

Modelling direct perceptual constraints on action selection: The Naming and Action Model (NAM)

Eun Young Yoon, Dietmar Heinke, and Glyn W. Humphreys

*Behavioural and Brain Sciences, School of Psychology,
University of Birmingham, UK*

There is increasing experimental and neuropsychological evidence that action selection is directly constrained by perceptual information from objects as well as by more abstract semantic knowledge. To capture this evidence, we develop a new connectionist model of action and name selection from objects—NAM (Naming and Action Model), based on the idea that action selection is determined by convergent input from both visual structural descriptions and abstract semantic knowledge. We show that NAM is able to simulate evidence for a direct route to action selection from both normal subjects (Experiments 1 and 2) and neuropsychological patients (Experiments 3–6). The model provides a useful framework for understanding how perceptual knowledge influences action selection.

SEMANTICALLY MEDIATED ACTION SELECTION

What are the mental operations involved when we select an action to an object, when we make a cutting action with a knife? Traditional cognitive models have emphasized that this process involves the retrieval of semantic knowledge about the object, with the semantic knowledge then being used to guide retrieval of the action (e.g., Roy & Square, 1985, for one example). For instance, for the knife, this may involve accessing knowledge that it is a utensil frequently found in the kitchen, that it is used in the preparation and eating of food, that it is employed along with a fork, and so forth. This semantic information, based on contextual and associative knowledge, is then used to “look up” a

Please address all correspondence to E.Y. Yoon, Behavioural and Brain Sciences, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK.
Email: eyy844@bham.ac.uk

This work was supported by grants from the Medical Research Council and the Wellcome Trust.

learned action that would normally be made with that utensil. Visual information may be used directly to “parameterize” the specific motor response (e.g., so that the grip aperture is scaled correctly; see Marr, 1982; Milner & Goodale, 1995), but the selection of “cutting”, as opposed to any other class of action, is semantically mediated. Furthermore, many models assume that the same semantic knowledge is accessed irrespective of the modality in which the object is presented—whether presentation involves sight of the actual object, hearing its sound, reading its name, or touch (see Caramazza, Hillis, Rapp, & Romani, 1990; Humphreys, Lamote, & Lloyd-Jones, 1995, for examples). Hence, according to such models, actions will be selected in the same manner from visual, verbal, and tactile input, with modality differences emerging only in how easily semantic information itself is accessed in the first place—for example, there may be “privileged” access to semantics from vision since the physical properties of objects may covary with their semantic category (see Hillis & Caramazza, 1991). For such models, selection of an action to an object involves a similar set of processes to selecting a name for speech production, which is also thought to be contingent on access to semantic knowledge (e.g., see Humphreys, Price, & Riddoch, 1999; Levelt, Roelofs, & Meyer, 1999). Action and name retrieval from objects are often thought to contrast with tasks such as word naming, which, in addition to being semantically mediated, may also involve direct associations between input and output representations (e.g., orthographic representations and phonology for either the whole word or for sub-parts; see Coltheart, Curtis, Atkins, & Haller, 1993; Plaut, McClelland, Seidenberg, & Patterson, 1996).

A DIRECT VISUAL ROUTE IN ACTION SELECTION

The proposal that we select actions on the basis of semantic knowledge has had some success in accounting for disturbances of action found in neurologically damaged patients. For example, some patients have what appears to be a central disturbance in their conceptual knowledge of particular objects, so that not only do they use the objects incorrectly but they also can fail to judge when the objects are used correctly (Rothi, Mack, & Heilman, 1986). In such cases, the problem in selection of action can be tied to a deficit in action recognition, consistent with a core semantic disturbance. In other patients, though, the deficit in action seems more related to retrieving the associated motor plan; thus, objects and actions may both be recognized (even named), but actions cannot be effected normally (e.g., in the syndrome of ideomotor apraxia; see DeRenzi & Faglioni, 1999). This pattern can be attributed to impaired access to learned knowledge of action after semantic access is achieved.

