tag{2}$$

might be used to quantify superadditivity. The optic aphasia patients we described earlier show ratios roughly between 2 and 8.

At a descriptive level, the conditions under which superadditivity will be observed are straightforward to characterize: Superadditivity occurs if the error rate for some task is determined by the total amount of damage along critical pathways required for performing the task, and the curve relating damage to error rate is positively accelerated (Figure 7). The focus of the present work is to move beyond the descriptive level and present a mechanistic explanation. It turns out that the sort of curve shown in Figure 7 falls out naturally from nonlinear connectionist systems.

Superadditivity implies some type of interaction or dependence between the two loci of damage. To argue this point, consider the probabilistic notion of independence, $p(X \,\&\, Y) = p(X)\, p(Y)$. If $1 - e_A$ is the probability that pathway A operates successfully, and $1 - e_B$ is the probability that pathway B operates successfully, then under independence, the probability that both operate successfully is $(1 - e_A)(1 - e_B)$. Can superadditivity occur under this assumption? That is, are there values of e_A and e_B for which Equation 1 holds? Under independence, Equation 1 is equivalent to

$$1 - (1 - e_A)(1 - e_B) > e_A + e_B \ ,$$

which reduces to

$$0 > e_A e_B \, .$$

This inequality is not satisfied for any positive error rates e_A or e_B, proving that superadditivity cannot occur with independent pathways.

A computational model of superadditive impairments

In this section, we instantiate the superadditive-impairment hypothesis of optic aphasia in a computational model. We constructed this model for two primary reasons. First, the plausibility of the hypothesis will be bolstered by a concrete simulation model as an existence proof. Second, by embodying the hypothesis in a computational model, we can determine whether it can account for some of the more subtle phenomena associated with optic aphasia, and can use the model to predict additional patterns of data.

Our modeling efforts are based on a connectionist framework. Connectionist systems seem particularly well-suited to embody the superadditive-impairment hypothesis because they have the ability to "clean up" mildly corrupted representations and therefore show robustness to partial damage, and their behavior can be highly nonlinear and therefore allows interactions of the sort proposed by the hypothesis.

Basic architecture of the model

Figure 6 shows the model's basic architecture. Following the canonical model, visual and auditory inputs converge on a unified semantic system, which in turn feeds naming and gesturing systems. Each arrow in the Figure, along with the source and destination representations, is referred to as a *pathway*. The four pathways in the model are: visual input to semantics (V→S), auditory input to semantics (A→S), semantics to naming (S→N), and semantics to gesturing (S→G). Pairing an input pathway with an output pathway, four different tasks can be performed: visual naming (V⇒N), auditory naming (A⇒N), visual gesturing (V⇒G), and auditory gesturing (A⇒G).

Pathways

The notion of a pathway—a mapping from one level of representation to another—is fundamental to any connectionist architecture. A *feedforward pathway* achieves the mapping by propagating activity from the input representation to the output representation using feedforward connections, possibly through layers of hidden units (Figure 8a; e.g., Norris, 1993; Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989; Sejnowski & Rosenberg, 1987). In comparison, an *attractor pathway* includes recurrent connections, allowing cooperation and competition among the units to determine the eventual pathway output (Figure 8b). The McClelland and Rumelhart (1991) interactive-activation model is a well-known example with two cascaded attractor pathways, mapping to feature and word level representations; the Farah, O'Reilly, and Vecera (1993) model of overt and covert face recognition is another with cascaded attractor pathways, mapping between name and face representations via a semantic representation. The term "attractor" comes from the fact that, under certain conditions on the connectivity, the pathway output will be drawn to specific representations—the attractors. Assuming that prior training has produced attractors that correspond to meaningful states in the domain, one can think about the attractor pathway as performing *clean up*: the recurrent connections force the pathway to eventually settle into an output representation that has meaning in the domain. For example, in the McClelland and Rumelhart interactive-activation model, activities at the word level tend to settle into

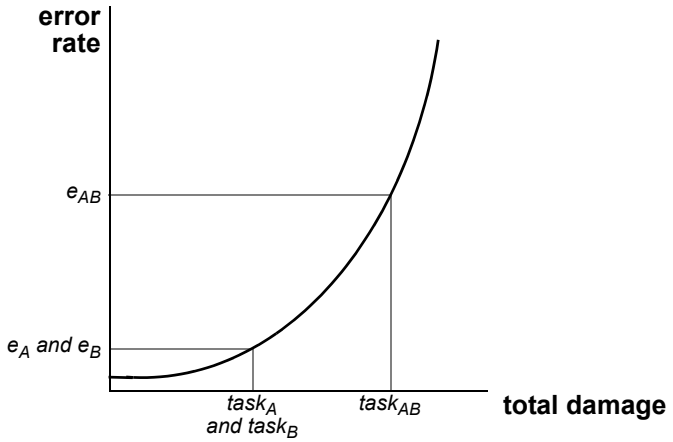

FIGURE 7. A hypothetical graph relating amount of damage along pathways required for performing some task and the error rate on the task.

a localist representation wherein one word unit is highly active and all others are mostly suppressed, due to inhibition between word units; the model of Farah et al. settles into distributed representations that correspond to previously learned names or faces.

Although our description suggests the mutual exclusivity of feedforward and attractor pathways, quite the contrary is true. The output produced by a feedforward pathway can be interpreted only with reference to meaningful patterns in the domain. Consequently, modelers often perform this interpretation as a postprocessing step in which the distance from the output to each of the meaningful patterns is computed, and the output is approximated as the nearest meaningful pattern, or the response probability or reaction time is assumed to be related to the distance to the correct output. Because these read out assumptions essentially duplicate the operation of an attractor net, one could argue that an attractor net is implicit in the feedforward pathway.

Similarly, an attractor pathway requires a feedforward network, because feedforward connections are necessary to map from one representational space to another. Such feedforward connections are present in Figure 8b, although they do not pass through a hidden layer. In the McClelland and Rumelhart interactive-activation model, feedforward (and feedback) networks map between levels, e.g., from the letter level to the word level.

Because a feedforward pathway is incomplete without an attractor net, and an attractor pathway is incomplete without a feedforward net, we suggest the generalized pathway architecture in Figure 8c that includes both feedforward and attractors nets and subsumes the architectures in Figures 8a and 8b. Many existing models explicitly stack the two components as a basic computational module (e.g., Hinton & Shallice, 1991; Mathis & Mozer, 1996; Mozer & Behrmann, 1990; Plaut & Shallice, 1993a, 1993b).

Of course, Figure 8 does not exhaust the architectural possibilities; e.g., the output layer could pass activity back into the hidden layers. However, all that is required for calling Figure 8c a "generalized" pathway is that it is not restricted in its computational power or in the nature of mappings it can achieve relative to other architectures. We have no a priori reason to believe this conjecture false. Additionally, the generalized pathway has the advantage of having components with clearly defined functional roles: The feedforward (black) connections implement a mapping from the input space to the output space, and the recurrent (grey) connections perform clean-up in the output space, by forcing the output representation to one of a predefined set of meaningful alternatives (the attractors). These two distinct stages are depicted in Figure 9.

In a connectionist system, items are represented by a pattern of activity over connectionist units. A pattern of activity over n units can be depicted as a point in an n-dimensional space. In Figure 9, the rectangle on the left depicts a two-dimensional input space, and the rectangle on the right a two-dimensional output space. Each attractor also corresponds to a point in the output space, as defined by previous training.

The pathway architecture is premised on the assumption that not all points in the output space correspond to meaningful entities in the domain. This assumption is valid in any representational domain having discrete entities

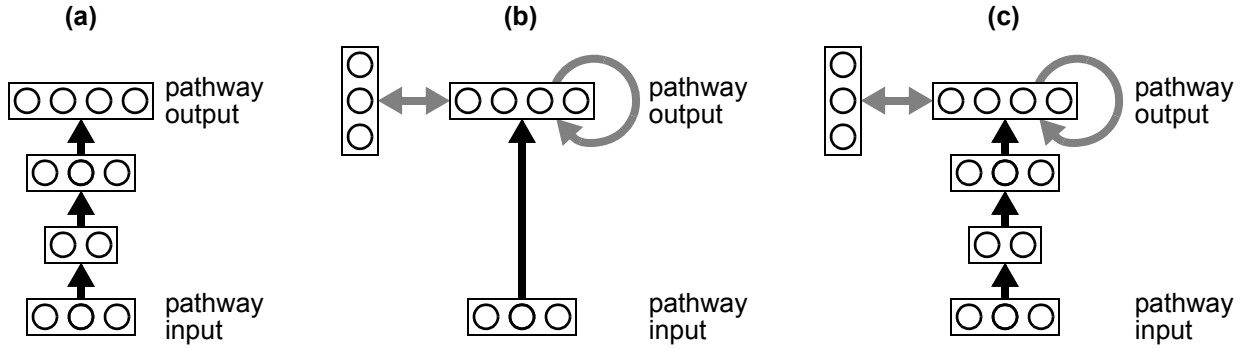

FIGURE 8. Connectionist architectures for implementing a pathway. Each rectangle with circles inside depicts a layer of connectionist processing units, and the arrows depict connectivity from one layer to another (black = feedforward connections; grey = recurrent connections). (a) a feedforward pathway, in which activation propagates upward from the input layer to the output layer; (b) an attractor pathway, in which cooperation and competition among units in the output representation (and the hidden layer drawn to the side) results in attractor dynamics; and (c) a generalized pathway, which includes both multilayer feedforward connections from the input to the output, and recurrent connections among units in the output representation that results in attractor dynamics.

(e.g., phonological patterns). More generally, the assumption is motivated on computational grounds: it bestows upon the architecture a degree of noise resistance. If the input is noisy, or if the feedforward mapping is inaccurate due to insufficient training, or if there is intrinsic noise in the system, the attractor dynamics still force the output to a meaningful state. Because the output of one pathway is input to another, the architecture prevents the accumulation of noise as information is transmitted through multiple pathways and therefore leads to better performance than an architecture like Figure 8a without attractor dynamics (Mathis & Mozer, 1995).

Temporal dynamics of pathways

Consider the situation in which the feedforward network reliably lands the output state in the right attractor basin for a given input, but is not always able to produce exactly the right output for every input. This situation is typical, because a feedforward network will generally show some variation in its output when an input is corrupted by noise or missing features, unless the network has received a huge amount of training and has a very large capacity (i.e., hidden units). Fortunately, moderate inaccuracies in the feedforward network can be compensated for by the attractor dynamics, with a cost in *time*. The feedforward network operates in a single shot: activity is propagated from the input representation to the output representation. In contrast, the attractor network operates through repeated iterations: each unit updates its activity based on the current activity of all other units; as this process iterates, the state moves toward and eventually stabilizes on an attractor. The grey curved line in Figure 9 reflects this gradual convergence on the correct state over time—a *speed-accuracy trade off* of sorts. Speed-accuracy trade offs are ubiquitous in interactive networks that produce stable states. Convergence is attained via an update rule that moves the state at iteration i to a state at iteration i+1 which better approximates some target state, thereby guaranteeing an improvement in the state over time, or equivalently, a speed-accuracy trade off.

The speed-accuracy trade off obtained in a single pathway has important consequences when several pathways are placed in cascade, as in the superadditive-effects model of Figure 6. If a pathway A feeds into a pathway B, such as V→S feeding into S→N, then the state unit activities of A serve as the input to B. Because these activities change over time as the state approaches an attractor, the dynamics of pathway B can be quite complex as it is forced to deal with an unstable input. In our simulations, we show that mild damage to A adds noise to the early output of the pathway, and this noise can combine with mild damage to B to produce superadditive effects.