However, other evidence is less easy to reconcile with the semantic account. We list next five main findings that suggest that actions are selected not only on

the basis of semantic knowledge (an indirect route, for visually presented objects), but also directly from the visual information present in the environment. We use this evidence to motivate a “dual-route” model of object action and naming (the Action and Naming Model; NAM), which, we show, can simulate pieces of evidence that are difficult for pure semantic accounts. According to NAM, there is a close coupling between the visual properties of objects and actions, which determines not only prehensile movements to objects (reaching and grasping) but also the selection of which class of action to perform with a given stimulus. Perception and action are coupled in the selection as well as the parameterization of action.

Before presenting NAM, we first present the evidence favouring a direct route from vision to the processes mediating action selection. We begin with two sets of experimental findings before discussing three pieces of neuropsychological evidence.

Action slips

Reason (1984) used diary studies to document “action slips” that occur in everyday life, when we perform incorrect actions on objects. These slips can include omissions and perseverations of action (e.g., stirring three rather than two spoonfuls of coffee into a cup), that may not be informative about the kinds of cues that are used in the selection of action. Other slips, however, can include actions that seem to be based on the visual properties of objects—an example being using an air freshener as a hairspray. In such instances, the action seems to be selected according to the shape of the object, without our necessarily taking account of the context where the object is found (e.g., along with cleaning material in a bathroom).

One problem with action slips, documented in diary studies, is that it is difficult to judge how frequently they occur, or even whether their report reflects a selective memory bias on the part of the subject (only certain kinds of errors being remembered). Action errors can be found in laboratory conditions, however, if subjects are required to respond to an unusually fast deadline. Rumati and Humphreys (1998) had subjects make gesture or naming responses to objects and words under deadline conditions. They found that, in both gesturing and naming, some errors could be classified as being visually related to target objects (e.g., making a “hammering” gesture to a razor, presumably because these objects have similar shapes), whereas others were classed as semantically related (e.g., making a gesture to the razor as though it were a shaving brush, with a loose rather than a firm wrist action). When pictures of objects were presented and gesture responses required of subjects, visual errors were proportionately greater in number than errors that were “purely” semantic or judged as both semantically *and* visually related to

targets. In contrast, when naming responses were required, visual errors were less frequent than semantic and semantic/visual errors. Rumiati and Humphreys proposed that this cross-over result was consistent with actions tending to be selected directly from the visual properties of objects without mediation by semantic knowledge. Object naming, on the other hand, tended to be driven by semantic knowledge, leading to relatively higher proportions of semantic and semantic/visual errors (see also Vitkovitch & Humphreys, 1991; Vitkovitch, Humphreys, & Lloyd-Jones, 1993, for similar evidence from naming to deadline).

Rumiati and Humphreys also had subjects make gestures-to-deadline to written words corresponding to objects. In this case they found that visual gesture errors were minimal, although semantic and semantic/visual errors still occurred. This indicates that visual gesture errors are not generated on the basis of the similarity of the actions for different objects, but only when prompted by the visual properties of objects. With words, semantic and semantic/visual errors predominate because actions are then semantically mediated.

Rumiati and Humphreys' study provides the first documentation of the relative proportion of visual to other types of action errors made by normal subjects, and it suggests that actions are selected directly from visual representations.

Response priming

Tucker and Ellis (1998, 2001) had subjects make left- or right-hand responses according to whether objects were depicted as upright or inverted in pictures. They found that responses were affected by the position of the handle of the objects with respect to the hand used for the response, even though the horizontal orientation of the object was irrelevant for the task. Thus a right hand response (e.g., for "upright") was faster if the object's handle was oriented to the right, and vice versa for a left hand response (e.g., for "inverted"). They propose that visuo-motor relations between objects and actions are coded automatically, even when these are not necessary for the task.

Similar effects of compatibility between the orientation of the object and the hand used for response has been reported by Craighero, Fadiga, Rizzolatti, and Umiltà (1998, 1999). They had subjects make a simple reaction time (RT) response by reaching and grasping an oriented bar stimulus. The initiation of the response was cued by a picture of a bar in the same or different orientation to the response bar. Although the orientation of the initiation cue was irrelevant, Craighero et al. found that RTs were faster when the cue and the response bar had the same orientation than when they had different orientations. They suggest that there was automatic activation of the oriented grasp action by the orientation of the cue.