When pathways are cascaded, the depiction of pathway operation in Figure 9 is misleading in suggesting that the two stages of a pathway operate sequentially. Sequential operation is a sensible way of thinking about the architecture when the pathway input is fixed, because once the feedforward net has produced its output, its job is complete. However, when the pathway input changes over time—as when the pathway is the second in a cascade—the output of the feedforward net also changes over time, and if the attractor net is to make effective use of the information it is provided, it needs to remain sensitive to the state output by the feedforward net. Consequently, the attractor net is faced with a dilemma: to what extent should it remain sensitive to its changing input and to what extent should it operate under its own dynamics to perform clean up? This dilemma is fundamental and implicit in any architecture containing cascaded attractor nets. One solution is to treat the cascade as a single, large, undifferentiated attractor net,

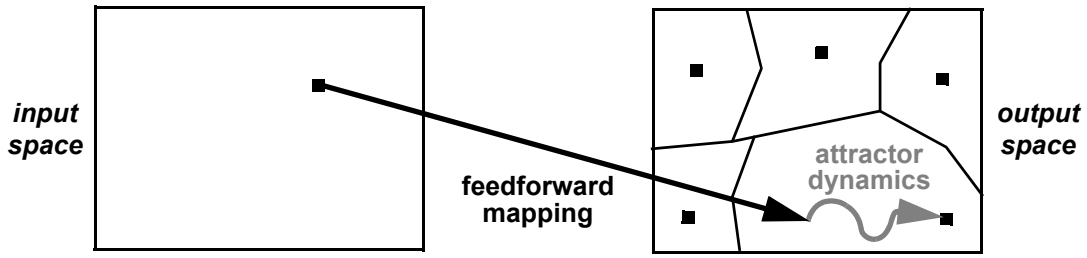

FIGURE 9. An intuitive depiction of the operation of a generalized pathway, adapted from Hinton and Shallice (1991). The rectangle on the left denotes the input space of the pathway and the rectangle on the right denotes the output space. The feedforward mapping performs a transformation from the input space to the output space. The recurrent connections implement attractor dynamics in which the output is drawn to an attractor, denoted by the small squares in the output space. The attractors correspond to the meaningful states in the output domain, as defined by prior training. The lines carving up the output space into nonoverlapping regions indicate the *attractor basins*. Any initial state that lies in one of these regions will be drawn to the attractor corresponding to that region.

but doing so discards any notion of modularity or functional specialization within the architecture (similar to treating the entire neocortex as a homogeneous entity), and consequently limits our ability to understand the operation of the system. Instead, we present an alternative approach to addressing the dilemma in our implementation of a pathway. Returning to Figure 9, perhaps a better way of intuiting the operation of a pathway is to think of the feedforward net output (the tip of the black arrow) not as the initial state of the attractor net, but as a force which continually draws the attractor net state (the grey arrow), much as do the attractors.

Connectionist implementation of a generalized pathway

In this section, we describe the implementation of a generalized pathway's two connectionist components: the feedforward net and the attractor net.

The feedforward net is a standard multilayer perceptron with a single hidden layer and symmetric sigmoidal activation functions (i.e., activity of a unit ranges from –1 to +1). The input units are fully connected to the hidden units, which in turn are fully connected to the output units, as in Figure 8a with one hidden layer removed.

The attractor net architecture commonly used in connectionist cognitive models (e.g., Hinton & Shallice, 1991; Plaut & Shallice, 1993a, 1993b) might be termed *fully distributed*, in that the knowledge about where attractors are located in the state space is spread over connections in the network; consequently, one cannot hand wire a network to have a certain set of attractors, and training procedures to sculpt an attractor landscape are laborious, unreliable, and generally result in *spurious* attractors (i.e., attractors in locations other than the ones where they are desired). For this reason, Zemel and Mozer (2000) have developed an alternative, termed a *localist* attractor network, in which information pertaining to a particular attractor is localized in the weights of the network. Consequently, setting the connections in these networks to achieve a particular set of attractors is a simple procedure; also, spurious attractors are avoided. We emphasize that the difference between distributed (standard) and localist attractor nets is at the implementation level; the qualitative dynamical properties of the two architectures are the same.[2]

A localist attractor network consists of a set of *state* units and a set of *attractor* units. The pattern of activity over the state units is the current state of the system—a point in the *n*-dimensional space on the right side of Figure 9. For each attractor in state space, there is one attractor unit. Intuitively, the attractor units draw the state toward their corresponding attractor in inverse proportion to their distance from the state. The update procedure involves two steps. First, each attractor unit i measures the distance from the location of its attractor, denoted by the vector μ_i, to the state at the current time t, $\boldsymbol{s}(t)$:

$$dist_i(t) = \left\| \boldsymbol{s}(t) - \mu_i \right\|^2 / \beta_i ,$$

where β_i is a *strength* parameter that influences the shape of the attractor basin—the region of state space over which an attractor will exert its pull, and also the rate at which the state will converge to the attractor. The attractor unit activities are then computed via a normalized exponential transform:

$$a_i(t) = \frac{e^{-dist_i(t)}}{\sum_j e^{-dist_j(t)}} .$$

This transform ensures that the attractor activities sum to 1, and that attractors which are close to the current state will have exponentially stronger draw on the state than those further away.

In the second step of the update procedure, the state is pulled toward each attractor in proportion to the attractor unit activity (the proximity of the attractor). The vector representing the combined influence of all attractor forces is:

$$\boldsymbol{influence}(t+1) = \sum_i a_i(t)\, \mu_i$$

2. The spurious attractors in a distributed attractor network can be viewed in a positive light in some cognitive models. The spurious attractors are not random states, but tend to be states that are a mixture of two or more attractors the network has been trained on. This type of compositionality allows the network to generate novel responses that share structure with the training examples. Although the ability to generate novel responses can be important in some cognitive domains, it is not relevant for the tasks we are modeling, and hence, the use of a localist attractor network is without a down side.

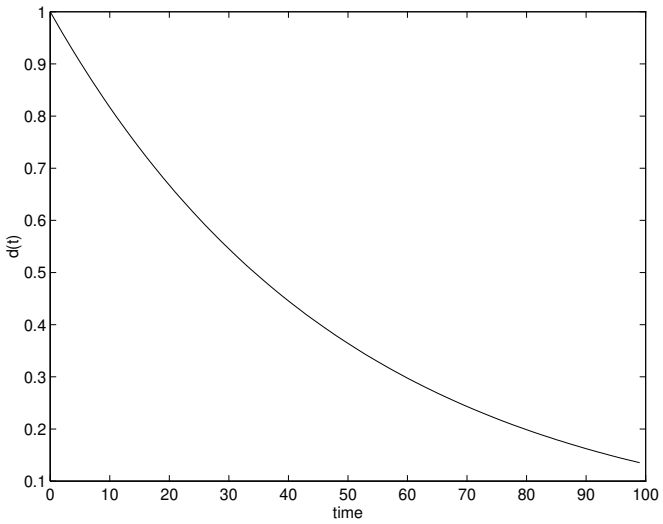

FIGURE 10. The relative influence of the external input versus the internal attractors of the attractor network, as a function of time. This curve assumes that the external input appears a time step 0 and does not change.

The state depends not only on the forces exerted by the attractors, but also on the external input, ***ext***, which is the output of the feedforward net. The new state is a weighted combination of the influence and the external input, modulated by a parameter ω in the range [0, 1]:

$$\boldsymbol{s}(t) = \omega\, \boldsymbol{ext}(t) + (1 - \omega)\, \boldsymbol{influence}(t).$$

As we alluded to earlier, this architecture poses the dilemma of how to set ω. Zemel and Mozer (1998) present a mathematically principled approach in which ω as a function of time is derived from a maximum likelihood formulation of the search task, assuming that the external input is fixed over time. Because we cannot assume the external input is fixed—it may propagate from another pathway whose output is changing over time—we adopted a heuristic rule for setting ω, based on the following argument.

Essentially, ω determines the degree to which attention is paid to the external input as opposed to the internal dynamics of the attractor network. The net should remain responsive to the external input as long as it is changing, i.e., ω should be close to 1.0, because the input is apt to be increasingly accurate as time passes. Once the input has stopped changing, however, then the internal dynamics of the attractor net should take over to interpret the input, and ω should be reduced toward 0.0. Rather than waiting for the entire input vector to stabilize, the model can increase its speed without loss of accuracy by specifying ω independently for each dimension of the state space; the parameter for dimension i will be denoted ω_i. The model automatically sets ω_i based on a comparison between the instantaneous input, $ext_i(t)$, and a recent-time average of the input, $\overline{ext}_i(t)$:

$$\omega_i(t) = h\left[1 - \frac{\overline{ext}_i(t-1)}{ext_i(t)}\right],$$

where $h[.]$ is a linear saturation function that bounds activity between –1 and +1. As the external input stops changing, $\overline{ext}_i(t-1)$ approaches $ext_i(t)$, and ω_i drops to 0.0. Figure 10 shows a graph of ω_i as a function of time, when the input ext_i changes instantaneously from 0 to 1 at time 0. The exponential drop-off in the Figure is due to an exponentially-weighted rule for computing the average input:

$$\overline{ext}_i(t) = \alpha\, ext_i(t) + (1 - \alpha)\overline{ext}_i(t-1) \quad .$$

where α is the averaging constant.

Simulation methodology

Pattern generation

The simulation requires items to be represented in five spaces: *visual* input, *auditory* input, *semantic*, *name* responses, and *gesture* responses. Items are encoded as patterns of activity over connectionist units. One approach to pattern construction is to specify meaningful features in each space—e.g., the visual feature of elongation or the semantic feature of "usefulness as a weapon"—and then determine the appropriate feature values for a set of items. This approach is not only laborious, but with a small set of features and a small sample of objects, one cannot be certain that the statistics of the patterns match the statistics of the knowledge people have about objects. We thus took an alternative approach of generating *random* patterns in these representational spaces subject to certain constraints. With this approach, we know exactly what assumptions are being designed into the representations and what the relationship is between patterns in different representational spaces.

Each representational space was 200 dimensional, meaning that the pattern of activity in each space was encoded over 200 units. In each representational space, we generated 200 different binary-valued (-1,+1) patterns, which were meant to correspond to known entities of that representational domain.[3] The semantic patterns served as the attractors for the V→S and the A→S pathways; the name patterns served as the attractors for the S→N pathway; and the gesture patterns served as the attractors for the S→G pathway.

For the visual, auditory, and semantic spaces, patterns were partitioned into 50 similarity *clusters* with 4 *sibling*s per cluster. Patterns were chosen randomly subject to two constraints: (1) patterns in different clusters had to be at least 80° apart;[4] (2) siblings had to be between 25° and 50° apart. In connectionist networks, pattern vectors separated by a small angle will tend to have similar effects in processing; pattern vectors that are orthogonal (i.e., 90°) will tend to have unrelated effects; and pattern vectors that are separated by large angles (close to 180°) will tend to have opposite effects. Thus, the constraints on the angle between patterns enforced a similarity structure on the representational space. This allows us to ascertain whether the model is more likely to produce, say, semantically-related items as error responses.

Because similarity of patterns in the name and gesture spaces was irrelevant to our modeling, we did not impose a similarity structure on these spaces. Instead, we generated patterns in these spaces at random subject to the constraint that every pattern had to be at least 60° from every other.[5]

After generating patterns in each of the representational spaces, we established arbitrary correspondences among the patterns such that visual pattern *n,* auditory pattern *n*, semantic pattern *n*, name pattern *n*, and gesture pattern *n* all represented the same concept. That is, the appropriate response in a visual naming task to visual pattern *n* would be semantic pattern *n* and name pattern *n*.

Training procedure

Our goal is to model of adult competence and the deficit in performance that results from brain damage, not to model the course of human learning and development. Nonetheless, before the model can be tested, each pathway in the model must be trained to achieve adult competence, i.e., to map a given input to the pathway to the corresponding output. The particular procedure used for training is not an essential aspect of the model, so long as the resulting model exhibits adult competence. One approach to training the generalized pathway architecture of Figure 8c is to

3. By selecting a dimensionality for the space that was on the same order as the number of alternative patterns, we were assured that the patterns would not be packed together too closely.

4. The angle between two patterns can be computed from the identity

$$\cos\theta = \frac{\boldsymbol{p}_1^{\mathrm{T}}\boldsymbol{p}_2}{||\boldsymbol{p}_1||\,||\boldsymbol{p}_2||},$$

where $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are the two pattern vectors, and θ is the angle between them.

5. The choices we made for the minimum angle between unrelated patterns in the five representational spaces were not critical to the model's behavior. However, as the minimum angle is increased, the model becomes more robust to damage. This makes sense because it is easier to clean up a corrupted pattern the further apart the target patterns are. Additionally, as the minimum angle is increased, we found that superadditive effects of damage increased. This is related to the noise robustness property of dissimilar patterns, as will hopefully become clear in the discussion of superadditivity that follows.

use a supervised learning procedure such as back propagation to determine connectivity within the entire architecture. However, because each pathway consisted of two distinct components—the feedforward net and the attractor net—with distinct functional characteristics, it seemed sensible to determine the connectivity within each component via two distinct procedures. We emphasize that this is simply a matter of convenience, not a central feature of the theory, and that the model is neutral with regard to the developmental processes that give rise to adult competence.

The feedforward networks in the four pathways (V→S, A→S, S→N, and S→G) were independently trained on all 200 associations using the on-line back propagation algorithm (Rumelhart, Hinton, & Williams, 1986). Each of these networks contained a single hidden layer of 150 units, and all units in the network used the symmetric activation function to give activities in the range [–1,+1].

The amount of training was limited to embody the architectural assumption that the feedforward net does not have the capacity to map every input to exactly the right output, and hence, the clean-up process is required. We could have embodied this assumption just as well by using a far smaller number of hidden units. Following training, when the network does not produce the correct output, it does produce an answer similar to the correct output (versus, for example, producing an output that would be correct for a similar input). The proximity of the actual output to the correct output is due to the training procedure, which attempts to minimize the discrepancy between the actual and correct output. Typically, one or two of the 200 units in the output pattern would have the wrong sign, and the others would have slight errors, e.g., activity of 0.92 instead of 1.00, or –.85 instead of –1.00.

Due to the localist representation of attractors in the attractor network, it was straightforward to algorithmically wire up each network with the 200 attractors for its domain, using 200 hidden units. In addition, we included one *rest-state* attractor located at the origin of the output. The rest state is the default state the attractor net falls into when no input is present.

The averaging constant, α, was 0.02 in all simulations. All attractor strengths, β, were initialized to the value 15.0 in all simulations, except for the rest-state attractor, whose β was 5.0. The rest-state attractor required a lower strength so that even a weak external input would be sufficient to kick the attractor network out of the rest state. The performance of the model was not terribly sensitive to the choice of α and β.

Priming mechanism

Previous simulations of neuropsychological impairments have accounted for perseveration using priming mechanisms (e.g., Kimberg & Farah, 1993; Plaut & Shallice, 1993b). Priming, by which we mean the increased availability of recently activated states of the system, has been found across a wide variety of tasks in normal subjects. This provides an independent motivation for including it in neuropsychological models, as a parsimonious way of explaining perseveration without hypothesizing any new components or characteristics of the cognitive architecture. In the present model, we implement priming as a strengthening of recently visited attractors. McClelland and Rumelhart (1985) first proposed the idea of priming as a strengthening of weights in a network, and was followed up by Becker, Behrmann, and Moscovitch (1993; Becker, Behrmann, Moscovitch, & Joordens, 1997) and Mathis and Mozer (1996). Priming as weight strengthening differs from the more common approach of allowing residual activation carry over from one trial to the next primarily in that priming effects can persist over several intervening trials.

In a distributed attractor net, the strength of an attractor is determined by connections throughout the network. In contrast, a localist attractor net has the virtue that one particular parameter, β, controls the strength of an attractor. Following an earlier simulation of priming in a localist attractor network (Mathis & Mozer, 1996), we implemented priming by increasing the strength of attractor i, β_i, in rough proportion to its activity,

$$\Delta\beta_i = \kappa(\beta_0 - \beta_i) + \eta\left(\frac{eligibility_i}{\sum_{j \in RBF} eligibility_j}\right) \tag{3}$$

where κ and η are constants, β_0 is the initial attractor strength, and $eligibility_i$ is a quantitative measure of how "eligible" an attractor is for modification, based on how active it has been over the time course of processing:

$$eligibility_i = \sum_t a_i(t) \ .$$

The first term in Equation 3 attempts to pull the attractor strength back to its original value, while the second term normalizes the *eligibility* to prevent the strengths from increasing without bound (due to long exposure of stimuli).

Both *eligibility* and $\Delta\beta$ are computed after the system has relaxed into a well formed-state and before the presentation of the next stimulus. The rest-state attractor was not included in this strength adjustment procedure, as any increase to its strength would have made it difficult for the model to escape the rest state. All attractor nets in the model were fitted with this priming mechanism. In all simulations, κ=0.6, and η=2.2.

Damaging the model

In principle, the model could be damaged by lesioning either the feedforward or the attractor networks, or both. However, we chose to lesion only the feedforward networks, for two reasons. First, patients show no intrinsic deficit in semantics or in naming, suggesting that the components that actually produce semantic and name representations—the attractor nets—are intact. Second, lesioning an attractor net can cause the attractors to shift, making it difficult to interpret the response of the model.

The model was damaged by removing a fraction γ of the connections in the V→S and S→N feedforward networks. The removed connections were chosen at random and an equal fraction was removed from the two pathways. We also explored a second technique for damaging the model: adding Gaussian noise to connections in the V→S and S→N pathways instead of removing connections altogether. Because the two lesioning techniques yielded qualitatively similar results, we report only on the first.[6]

Response classification

Responses were determined after the model was allowed sufficient time to relax into an attractor in either the name or gesture space, depending on the task being performed. The asymptotic attractor state was taken to be the response. Each response was classified as one of the following mutually exclusive response types:

- *correct*: response (output attractor state) corresponds to the presented input pattern.
- *no response*: response is the rest-state attractor.
- *perseveration*: response is the same as that produced on any of the three immediately preceding trials.
- *visual*: the visual pattern corresponding to the incorrect response is a sibling of the visual pattern corresponding to the correct response.
- *semantic*: the semantic pattern corresponding to the incorrect response is a sibling of the semantic pattern corresponding to the correct response.
- *visual+semantic*: an error response that is both *visual* and *semantic*.
- *other*: all other errors.

If a response was both a *no response* and a *perseveration*, it was classified as a *no response*. However, if it was both a *perseveration* and another type of error, it was classified as a *perseveration*.

Testing procedure

After all pathways had been trained, the V→S and S→N pathways were damaged as described earlier. The architecture was damaged a total of 30 different times, creating 30 simulated patients who were tested on each of the four tasks. For each of these simulated patients, the following algorithm was performed:

6. The Gaussian noise damage differed from the connection removal damage in only one significant respect. As we will show in Figure 18, as more connections are removed from the model, and errors on the visual naming task rise above 80%, semantic errors decrease and perseverations an no-responses increase. With Gaussian noise damage, semantic errors decreased but perseverations and no-responses did not significantly increase. This is due to the fact that connection removal results in less overall bottom-up activity, while Gaussian noise damage does not. The less bottom-up activation, the easier it is for a perseverative response or the rest attractor to commandeer the attractor net.

```
Loop through all tasks (V⇒N, V⇒G, A⇒N, A⇒G)
        Loop through all 200 input patterns
                Clamp input pattern
                Run model until output has settled
                Classify response
                Update attractor strengths based on priming mechanism
        Repeat
Repeat
```

The results that we report come from averaging the performance across the simulated patients.

Results

Error rate for visual naming as a function of severity of damage

Figure 11 presents the error rate for the V⇒N task as a function of the amount of damage. The amount of damage is quantified by the parameter γ. With no damage (γ = 0), the model performs perfectly. Even at 10% damage, the model performs without deficit, due to the ability of the attractor nets to compensate for the small amounts of noise introduced by the relatively minor damage. Up to about 50% damage, the error rate curve is positively accelerated. Above 50% damage, the error rate in Figure 11 approaches the ceiling, such that beyond about 70% damage, the connectivity of the model is sufficiently disrupted that the model cannot perform the task at all.

Based on the theoretical curve in Figure 7, we argued that any model that shows positively accelerated error as a function of damage is a likely candidate to produce superadditive effects. Although positively accelerated, the nonsaturating region of the curve in Figure 11 (i.e., up to about 50% damage) does not provide definitive evidence of the sort of superadditivity we have conjectured: Figure 11 only provides evidence of superadditivity of errors on a simgle task across simulated patients or lesions. Our goal is to show superadditivity of errors across tasks for a given simulated patient or lesion. To explore this issue, we focus on simulated patients having an intermediate amount of damage, γ=.30, because this level of damage yields no floor or ceiling effects, and also produces error rates of 30–40% for

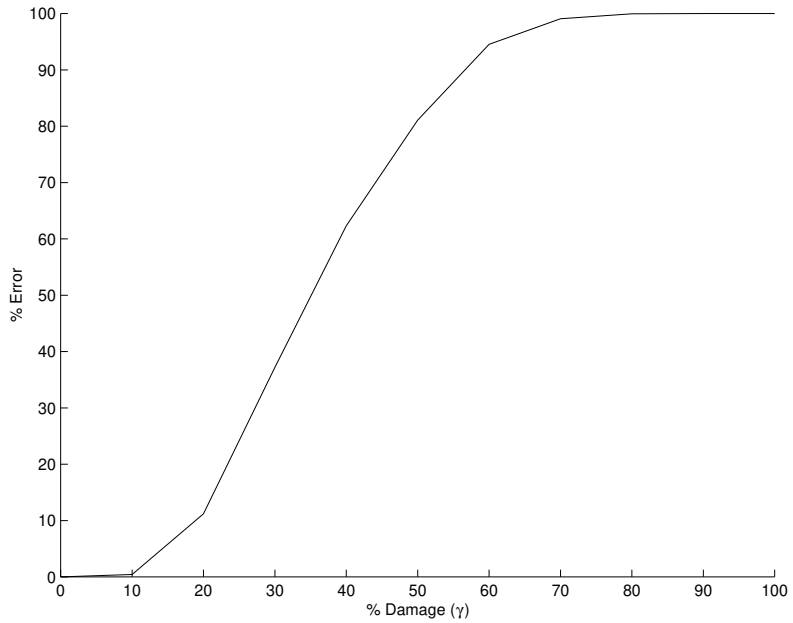

FIGURE 11. Error rate on the visual naming task as a function of the amount of damage to the V→S and S→N pathways.

the V⇒N task, roughly the median performance of patients in the literature.

Superadditivity of errors across tasks

The second column of Table 2 presents the error rates of the generalized-pathway model on four tasks, averaged over simulated patients with a damage level of γ=.30. No errors were produced on the A⇒G task because the two component pathways (A→S and S→G) were undamaged. Relatively few errors were made on the A⇒N and V⇒G tasks, each of which involved one damaged pathway, because the attractor nets were able to compensate for the damage. However, the error rate for the V⇒N task was quite large, due to damage on both of its component pathways (V→S and S→N). The error rate for V⇒N cannot be accounted for by summing the effects of the damage to the two component pathways because the sum of the error rates for A⇒N and V⇒G, each of which involves one of the two partially damaged pathways, is nearly four times smaller. Rather, it appears that the effects of damage on these pathways do in fact interact, and their interaction leads to superadditive impairments.

What is the source of the interaction? One hypothesis is that the interaction arises from the feedforward networks, because the feedforward networks are nonlinear, and superadditivity is one manifestation of nonlinearity. This hypothesis can be evaluated by studying the feedforward-pathway model. The generalized-pathway model can be trivially transformed into a feedforward-pathway model by removing the attractor networks of each pathway. Because the output of the feedforward-pathway model hasn't been cleaned up, we must assume the model's response to be the closest (in Euclidean distance) of the 200 meaningful output patterns. (In other words, we assume a simple sort of clean up on the output to read out from the model.) Error rates on the four tasks for the feedforward-pathway model are shown in the third column of Table 2. Even without explicit clean up, the V⇒N error rate is significantly larger than the sum of the V⇒G and A⇒N error rates, indicating superadditivity. By the ratio measure (Equation 2), superadditivity is more pronounced in the generalized-pathway model than in the feedforward-pathway model; this finding is generally true regardless of the degree of damage (γ).

The comparison between the feedforward-pathway and generalized-pathway models suggests that the nonlinearity of the feedforward nets is only partially responsible for the observed superadditivity in the generalized-pathway model. Superadditivity also arises from the temporal dynamics of cascaded attractor networks, a point we explain further.

When a visual pattern is presented to the model, it is mapped by the damaged V→S pathway into a corrupted semantic representation which is then cleaned up. Although the corruption is sufficiently minor that clean up will eventually succeed, clean up is slowed considerably by the corruption. While the semantic attractor network is searching for the correct attractor, the corrupted semantic representation is being processed by the S→N pathway. Although an undamaged pathway may be able to handle a corrupted input, the damaged S→N pathway is not. Thus, the combination of slowed convergence of the semantic representation *and* damage to the S→N pathway causes corruption of the naming representation beyond the point where it can be cleaned-up properly.

An interaction between the two loci of damage in the model is inevitable, and is not merely a consequence of some arbitrary assumption that is built into our model. To argue this point, we consider two modifications to the architecture which might eliminate the temporal interaction in the damaged model. First, if we allowed the V→S pathway to relax into a attractor state before feeding its output into the S→N pathway, there would be no interaction; the sequential operation of the pathways yields a correct response only if each pathway independent of the other produces a correct output. However, cortical pathways do not operate sequentially, one stage finishing its computation

TABLE 2. Error of damaged model on four tasks

task	error rate	
	generalized pathways	feedforward pathways
A⇒G	0.0%	0.0%
A⇒N	0.5%	0.5%
V⇒G	8.7%	37.8%
V⇒N	36.8%	67.5%

and then turning on the next stage. Moreover, in the undamaged brain, such a processing strategy is maladaptive, as cascading partial results from one pathway to the next can speed processing without the introduction of errors (McClelland, 1979). Second, the interaction might be eliminated by making the S→N pathway continually responsive to changes in the output of the V→S pathway. Then, the rate of convergence of the V→S pathway would be irrelevant to determining the eventual output of the S→N pathway. However, because the output of the S→N pathway depends not only on its input but its internal state (the state of the attractor net), one cannot design a pathway that is continually responsive to changes in the input and is also able to clean up noisy responses. Thus, the two modifications one might consider to eliminate the interactions in the damaged model seriously weaken the computational power of the undamaged model. We therefore conclude that the framework of the generalized-pathway model makes it difficult to avoid a temporal interaction of damage between two pathways.