Optic aphasia

Neuropsychological evidence for direct linkages between visual information and action comes from the syndrome of optic aphasia. The term optic aphasia is applied to patients who seem to have a problem in naming objects that is specific to the visual modality. Such patients will fail to name an object when shown to them visually (e.g., a hammer), but will name it when given a verbal definition (what do you use to knock a nail into wood?), indicating that there is not a general problem in naming. However, and in contrast to their impaired naming from vision, the patients will often make appropriate gestures to the objects (e.g., moving a hand up and down in a gesture of “striking”, when shown a hammer; e.g., Lhermitte & Beauvois, 1973; see Riddoch, 1999, for a recent summary). The presence of good gesturing has often been taken to indicate that the patients can retrieve semantic information from objects, but then fail to find the associated name (e.g., Beauvois, 1982). Against this, patients have been shown to be impaired when required to make semantic matches to visually presented objects (see Hillis & Caramazza, 1995; Riddoch & Humphreys, 1987; Sirigu, Duhamel, & Poncet, 1991). For example, patient JB documented by Riddoch and Humphreys (1987) was asked to judge which two of the following three objects were used together: “hammer, chisel, screw”. From vision alone he made around 30% errors. When presented with the names for the objects, though, he performed at ceiling. JB’s good performance when given the names of the objects indicates that he had reasonably intact semantic knowledge, but that he had difficulty in accessing this from vision. From this it follows that the gestures he made were not necessarily based on semantic knowledge but rather were accessed directly from vision. Interestingly some of the gestures made by JB were highly specific. For instance, he gestured with his left hand to a fork and his right to a knife. The specificity of these gestures suggests that they were contingent on access to stored knowledge about the objects. Riddoch and Humphreys thus proposed that there were direct links between stored knowledge of the structural properties of objects (represented in a stored “structural description system”) and learned actions. This direct route by-passed semantic (associative, contextual) knowledge about objects, and so could be used in optic aphasia even when visual access to semantic knowledge was impaired.

A similar dissociation, involving the relative preservation of action compared with impaired object recognition, has also been reported in patients with semantic dementia. Hodges, Spatt, and Patterson (1999), Lauro-Grotto, Piccini, and Shallice (1997), and Riddoch, Humphreys, Heslop, and Castermans (in press), for instance, have documented cases where the patients showed poor matching based on inter-object associations. The patients reported by Lauro-Grotto et al. and Riddoch et al. remained able to perform well-learned everyday tasks. The patient of Hodges et al. made accurate

functional judgements about how objects could be used. Such deficits can be accounted for if a direct route to action is left undisturbed when semantic knowledge deteriorates.

Modality-specific apraxias

The term apraxia is used to describe disorders of motor performance that are not contingent on impaired intellect, object recognition, or motor responses (e.g., hemiplegia; see DeRenzi & Faglioni, 1999). As we have noted earlier, at least some apraxic disorders seem to reflect a central problem in retrieving stored knowledge about actions, whereas others seem contingent on impaired production, though comprehension of actions is intact. More relevant to our present purposes, some patients also manifest modality-specific deficits in action. Most frequently, patients may be impaired at making gestures to verbal input (“show me how to use a hammer”) but are rather better at gesturing to the presence of a visually presented object (when the hammer is actually present). In other instances, however, an opposite pattern is apparent. Patients are worse at gesturing to visually presented objects than they are at gesturing to verbal input (e.g., DeRenzi, Faglioni, & Sorgato, 1982; Pilgrim & Humphreys, 1991; Riddoch, Humphreys, & Price, 1989). This last pattern, which may be labelled “visual apraxia”, seems to comprise the opposite form of impairment to optic aphasia. In optic aphasia, patients can make actions to but are impaired at naming visually presented objects. In visual apraxia, patients are impaired at making actions to visually presented objects, but this can be in the presence of preserved naming and comprehension of the objects concerned (see Riddoch et al., 1989). To account for such a disorder, Riddoch et al. proposed that there was damage to a direct route to action from vision, which was sufficient to block actions to visually presented objects despite a semantic route to action being intact (there is good gesturing to object names), and despite good visual access to semantic knowledge (e.g., object naming being relatively preserved). In patients showing impaired gesturing to verbal input, and an improvement with visually presented objects, it can be argued that preservation of a visual route to action supports performance even when the semantic route is damaged. The double dissociations, between visual apraxia and optic aphasia and between the different modality-specific apraxias, are difficult to account for in terms of a single semantic route to action.