To summarize, two properties contribute to the observed superadditivity: (1) nonlinearity of the feedforward networks, and (2) the temporal dynamics of cascaded attractor nets, in which the second attractor net in a cascade begins processing the output of a first attractor net before the first net has completed processing. The first property is standard; and the second property is true of any model composed of cascaded attractor nets, such as McClelland and Rumelhart (1981) and Farah et al. (1993).[7] Because at least one of these two properties comes into play in essentially every architecture, and because the two models we explored—the generalized-pathway and feedforward-pathway models—are typical of connectionist architectures in the cognitive modeling literature, we see superadditivity of multiple loci of damage as a general property of connectionist models, not as a curious byproduct of a specific modeling approach.

Why hasn't superadditivity been observed in earlier models? The simple answer is that, although some researchers have exhaustively explored the consequences of lesioning each set of units or connections in a model (e.g., Plaut & Shallice, 1993a), few if any have explored multiple loci of damage to a model. Because many models in the literature are narrow in scope, corresponding to what we have termed a single pathway, it does not make a lot of sense to explore multiple lesions. However, as models of greater complexity and broader scope are developed—models that have multiple functional components or pathways—we expect that increasingly the theoretical justification will emerge to investigate multiple loci of damage.

Modeling the performance of individual patients

Significant variability in performance is observed from one optic-aphasia patient to another (Table 1). In eleven case studies we reviewed, error rates on V⇒N range from 27% to 100%, error rates on V⇒G range from 0% to 75% of the V⇒N error rate, and error rates on A⇒N range from 0% to 40% of the V⇒N error rate. One should not be surprised by this variability, considering that each patient has a distinctive lesion and each study used its own stimulus materials and testing procedure. What the reported cases share in common is (1) a higher error rate on V⇒N than on either V⇒G or A⇒N, and (2) superadditivity (where quantitative results are reported for all three tasks). These two phenomena characterize optic aphasia, and are accounted for by our model via the simulations reported in the previous section.

One might sensibly ask whether our model can account for the full range of variability among patients. However, modeling the behavior of individual patients is of questionable value, for the following reason. The simulation results reported thus far make three key assumptions. First, concerning the nature of damage to the patient, we assumed that the V→S and S→N pathways have lesions of comparable severity. Second, concerning the nature of the experimental tasks, we assumed that the naming (V⇒N and A⇒N) and gesturing (V⇒G and A⇒G) tasks are matched in difficulty and systematicity; our simulation embodies these assumptions by virtue of the fact that the gesturing and naming spaces each have 200 response alternatives, and the mapping from semantics to names or gestures is arbitrary. Third, concerning the scoring procedure, we assumed that same criterion was applied in scoring naming and gesturing responses; one can imagine that a human observer might apply a less stringent criterion in scoring a gesturing response to be correct.

To model the data collected from a specific patient with specific stimulus materials and a specific scoring procedure, we should be allowed to adjust at least six parameters of the model: the degree of damage to the V→S pathway and the possibly nonidentical degree of damage to the S→N pathway, reflecting the lesion of a particular patient; the

7. For example, in the McClelland and Rumelhart interactive-activation model, activation continuously feeds through from the letter level to the word level, even while the letter level continues to "clean up" via the mutually-inhibitory connections among the letters in a given word position.

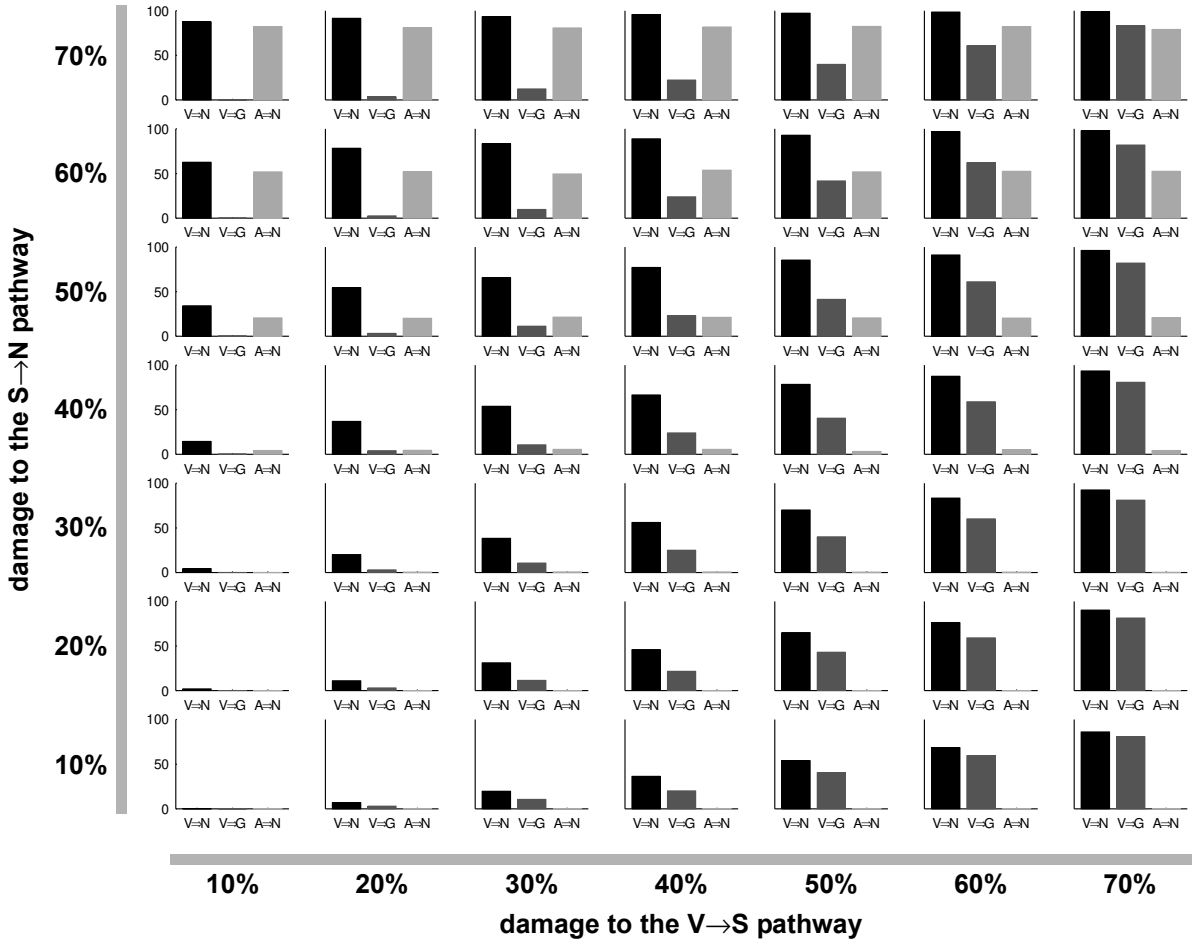

FIGURE 12. Error rates of the damaged model on the V⇒N, V⇒G, and A⇒N tasks based on the degree of damage to the V→S and S→N pathways ($\gamma_{V\to S}$ and $\gamma_{S\to N}$, respectively).

number of response alternatives for the naming and gesturing tasks, reflecting the relative degree of difficulty of each task; and the criteria for judging naming and gesturing responses to be correct, reflecting the scoring procedure.

With these six degrees of freedom, and the fact that the model shows robust superadditivity, accounting for the data from any particular patient does not seem a great challenge. To illustrate the flexibility that the model allows, we systematically varied two of the six degrees of freedom: the degree of damage to the V→S pathway, which we denote $\gamma_{V\to S}$, and the degree of damage to the S→N pathway, which we denote $\gamma_{S\to N}$. Each parameter was varied from 10% to 70% damage, resulting in 49 distinct patterns of damage. For each pattern, we computed the error rates on the affected tasks, averaged over trials and five simulated patients (Figure 12).[8] Where performance deficits are observed, the V⇒N error rate is higher than either the V⇒G or A⇒N error rates. Superadditivity is most pronounced for $\gamma_{V\to S}$ and $\gamma_{S\to N}$ in the 20–50% range, where the damage is moderate and the extent of damage to the two pathways is roughly equal.. For patterns of severe damage, superadditivity is hindered because of the ceiling on the error rate. For patterns in which the damage to the two pathways is unbalanced, the more severely damaged pathway has the bulk of the effect on errors; consequently, the V⇒N error rate does not stand out from the V⇒G error rate (when $\gamma_{V\to S}$ is large) or from the A⇒N error rate (when $\gamma_{S\to N}$ is large). Examining the spectrum of performance, the reader

8. All other results in this paper are based on an average over thirty simulated patients. However, because of the intense computation requirements of this simulation, we ran only five simulated patients per pattern of damage.

can understand why we chose $\gamma_{V\to S} = \gamma_{S\to N} = 30\%$ for the bulk of our simulations; this pattern of damage yields a canonical profile of optic aphasia.

Given the diverse performance reflected in Figure 12, it should not be difficult to account for the performance of individual patients. Consider patient DHY of Hillis and Caramazza (1995), who is interesting because Hillis and Caramazza assessed gesturing performance using stimuli that required highly specirfic, discriminable responses (Riddoch & Humphreys, 1987), thereby providing some assurance that naming and gesturing tasks were matched for difficulty. DHY's error rate was 75% on V⇒N, 65% on V⇒G, and "perfectly normal" on A⇒N. Our model, with $\gamma_{V\to S} = 60\%$ and $\gamma_{S\to N} = 20\%$ produces a similar pattern of errors: 76% on V⇒N, 59% on V⇒G, and 0% on A⇒N. Hillis and Caramazza emphasized that DHY's performance on nonverbal tasks requiring semantic access (e.g., gesturing, card sorting, word-picture verification) was normal when measured using conventional tests, but that significant deficits were observed on tests that required more detailed semantic information. Thus, they argued that DHY has a problem in accessing semantics from visual information. This analysis is consistent with the pattern of damage our model proposes to account for DHY's data—a severe lesion to the V→S pathway, and a mild lesion to the S→N pathway.

Because of the degrees of freedom available to specify the nature of damage, as well as the additional degrees of freedom available to specify the relative task difficulty and scoring procedure, we find it premature to quantitatively model the data from individual patients; doing so will require additional tests of the patient designed to evaluate specific aspects of the model, which we describe later.

In the remainder of this discussion, we focus on accounting for qualitative features of optic aphasia that are robust across patients. Our simulations utilize $\gamma_{V\to S} = \gamma_{S\to N} = 30\%$, although our discussions concerning the model's behavior apply more generally to patterns of damage in which $\gamma_{V\to S}$ and $\gamma_{S\to N}$ are roughly comparable.

Relative deficit in performance on V⇒G and A⇒N tasks

A subtle yet significant aspect of the model's performance is that the error rate on the V⇒G task is reliably higher than the error rate on the A⇒N task, despite the fact that each task makes use of one damaged pathway, and the pathways are damaged to the same degree. The difference in performance is due to the fact that the damaged pathway for the V⇒G task is the first in a cascade of two, whereas the damaged pathway for the A⇒N task is the second. The initially noisy response from a damaged pathway early in a cascade propagates to later pathways, and although the damaged pathway will eventually produce the correct response, subsequent pathways may be unable to attenuate the noise.

Table 1 suggests the same trend in the patient data. Averaging across patients, and excluding cells in the table for which no numerical values were given, one finds a striking difference in mean error rate: 25.0% on V⇒G but only 11.2% on A⇒N. As we discussed in the previous section, these error rates reflect a range of patient lesions, stimulus materials, and scoring procedures, and hence, one cannot declare this a clear victory for a model of a specific lesion based on specific experimental methodology. However, because the model embodies the principle that early lesions in a cascade are more detrimental than late lesions, the predominance of V⇒G errors over A⇒N errors is a strong prediction of the model, regardless of its specific instantiation.

Distribution of errors for visual object naming

Figure 13 presents the distribution of the types of errors made by the model on visual naming. In accordance with the patient data, the model produces many more semantic and perseveration errors than by chance. The chance error proportions were computed by assuming that if the correct response was not made, then all other responses had an equal probability of being chosen. *No response* errors were included in the *other* category as they were too infrequent (0.9%) to be represented clearly in the Figure. The proportion of *no response* errors can be made larger by increasing the strength of the rest-state attractor, and doing so does not appear to interact with any other qualitative properties of the model.

To understand the predominance of semantic errors, consider the effect of damage to the V→S pathway. Remember that the damage involves removal of a fraction γ of connections in the feedforward network. When γ is small, the output of each unit will be close to what it was originally, with small random perturbations. Consequently, the overall mapping produced will be close to the correct mapping (Figure 14). Most of the time, minor perturbation of the mapping will be compensated for by the attractor net. Occasionally, the perturbation will land the model in a different attractor basin, and a different response will be made. However, when the wrong attractor is selected, it will

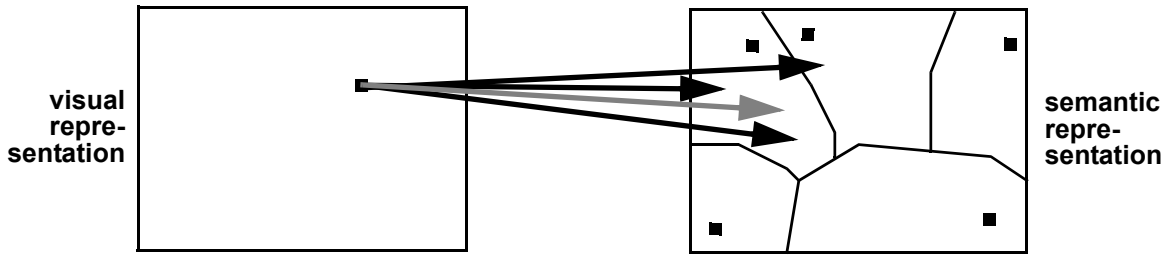

FIGURE 14. A schematic depiction of the V→S pathway. The grey arrow depicts the feedforward mapping in the undamaged pathway. The black arrows depict perturbations to this mapping that result from minor damage to the pathway. Two of the three perturbations land the model in the same attractor basin; the third perturbation lands the model in the attractor basin for a semantic sibling of the correct attractor.

be one "close" to the correct attractor, i.e., it will likely be a sibling in the same pattern cluster as the correct attractor. In the case of the V→S pathway, the siblings of the correct attractor are by definition semantically related. A semantic error will be produced by the model when a sibling semantic attractor is chosen, and then this pattern is correctly mapped to a naming response in the S→N pathway. From this account, one might predict that as the degree of damage increases, the perturbations to the mapping become larger, and semantic errors will no longer dominate. We verify this prediction in a later simulation.

In addition to semantic errors, the other frequent error type in visual naming is perseverations. The priming mechanism is responsible for the significant number of perseverations. By increasing the strength (β) of the most active attractors on one trial, those attractors will be more effective in drawing in states on subsequent trials (Figure 15a). The combination of increasing the strength of an attractor and inaccuracy in the feedforward mapping introduced by damage can give rise to perseverations on subsequent trials (Figure 15b). One cannot reasonably argue that the priming mechanism is merely a "perseveration mechanism," because neither the undamaged model nor the undamaged pathways in the damaged model show perseveration as a result of priming. In fact, in the undamaged model, the priming mechanism facilitates performance (Mathis & Mozer, 1996). Perseveration results from the combination of normal priming and the weakened influence of the current stimulus due to damage. When this influence is weakened, the model tends to return to recently active states. Additional simulations run without a priming mecha-

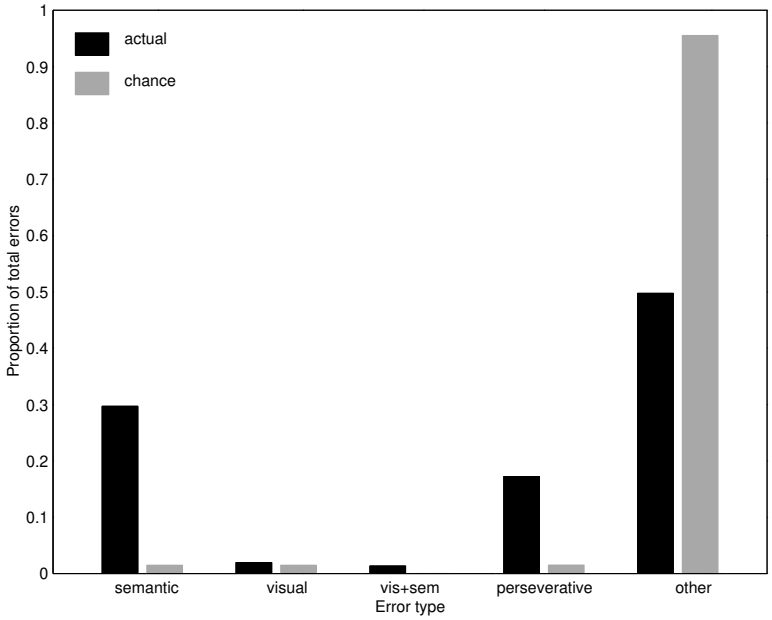

FIGURE 13. Distribution of error types made by model on the V⇒N task (black bars) relative to chance (grey bars). "No response" errors were placed in the "other" response category.

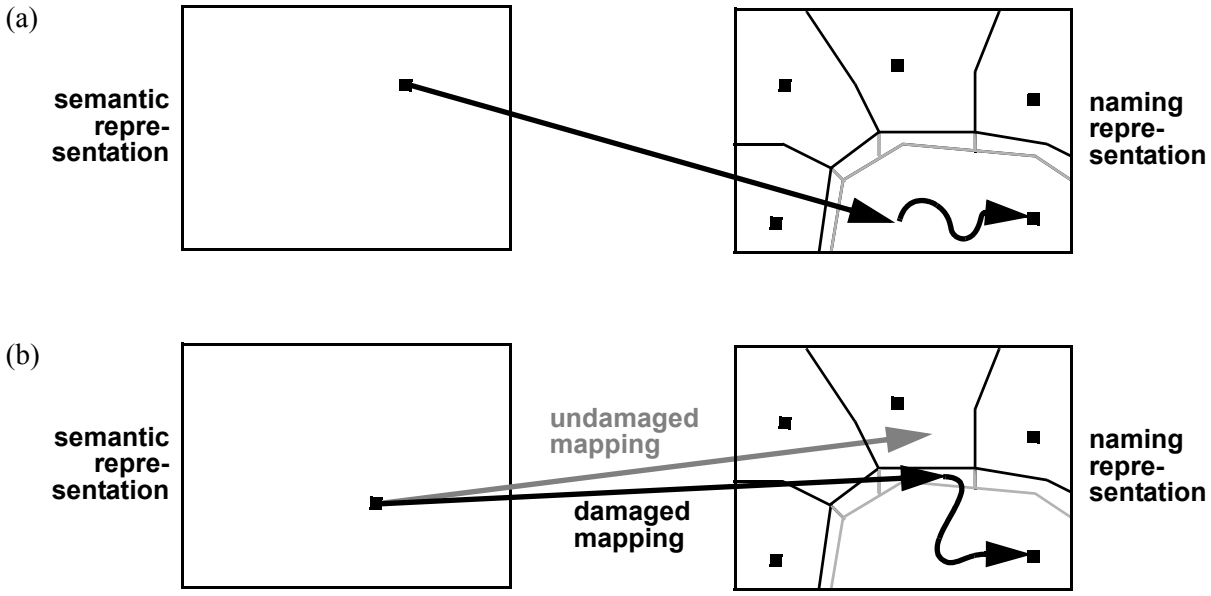

FIGURE 15. A schematic depiction of the sequence giving rise to a perseveration error. (a) On one trial, the model produces a particular naming response. As a result of this experience, the priming mechanism strengthens the selected attractor in the naming space. The lightly shaded region boundary indicates the original basin of attraction for the attractor in the lower right; the solid region boundary indicates the basin that results from the priming mechanism. (b) On presentation of a second trial, the damaged pathway produces a mapping that, in combination with the expanded attractor basin, causes the previously selected attractor to be selected again.

nism showed that its only effect on the model's behavior was in producing perseverations; none of the other features of the model's behavior reported below were changed in a qualitative manner.

Figure 16 shows the distribution of perseveration errors by trial. As expected from the dynamics of the priming mechanism, the model shows nonlocal perseverations, with more recent responses having a stronger perseverative influence. Also consistent with the patient data, perseverations did not occur on the A⇒G (undamaged) task, although the existence of the priming mechanism did allow the output to settle more rapidly if the same input was presented twice.

Just as important as the presence of perseverative and semantic errors is the absence of visual errors, a feature of optic aphasia that contrasts sharply with visual agnosia (Farah, 1990). The same mechanisms explain why the rate of visual errors is close to its chance value and why visual+semantic errors are above chance. V⇒N errors occur because there is an error either in the V→S or S→N mappings, or both. Because the erroneous outputs of these pathways show a strong tendency to be similar to the correct output, and because semantic and name similarity does not imply visual similarity (the patterns were paired randomly), visual errors should only occur by chance. When a visual error does occur, though, there is a high probability that the error is also semantic because of the strong bias that already exists toward producing semantic errors. Therefore, more visual+semantic errors occur than by chance, and the proportion of these errors is only slightly less than the proportion of visual errors.

Plaut and Shallice (1993b) have also proposed a connectionist model to account for the distribution of errors made by optic aphasia patients. Although their model was not designed to account for any of the other phenomena associated with the disorder, it has some features in common with the model we are proposing. Like our model, visual input patterns are fed through a feedforward network to a set of semantic state units. The representations appearing on the state units are then cleaned up by an iterative relaxation search performed by a recurrent network. Unlike our model, however, theirs does not include any kind of naming response system; responses were determined solely from the semantic state. Although the distribution of errors produced by their model was similar to ours in many respects (i.e., more semantic than visual errors and many perseverations), their model produced a larger proportion of visual+semantic errors than any other error type. The reason for this—as we elaborate below—is probably due to an

assumption that visually similar objects are likely to be semantically similar; we refer to this as the *S~V assumption*. Plaut and Shallice constructed semantic patterns such that about one-third of each semantic pattern is devoted to describing the visual characteristics of the object to which it corresponds. Consequently, visually similar patterns tend to be semantically similar as well, which could explain the predominance of visual+semantic errors and the consequent reduction in pure visual errors in their model. Without the S~V assumption, the Plaut and Shallice model would not yield a a low rate of pure visual errors. Evidence for this statement comes from the Hinton and Shallice (1991) model of deep dyslexia, from which Plaut and Shallice's model was adapted: The Hinton and Shallice model lacked the S~V assumption and shows a visual error rate comparable to the semantic error rate (Plaut & Shallice, 1993b).

Our model produces many semantic errors and few visual errors, yet it does not require the S~V assumption. (Incorporating the assumption would give rise to an even greater predominance of semantic errors.) Given that our model has a similar architecture as the Plaut and Shallice model, how can we explain the discrepancy—that our model does not require the S~V assumption to produce the dominance of semantic over visual errors whereas the Plaut and Shallice model does? One possible answer may lie in the different training procedures for the two models. In connectionist networks, similar inputs tend to produce similar outputs, and it usually requires a great deal of training to get them to do otherwise. In Plaut and Shallice's model, both the feedforward net and the attractor net were trained as a single network using back propagation through time (Rumelhart, Hinton, & Williams, 1986). With this training procedure, the responsibility for producing correct semantic states is distributed between the feedforward net and the attractor net. Consequently, if similar inputs need to be made into dissimilar outputs, the feedforward mapping can, for the most part, still get away with producing similar outputs to similar inputs and rely on the attractor net to pull them apart. Because the effects of inducing damage to the system can be likened to changing the boundaries of the attractor basins, damaged mappings will result in not only semantically similar mappings being captured by the same attractor, but also purely visually similar patterns as well (Hinton & Shallice, 1991). However, if the pattern spaces are constructed in such a way that visually similar patterns have corresponding semantically similar patterns, as in Plaut and Shallice's model, then visual errors will usually be accompanied by semantic errors, and hence will be classified as *visual+semantic.* In our model, the feedforward and attractor nets were trained separately—reflecting the distinct functions of the two components. The feedforward net was required to pull the visually similar patterns apart as best it could, minimizing the effect of visual similarity on subsequent processing.

Although we conjecture that the behavior of each model is a consequence of its learning procedures, neither model has an edge in terms of the developmental or biological plausibility of its learning procedures.[9] What must we

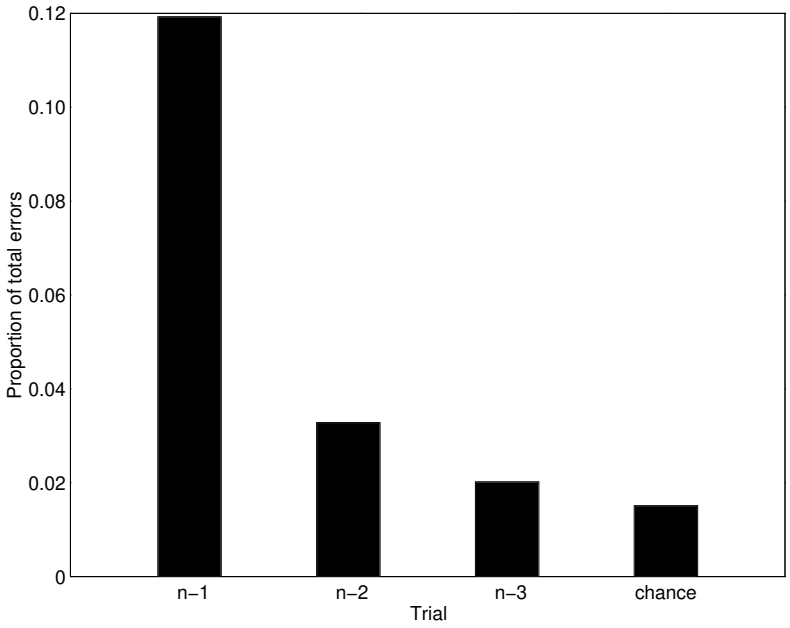

FIGURE 16. Proportion of total errors on trial *n* that were perseverations from trials *n*–1, *n*–2, or *n*–3. The fourth bar shows the chance rate of errors that would be classified as perseverations.

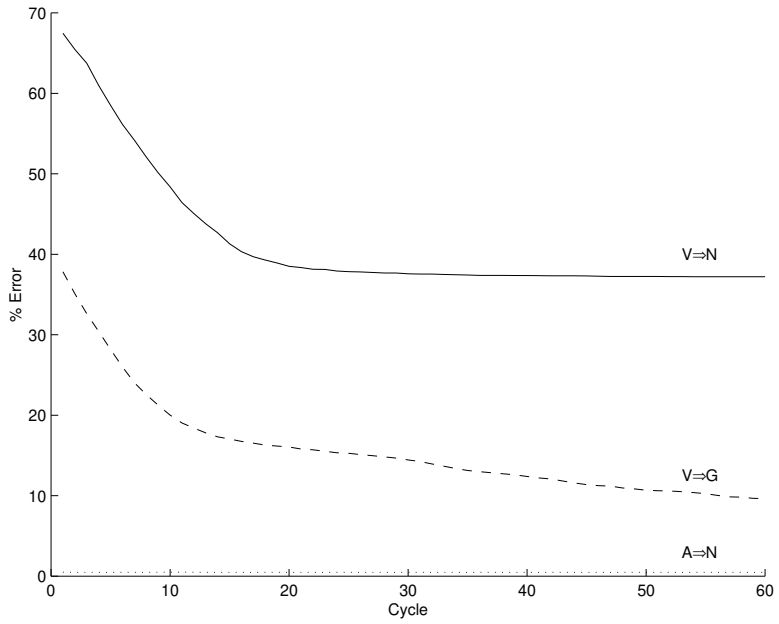

FIGURE 17. Probability of an error response for V⇒N, V⇒G, and A⇒N tasks as a function of the number of cycles the model is allowed to process an input stimulus. If the attractor net in the response pathway has not settled on an output by a given point in time, the response assumed is the attractor closest to the current state of the system. The V⇒N and V⇒G tasks show homing-in behavior, in that the model is more likely to produce the correct response given increased processing time.

do to resolve which model is the best account of the error distribution in optic aphasia? First, it must be determined whether the Plaut and Shallice S~V assumption is valid for visual semantics in the world. If it were true, we could incorporate the S~V assumption into our model with no qualitative change in performance; if it were not true, removing the S~V assumption from the Plaut and Shallice model would impair its ability to account for data. Second, although each model can explain certain qualitative aspects of the data, the models appear to make opposite predictions on another aspect of the data: visual+semantic errors dominated over pure semantic errors in the Plaut and Shallice model, whereas pure semantic errors dominate in our model. One might hope that these predictions could be evaluated via a careful examination of the patient data; however, for the absolute error rates produced by the model to be meaningful, the models must also produce the same chance error rates as patients. Thus, evaluating the seemingly contradictory predictions of the two models will require additional refinement of the models.

Homing in

Another distinct characteristic of optic aphasia is the tendency of a patient to "home in" on the correct name for a visually presented object when given sufficient time. Our model provides an interesting possible explanation for this phenomenon. The simulation results reported previously are obtained by allowing the model to settle into an attractor before a response is initiated. However, one could also force the model to respond earlier with its best guess. Figure 17 shows the error rate for three tasks as a function of the time at which the model is forced to respond. The response at a particular time is the most active attractor, excluding the rest state. As shown in the figure, the V⇒N task shows a speed-accuracy trade off. *This tradeoff is extended several orders of magnitude of processing time over that of the*

9. The critical difference in learning procedures is that the Plaut and Shallice model uses supervised learning for training both the feedforward and attractor nets, whereas our model uses a supervised procedure for training the feedforward net and—as an alternative to predetermined wiring—might use an unsupervised procedure for training the attractor net (to reflect the statistics of the environment). Such broad assumptions about learning are unlikely to be validated or invalidated.

undamaged model; for example, the A⇒G task, which involves no damaged pathways, gives the correct response after just a couple of processing cycles. The temporally extended performance on the V⇒N task is due in large part to the fact that the S→N pathway remains partially sensitive to the changing output of the V→S pathway as it converges on an attractor in semantic state.

Because guesses of the model improve with increased processing time, this behavior might be viewed as homing in on the correct response. However, this is not "homing in" in the strict sense that one guess helps improve accuracy on the next guess. Instead, this account suggests that patients are simply reporting their current mental states, and homing in is a reflection of the greatly slowed time course of processing in these brain damaged patients.

This account of homing in requires that we re-examine the simulation results we reported earlier in which model settles into an attractor before it initiates a response. It turns out that regardless of the point in time at which processing is terminated—whether the model be allowed to settle to an attractor or is stopped after only a small number of processing steps— the model's behavior is qualitatively the same (e.g., magnitude of the superadditive impairment, the distribution of errors, etc.)

Our explanation of the homing-in effect may not be the entire story. It also seems plausible that, given relatively intact auditory comprehension, patients hear their own responses and use the auditory feedback to verify or refine the responses. The use of auditory feedback seems unlikely to be the entire explanation for the homing-in effect, because patients occasionally *pass through* but not stop at the correct name during their circumlocutions. In the next section, we propose an experimental test that distinguishes the two accounts of homing in, based on a test in which patients are asked to provide responses following a variable delay.

Novel predictions of the model

During the course of this simulation study, we noticed several robust behaviors displayed by the model that have not previously been associated with optic aphasia. In this section, we mention these novel behaviors which lend themselves to experimental testing. Many of the behaviors are conditional upon lesions of roughly equal magnitude to the the V→S and S→N pathways. Of course, selecting patients on the basis of the relative severity of multiple functional lesions is impossible. Nonetheless, useful predictions can be extracted from the model, because the behavior of the model with lesions of equal severity is similar to the mean behavior of the model with random variation in the relative severity of lesions. (For example, performance of the model with $\gamma_{\mathbf{V\to S}} = \gamma_{\mathbf{S\to N}} = 30\%$ is comparable to the average performance of two models, one with $\gamma_{\mathbf{V\to S}} = 40\%$ and $\gamma_{\mathbf{S\to N}} = 20\%$, and the other with $\gamma_{\mathbf{V\to S}} = 20\%$ and $\gamma_{\mathbf{S\to N}} = 40\%$.) Thus, predictions arising from the model with equal damage to the two pathways can be evaluated via the mean performance across multiple patients, even though the predictions might not hold true for every patient who is classified as an optic aphasic by somewhat unprincipled and nonquantitative neuropsychological criteria.

- Our model shows a high error rate on the V⇒N task due to a synergy between the two lesion sites. This synergy, and therefore the high error rate, is eliminated if the two damaged pathways operate sequentially rather than in cascade. With sequential operation, the error rate of our moderately-damaged model drops from 36.8% to approximately 1%. Patients might be encouraged to employ such a sequential operation strategy if shown a visual object and were prevented from immediate naming or gesturing by a distractor task. Following a delay, they would then receive a cue as to which task—naming or gesturing—to perform. Although we were somewhat skeptical about this prediction at first, there is evidence is at least consistent with this prediction. It turns out that some patients have difficulty with the V⇒G task because they cannot help but to try naming the object before producing the gesture, which causes them to gesture in accordance with the names they produce, which are usually incorrect. However, Schnider, Benson, and Scharre (1994) discovered a method to suppress naming during the V⇒G task. Using this technique, their patient made no errors on the V⇒G task. Moreover, if they asked the patient to name the object after gesturing, he always produced the correct name. Although this evidence is consistent with our prediction, an alternative interpretation is that the name was accessed via kinesthetic feedback during gesturing.
- Given comparable lesions to the V→S and S→N pathways, and controlling for the difficulty of the naming and gesturing tasks, our model shows a higher error rate for V⇒G than for A⇒N tasks. The V⇒G task taps a damaged pathway early in the processing stream, whereas the A⇒N task taps a damaged pathway late in the processing stream. The effects of damage early in the processing stream propagate to later pathways, and hence have a greater impact. The trend in the literature supports the proposition that the V⇒G error rate is higher than the A⇒N error rate (Table 1), all the more impressive considering that without controls on the

TABLE 3. Superadditivity as a function of amount of damage to model

Amount of Damage (γ)	Superadditivity
10%	25.00
20%	4.73
30%	3.99
40%	2.52
50%	1.68
60%	0.86
70%	0.61

relative difficulty of naming and gesturing tasks, gesturing is probably a coarser response than naming, and hence experimental tests are likely to have provided a lower bound for the error rate on the V⇒G task when equated for difficulty with the A⇒N task.

- Our model shows a positive correlation between the V⇒N and V⇒G error rates. The correlation in the model is premised on equating naming and gesturing for difficulty, and roughly equal damage to the two lesioned pathways. For the patients listed in Table 1, the relationship does seem to hold (correlation coefficient is 0.65), but more cases are required.
- As the amount of damage increases, the V⇒N error rate rises but the model shows less superadditivity, as quantified by Equation 2—the ratio of the V⇒N error rate to the sum of the V⇒G and A⇒N error rates. Table 3 shows superadditivity as a function of amount of damage in the model. From this Table, it is clear that there are limits as to how strong the interaction can be between the two damaged pathways. Unfortunately, this prediction cannot be evaluated, even tentatively, by the patient data in Table 1.
- As shown in Figure 16, the model produces homing in on the V⇒G task as well as the V⇒N task. In both tasks, this is due to the fact that the output pathways maintain sensitivity to the changing output of the damaged V→S pathway as it attempts to clean up its corrupted representations. However, because gestures are not produced instantaneously and take variable amounts of time to be recognized, this phenomenon may be difficult to observe in patients. Nonetheless, it might be possible to bring out homing-in with the V⇒G task using deadline procedures.
- As shown in Figure 17, accuracy of the model's responses increases with processing time. This improvement could be responsible at least in part for the homing-in effect. That is, even when patients are asked not to verbalize guesses and formulate hypotheses about a visual stimulus, the model predicts that the dependence of response accuracy on processing time should still be observed. Patients could be asked to give a single response at various delays following stimulus presentation, and the accuracy versus delay function could be compared to that of Figure 17 and that of the standard homing-in behavior.
- As the amount of damage increases, our simulations show that the distribution of error types changes in the V⇒N task. In particular, the errors contain a lower proportion of *semantic* errors and greater proportion of *no response* and *perseveration* errors (Figure 18). Further, as the amount of damage approaches full disconnection (100% damage), *no response* errors dominate the distribution. The reason the proportion of *semantic* errors begins to fall away with increased damage has to do with the fact that when the V→S pathway is severely damaged, visual patterns have a smaller probability of being mapped to a semantic sibling. *Perseveration* and *no response* errors become more prominent because with increased damage, feedforward nets produce less overall distinctive activation. This reduction in activity amplifies the bias toward previous states (leading to *perseverations*) and the rest-state attractor (leading to *no response*).
- A final prediction suggested by the model is that partial damage to cascading pathways other than V→S and S→N could result in superadditive impairments. According to the model, it is possible, at least in principle, that partial damage to, say, the V→S and S→G pathways could result in an "optic apraxia," wherein patients would show an isolated deficit in gesturing the use of a visually presented object. Similar possibilities exist for partial damage to the A→S and S→G pathways, or the A→S and S→N pathways. Have there been any

reported cases of individuals with such isolated deficits on any of these tasks? The answer appears to be yes for at least two of them. Facial apraxia, a rather common clinical phenomenon, is marked by a deficit in carrying out facial movements upon verbal command, despite relatively intact comprehension of such commands and the ability to produce the correct facial movement if the stimulus is presented visually. For example, a patient may not produce the correct facial movements for blowing out a match or sucking through a straw when asked to do so, but can readily do so when a match or straw is shown to them (Geshwind, 1974). An analogous case of "auditory aphasia" was described by Denes and Semenza (1975), in which the patient showed an isolated deficit on auditory naming. Also, if we assume that there is a pathway from somatosensory input to semantics, it appears that a case of "bilateral tactile aphasia" described by Beauvois et al. (1978), could also be accounted for by superadditive effects of damage. There does not appear to be any case displaying the type of "optic apraxia" described above, but the possibility of such cases is a prediction of the model. Actually finding such cases may be another matter; their apparent rarity may simply be due to anatomical considerations. Because the model does not make direct claims about the neuroanatomical placement of its components, it could be that the A→S and S→G pathways do not lie next to each other in the brain, and are therefore unlikely to be affected by a single locus of damage.

Discussion

By lesioning a connectionist model, we have verified the sufficiency of Farah's (1990) hypothesis that partial damage to two processing pathways may result in close-to-normal performance on tasks involving one pathway or the other while yielding a severe performance deficit on tasks involving both damaged pathways. In demonstrating superadditive effects of damage, we have offered an account that explains the primary phenomenon of optic aphasia: severe impairments in visual naming in conjunction with relatively spared performance on naming from verbal description or gesturing the appropriate use of a visually presented object. In addition, several unanticipated features of the model's performance paralleled the performance of patients, further supporting the superadditivity hypothesis. Finally, our model makes predictions about optic aphasia that can be experimentally evaluated in the future.

In the remainder of this discussion, we elaborate the implications of the model for optic aphasia in particular and for cognitive neuropsychology more generally.

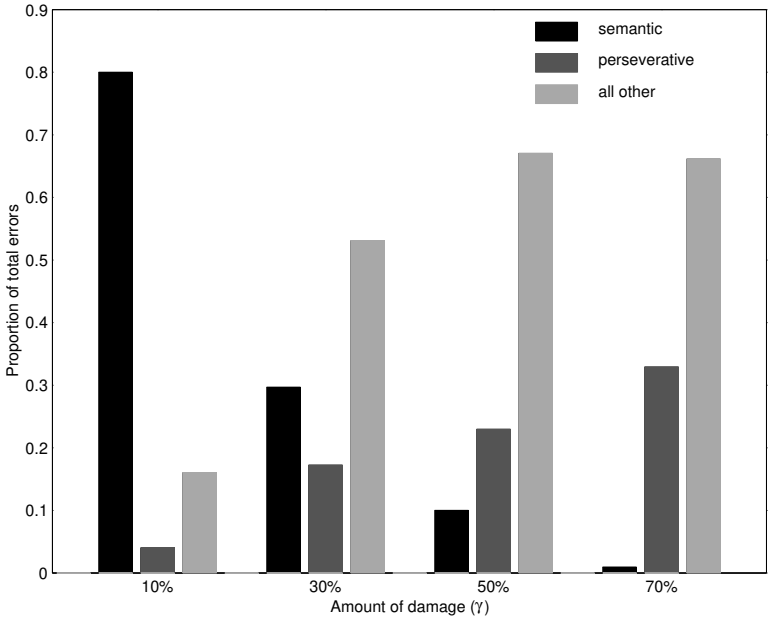

FIGURE 18. Distribution of errors produced by the model as a function of the amount of damage to the model.

Implications for optic aphasia

The superadditive-impairment hypothesis joins a small number of other hypotheses, reviewed earlier, in providing an explanation of the core, defining features of optic aphasia. Although further development and testing of these hypotheses is needed to decide which, if any, is the correct explanation in any given case, the superadditive-impairment hypothesis has some additional points in its favor, beyond merely accounting for the disproportionate impairment of visual naming that is the defining feature of the disorder.

One point in its favor is theoretical parsimony, specifically the parsimonious view of the normal cognitive architecture entailed by this account of optic aphasia. We need hypothesize only a single semantic memory system, interposed in a straightforward way between perceptual input modalities and output response systems such as language and gesture, as illustrated in Figure 6. In contrast, the other accounts require that the normal cognitive architecture contains specialized subdivisions of semantic memory or specialized pathways, as shown in Figures 2 through 5.

The superadditive-impairment hypothesis does, of course, require two functional lesions rather than one, and in this sense is less parsimonious. The loss of parsimony resulting from two hypothesized lesions is minimal, however, because of two further considerations: the independent evidence of multiple functions associated with the left occipito-temporal region and the rarity of optic aphasia. The first consideration implies that a single anatomical lesion is likely to cause multiple functional lesions, and therefore eliminates the need for multiple independent neurological events. All that is required is a particular correspondence between the premorbid organization of this brain area with respect to visual and semantic processes and the location and extent of a single lesion. Although such a correspondence would not necessarily be expected to occur routinely with lesions in this area, the second consideration eliminates the need to assume that it does.

In addition to preserving the parsimonious canonical model of the cognitive architecture underlying naming, the superadditive-impairment hypothesis has a further advantage over competing hypotheses in that it accounts naturally for five associated features of optic aphasia: (1) the semantic nature of the naming errors made by optic aphasics, (2) the absence of visual errors, (3) the tendency to perseverate responses from one trial to the next, (4) the homing-in process whereby a correct response is often preceded by semantically related incorrect responses, and (5) the trend for naming verbally-defined objects to be more accurate than identification by gesture of visually presented objects. These features are not simply accomodated by the model; the model cannot help but produce them. They can therefore be viewed as a series of opportunities to disconfirm the model, which by surviving them gains further credibility.

A longstanding issue concerning optic aphasia is its relation to visual agnosia, specifically, whether they are distinct syndromes or on a continuum of some sort (e.g., Bauer & Rubens, 1985; Davidoff & Bleser, 1993; Farah, 1990; Geschwind, 1965; Humphreys & Riddoch, 1987; Kertesz, 1987; Schnider, Benson, & Scharre, 1994). The present account offers an answer to this question, which reconciles many of the distinctive features of optic aphasia (its dependence on output modality, its semantic errors, and absence of visual errors) with its similarities to visual agnosia (visual-modality specificity) and the tendency for some cases of visual agnosia to evolve longitudinally into optic aphasia. Our account also addresses the relation between optic aphasia and anomia, evident in the work of Bisiach (1966) described earlier.

Figure 19 shows the relation between optic aphasia, visual agnosia, and anomia according to the model. The horizontal axis corresponds to $\gamma_{\mathbf{V\rightarrow S}}$, the degree of damage to the pathway that maps visual information to semantic representations, and the vertical axis corresponds to $\gamma_{\mathbf{S\rightarrow N}}$, the degree of damage to the pathway that maps semantic representations to naming responses. Figure 19 is intended to capture the general pattern that emerges from the simulation results in Figure , in which Optic aphasia results from roughly equal and moderate damage to the two pathways. If $\gamma_{\mathbf{V\rightarrow S}}$ is large, the model performs poorly on all visual tasks, and therefore behaves as a visual agnosic. If $\gamma_{\mathbf{S\rightarrow N}}$ is large, the model performs poorly on all naming tasks, and therefore behaves as an anomic. Thus, optic aphasia can be placed on a continuum with other disorders, but the continuum lies in a two dimensional space.

It should be noted that the current implementation of the model fails to capture one key property of visual agnosia, namely the high incidence of visual errors. This failure is not due to the model being wrong; rather, the model is neutral because we have focused on only those pathways directly relevant to explaining optic aphasia. Visual errors could arise from damage to early visual processing, which is not part of the present model. To simulate the high incidence of visual errors in agnosic performance, the model would have to be expanded to include earlier visual pathways, feeding into the V→S pathway, and visual errors would result from damage encompassing this input. The evolution of visual agnosia into optic aphasia, which is the prime evidence favoring the idea of a continuum, can then be understood in terms of the recovery of the V→S pathway and its inputs.

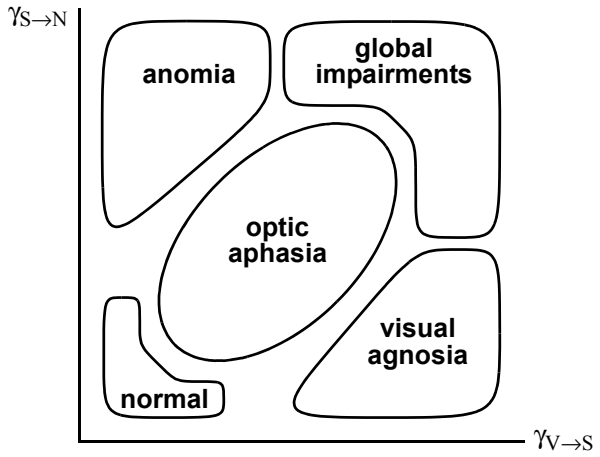

FIGURE 19. A two dimensional space representing the possible combinations of damage to the V→S pathway ($\gamma_{V\rightarrow S}$) and damage to the S→N pathway ($\gamma_{S\rightarrow N}$). Optic aphasia results from roughly equal and moderate damage to the two pathways, whereas visual agnosia and anomia result from other patterns of damage. This sketch is an abstraction from Figure 12, in which $\gamma_{V\rightarrow S}$ and $\gamma_{S\rightarrow N}$ were manipulated independently in simulations.

More generally, this account suggests a continuum of disruption to visual and naming processes. Rather than labeling patients as "optic aphasics," "visual agnosics," or "anomics," one might use this continuum to place each patient more precisely in the two-dimensional space.

Implications for cognitive neuropsychology

Beyond the explanation of optic aphasia, the present model also has broader implications for neuropsychology. As noted earlier, a selective impairment in a specific ability need not result from damage to a module specialized for that ability in the normal cognitive architecture. Rather, the manifest behavioral impairment may result from a complex interaction between the damaged component and the intact components.

Our model of optic aphasia illustrates yet another way in which behavioral deficits to a damaged cognitive system can be attributed to nonlocal effects of local lesions. In this case, a lesion violates locality not by altering the functioning of intact components, but by altering the functioning of a second lesioned component: Whereas the effects of the S→N lesion in isolation are minimal, this lesion results in erroneous mappings of semantics to names when there is also a V→S lesion (which, in isolation, is also benign).

It may well be possible to understand other highly selective cognitive impairments in this way, in terms of the superadditive effects of multiple lesions. Candidate syndromes for explanation in terms of superadditive lesion effects are all those in which an impairment appears selective along two dimensions that are associated with different components or levels of processing within the cognitive architecture. Optic aphasia fits this description because the impairment is selective for both the dimension of input modality (visual as opposed to auditory or tactile) and the dimension of response mode (verbal as opposed to gesture or sorting action), and these dimensions are associated with blatantly different components, namely perceptual input components and response output components. A variety of other neuropsychological impairments present similarly puzzling combinations of selectivity.

In the domain of spatial attention, some cases of hemispatial neglect have been found to be highly selective for certain stimulus categories. For example, neglect may be manifest only for the left sides of faces and not other kinds of objects (Young et al., 1990). This is puzzling when considered within the framework of single-lesion hypotheses, for the following reason. Although there is considerable evidence that visual recognition mechanisms can be subdivided along the dimension of face versus nonface processing, and that the left-right axis is an important organizing dimension for the spatial attention system, there is no known component of the cognitive architecture that is associated with both the face-nonface dimension and the left-right dimension. Face-specific recognition processes are part of the high-level visual processing of ventral visual areas, whose visual representations are abstracted from the spatial topography of earlier visual areas (Desimone & Ungerleider, 1989). Visuospatial attention processes are part of the

dorsal visual system, whose representations of spatial location are neutral with respect to the type of stimulus (Desimone & Ungerleider, 1989).

One interpretation of facial neglect is that it provides the first evidence of a face-specific spatial attentional mechanism, and this is indeed the interpretation favored by Young and colleagues. However, a simpler alternative is suggested by the phenomenon of superadditive lesion effects. Perhaps facial neglect is the result of subclinical impairments of both face perception and visuospatial attention, whose effects are synergistic and manifest only when faces are the focus of attention. The appeal of this account is that it requires only the two components already known to exist, namely general-purpose spatial attention and face-specific perceptual mechanisms, and does not require us to hypothesize a new component dedicated to facial attention.

Similar considerations arise with findings of neglect specific to representations of the human body (Guariglia & Antonucci, 1992), to number (Cohen & Dehaene, 1991), and even separately to reading (Costello & Warrington, 1987) and writing (Baxter & Warrington, 1983). Application of the standard pattern of inference in cognitive neuropsychology leads to a profusion of stimulus specific attentional systems, and this has been the conclusion endorsed by the authors just cited, as well as by Umilta (1994) in a general review of selective forms of neglect. Alternatively, it may be possible in each case to avoid hypothesizing new attentional systems by instead hypothesizing superadditive effects of lesions to a general-purpose spatial attentional mechanism and to other systems for which we already have independent evidence, including systems for representing the body schema, numbers, visual word recognition, and writing.

Among the stimulus-specific forms of neglect, the most challenging cases to explain without recourse to stimulus-specific attentional systems are those in which a single patient has neglect for one side of words and for the opposite side of nonword stimuli. However, a single general-purpose spatial attention mechanism can be maintained, in principle, by a different (but not superadditive) dual lesion mechanism, in this case involving bilateral lesions to the spatial attention system. Reading is known to activate the left hemisphere and thereby exacerbate left neglect (Bowers & Heilman, 1980). Consider the distribution of attention resulting from bilateral lesions, in which the left hemisphere lesion is slightly larger than the right. Under default conditions, attention will be biased to the left, producing right neglect. Presentation of verbal materials and consequent activation of the left hemisphere will counteract this imbalance, and could in some cases result in neglect of the left. This account makes a strong prediction: the neglect for words must always be for the left side and the neglect for nonwords must always be for the right side. In all published cases this prediction is confirmed.

In the domain of language, similarly perplexing impairments have been reported, combining selectivity along dimensions normally associated with different components of the cognitive architecture. For example, Caramazza and Hillis (1991) describe two patients with strikingly restricted language impairments. Case HW made errors in oral reading and picture naming, in which a spoken response was required, but made no errors when writing to dictation or writing the name of an object. In contrast, case SJD was good at oral reading and naming, but made errors when responses had to be written. These patterns of performance are consistent with damage to phonological and orthographic output lexicons, respectively. However, the patients' patterns of performance showed an additional dimension of selectivity, which cannot be accounted for in this way. HW was substantially more impaired with verbs than with nouns, and SJD showed the reverse pattern. Thus, the impairments were selective along both grammatical and output-modality dimensions.

The authors conclude that each output lexicon, phonological and orthographic, is subdivided into separate components representing nouns and verbs. However, these patterns of behavior can be explained within a simpler cognitive architecture by dispensing with the assumption that only a single cognitive component has been damaged in each case, and invoking the superadditive effects of dual lesions. Specifically, HW can be hypothesized to have lesions in the phonological output lexicon (organized along purely phonological dimensions) and in some syntactic component specialized for verbs (and unrelated to a particular input or output modality), and SJD can be hypothesized to have lesions in the orthographic output lexicon (organized along purely orthographic dimensions) and in some syntactic component specialized for nouns (and unrelated to a particular input or output modality).

Acknowledgements

This research was supported by grant 97-18 from the McDonnell-Pew Program in Cognitive Neuroscience and NSF award IBN-9873492 to MM, and grant R01NS34030 from NIH to MF. This work benefited from the helpful feed-

back of John Duncan, David Plaut, Marlene Behrmann, Richard Zemel, and an anonymous reviewer. All three authors contributed equally to this work.

References

Assal, G., & Regli, F. (1980). Syndrome de disconnection visuo-verbale et visuo-gesturelle. *Revue Neurologique*, *136*, 365–376.

Bauer, R. M., & Rubens, A. B. (1985). Agnosia. In K. M. Heilman & E. Valenstein (Eds.), *Clinical Neuropsychology*. New York: Oxford University Press. 2nd edition.

Baxter, D. M., & Warrington, E. K. (1983). Neglect dysgraphia. *Journal of Neurology, Neurosurgery, & Psychiatry*, *46*, 1073–1078.

Beauvois, M. F. (1982). Optic aphasia: A process of interaction between vision and language. *Philosophical Transactions of the Royal Society B, 298*, 35–47.

Beauvois, M. F., & Saillant, B. (1985). Optic aphasia for colours and colour agnosia: A distinction between visual and visuo-verbal impairments in the processing of colours. *Cognitive Neuropsychology*, *2*, 1–48.

Beauvois, M. F., Saillant, B., Meininger, V., & Lhermitte, F. (1978). Bilateral tactile aphasia: A tactoverbal dysfunction. *Brain*, *101*, 381–401.

Becker, S., Behrmann, M., & Moscovitch, K. (1993). Word priming in attractor networks. *Proceedings of the Fifteen Annual Conference of the Cognitive Science Society* (pp. 231–236). Hillsdale, NJ: Erlbaum

Becker, S., Behrmann, M., & Moscovitch, K., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1059–1082.

Bisiach, E. (1966). Perceptual factors in the pathogenesis of anomia. *Cortex*, *2*, 90–95.

Bowers, D., & Heilman, K. (1980). Material-specific hemispheric activation. *Neuropsychologia*, *18*, 309–318.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*, 788–790.

Cohen, L., & Dehaene, S. (1991). Neglect dyslexia for numbers? A case report. *Cognitive Neuropsychology*, *8*, 39–58.

Coltheart, M. (1984). Editorial. *Cognitive Neuropsychology*, *1*, 1–8.

Coslett, H. B. (1997). Acquired dyslexia. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral Neurology and Neuropsychology*. New York: McGraw Hill.

Coslett, H. B., & Saffran, E. M. (1989). Preserved object recognition and reading comprehension in optic aphasia. *Brain*, *112*, 1091–1110.

Costello, A. D., & Warrington, E. K. (1987). The dissociation of visuospatial neglect and neglect dyslexia. *Journal of Neurology, Neurosurgery, & Psychiatry*, *50*, 1110–1116.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. (1996). A neural basis for lexical retrieval. *Nature*, *380*, 499–505.

Davidoff, J., & Bleser R. D. (1993). Optic aphasia: A review of past studies and reappraisal. *Aphasiology*, *7*, 135–154.

Démonet, J., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J., Wise, R., Rascol, A. Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, *115*, 1753-1768.

Denes, G., & Semenza, C. (1975). Auditory modality-specific anomia: Evidence from a case of pure word deafness. *Cortex*, *11*, 401–411.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology*. Amsterdam: Elsevier.

Farah, M. J. (1990). *Visual agnosia*. Cambridge, MA: MIT Press/Bradford Books.

Farah, M. J., O'Reilly, R. C., & Vecera, S. P. (1993). Dissociated overt and covert recognition as an emergent property of a lesioned neural network. *Psychological Review*, *100*, 571–588.

Farah, M. J. (1994). Neuropsychological inference with an interactive brain. *Brain and Behavioral Sciences*, *17*, 43–104.

Farah, M. J. (1995). Current issues in the neuropsychology of image generation. *Neuropsychologia*, *33*, 1455–1471.

Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioral Brain Research*, *76*, 181–189.

Feinberg, T. E., Schindler, R. J., Ochoa, E., Kwan, P. C., & Farah, M. J. (1994). Associative visual agnosia and alexia without prosopagnosia. *Cortex*, *30*, 395–412.

Ferro, J. M., & Santos, M. E. (1984). Associative visual agnosia: A case study. *Cortex*, *20*, 121–134.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. Part II. *Brain*, *88*, 585–645.

Geschwind, N. (1974). The apraxias: Neural mechanisms of disorders of learned movement. *American Scientist*, *63*, 188–195.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gil, R., Pluchon, C., Toullat, G., Michanau, D., Rogez, R., & Levevre, J. P. (1985). Disconnexion visuo-verbale (aphasie optique) pour les objets, les images, les couleurs et les visages avec alexie 'abstractive'. *Neuropsychologia*, *23*, 333–349.

Guariglia, C., & Antonucci, G. (1992). Personal and extrapersonal space: A case of neglect dissociation. *Neuropsychologia*, *30*, 1001–1009.

Hillis, A. E., & Carramaza, A. (1995). Cognitive and neural mechanisms underlying visual and semantic processing: Implications from "optic aphasia." *Journal of Cognitive Neuroscience*, *7*, 457–478.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*, 74–95.

Hodges, J. R., & Patterson, K. (1995). Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia*, *33*, 441–459.

Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C. Frackowiak, R. (1992). The cortical localization of the lexicons. *Brain*, *115*, 1769-1782.

Humphreys, G. W., & Riddoch, M. J. (1987). The fractionation of visual agnosia. In G. W. Humphreys, & M. J. Riddoch (Eds.), *Visual Object Processing: A Cognitive Neuropsychological Approach*. London: Lawrence Erlbaum Associates.

Kertesz, A. (1987). The clinical spectrum and localization of visual agnosia. In G. W. Humphreys, & M. J. Riddoch (Eds.), *Visual Object Processing: A Cognitive Neuropsychological Approach.* London: Lawrence Erlbaum Associates.

Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General*, *122*, 411–428.

Klein, D., Milner, B., Zatorre, R. J., Meyer, E., Evans, A. C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Sciences*, *USA*, *92*, 2899-2903.

Larrabee, G. J., Levin, H. S., Huff, F. J., Kay, M. C., & Guinto, F. C. (1985). Visual agnosia contrasted with visual-verbal disconnection. *Neuropsychologia*, *23*, 1–12.

Lhermitte, F., & Beauvois, M. F. (1973). A visual-speech disconnexion syndrome: Report of a case with optic aphasia, agnosic alexia and colour agnosia. *Brain*, *96*, 695–714.

Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior.* New York: Pergammon Press.

Manning, L., & Campbell, R. (1992). Optic aphasia with spared action naming: A description and possible loci of impairment. *Neuropsychologia*, *30*, 587–592.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L. Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, *270*, 102-105.

Mathis, D. W., & Mozer, M. C. (1995). On the computational utility of consciousness. In G. Tesauro, D. S. Touretzky,

& T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7* (pp. 10–18). Cambridge, MA: MIT Press.

Mathis, D. W., & Mozer, M. C. (1996). Conscious and unconscious perception: A computational theory. In G. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 324–328). Hillsdale, NJ: Erlbaum.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88*, 375–407.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.

Mozer, M. C., & Behrmann, M. (1990). On the interaction of spatial attention and lexical knowledge: A connectionist account of neglect dyslexia. *Cognitive Neuroscience*, *2*, 96–123.

Mummery, C. J., Patterson, K., Hodges, J. R. Wise, R. J. S. (1996). Generating 'tiger' as an animal name of a word beginning with T: differences in brain activation. *Proceedings of the Royal Society of London, Series B*, *263*, 989-995.

Norris, D. (1993). Bottom-up connectionist models of 'interaction'. In. G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 211–234). Hove, UK: Lawrence Erlbaum Associates.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585-589.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. A. Raichle, M. E. (1989). Positron Emission Tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, *1*, 153-170.

Plaut, D., & Shallice, T. (1993a). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377–500.

Plaut, D., & Shallice, T. (1993b). Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist approach. *Journal of Cognitive Neuroscience*, *5*, 89–112.

Poeck, K. (1984). Neuropsychological demonstration of splenial interhemispheric disconnection in a case of optic anomia. *Neuropsychologia*, *22*, 707–713.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. K., Pardo, J. V., Fox, P. T. Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, *4*, 8-26.

Ratcliff, G., & Newcombe, F. (1982). Object recognition: Some deductions from the clinical evidence. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions*. New York: Academic Press.

Riddoch, M. J., & Humphreys, G. W. (1987a). A case of integrative visual agnosia. *Brain*, *110*, 1431–1462.

Riddoch, M. J., & Humphreys, G. W. (1987b). Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology*, *4*, 131–185.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: Foundations* (pp. 318–362). Cambridge, MA: MIT Press/Bradford Books.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press/Bradford Books.

Schnider, A., Benson, D. F., and Scharre, D. W. (1994). Visual agnosia and optic aphasia: Are they anatomically distinct? *Cortex*, *30*, 445–457.

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. E. Patterson, & J. C. Marshall (Eds.), *Deep Dyslexia*. London: Routledge and Kegan Paul.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, *1*, 145–168.

Shallice, T. (1988). *From neuropsychology to mental structure.* New York: Cambridge University Press.

Spreen, O., Benton, A. L., & Van Allen, M. W. (1966). Dissociation of visual and tactile naming in amnesic aphasia. *Neurology*, *16*, 807–814.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis of findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195–231.

Umilta, C. (1995). Domain-specific forms of neglect. *Journal of Clinical & Experimental Neuropsychology*, *17*, 209–219.

Vandenberghe, R., Price, C., Wise, R., Josephs, O. Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, *383*, 254-256.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E. Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, *114*, 1803-1817.

Young, A. W., de Haan, E. H., Newcombe, F., & Hay, D. C. (1990). Facial neglect. *Neuropsychologia*, *28*, 391–415.

Zaidel, E. (1985). Language in the right hemisphere. In D. F. Benson & E. Zaidel (Eds.), *The Dual Brain* (pp. 205–231). New York: The Guilford Press.

Zemel, R. S., & Mozer, M. C. (2000). Localist attractor networks. To appear in S. A. Solla, T. K. Leen, & K.-R. Mueller (Eds.), *Advances in Neural Information Processing Systems 12*. Cambridge, MA: MIT Press.

Zola, S. M. (1997). Amnesia: Neuroanatomic and clinical aspects. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral Neurology and Neuropsychology*. New York: McGraw Hill.