Utilization behaviour

Following damage to the frontal lobes, patients can manifest so-called utilization behaviour, in which they appear to act in a very environmentally driven manner. For example, Lhermitte (1983) first described cases in which patients placed successive pairs of spectacles on their faces, even when others were still present, if given the objects to hold. There is recent evidence that such

behaviours are strongly cued by the visual properties of the objects present. Riddoch and colleagues (e.g., Riddoch, Edwards, Humphreys, West, & Heafield, 1998; Riddoch, Humphreys, & Edwards, 2000) had patients carry out a novel task in the face of conflicting visual cues to make over-learned responses. The stimuli were cups placed either on the left or right side of the patient's body, and the task was to pick up each cup using the hand on the side of space matching the position of the cup (left hand to left-side cup, right to right-side cup). The handles of the cups could face left or right, but this was irrelevant to the task. Although patients were able to explain the rules to the task, Riddoch et al. found that many errors were made by patients responding with the hand congruent with the handle of the cup, even though this could contradict the task rule (e.g., using the right hand to reach to the right-side handle of a cup placed on the left of the patient's body). Interestingly, these utilization-like behaviours were modulated by the visual properties of the objects. For example, the action errors reduced if the cups were inverted. Now, since access to semantic knowledge about objects is relatively indifferent to the left-right orientation of objects (see Biederman & Cooper, 1991), the strong effects of left-right orientation observed by Riddoch et al. indicate direct visual constraints on the activation of the overlearned actions. This is also suggested by the effects of object inversion, which reduces the visual familiarity of stimuli.

Humphreys, Forde, and Francis (2000) demonstrated similar effects in multi-object tasks. They presented a patient with frontal lobe damage with an array of objects and instructed him to use them in relatively novel ways—for example, to place a saucer on a cup. They found that the patient frequently made an over-learned action rather than the action as instructed (e.g., putting the cup on the saucer), even though he could often recall the instruction when asked. Here the over-learned response overrode control from a verbal memory representation. The likelihood of this occurring was much reduced, though, if the objects were replaced by cards (then the patient might place the card saying "saucer" on the card saying "cup"), even though the words could be recognized and should cue a semantic route to action. Again it appears that utilization-like behaviours are most likely to be generated by visually familiar objects, which may directly cue visually associated actions.

THE NAMING AND ACTION MODEL (NAM)

The experimental and neuropsychological evidence cited in the previous section is consistent with the existence of a direct, visual route to action in addition to a route to action through semantic (associative and contextual) knowledge. There appear to be dynamic constraints on action selection from the visual properties of objects, so that visually related actions are activated rapidly and emerge under deadline conditions (Rumiati & Humphreys, 1998). These visually activated actions may subsequently be integrated with semantic

knowledge, so that, usually, an action is selected that is consistent with both the visual and semantic properties of objects. In patients, one or other route may be damaged, giving rise to different modality-specific apraxias, although impaired visual access to semantic knowledge can still leave a visual route to action able to support gestures, in optic aphasia. Utilization behaviours may also be contingent on fast-access to action through the visual route, which may be activated by the presence of objects even when this activation contravenes a representation of the task held in verbal working memory (Humphreys et al., 2000). A framework illustrating a “dual-route” model of action selection, first proposed by Riddoch et al. (1989), is presented in Figure 1.

The “boxes and arrows” framework given in Figure 1 is useful in accounting for the functional dissociations in action and naming found in brain-lesioned patients discussed in the previous section. However, it provides a less satisfactory account of dynamic aspects of action selection. For example, it is not clear how actions activated by the visual properties of objects can be integrated with

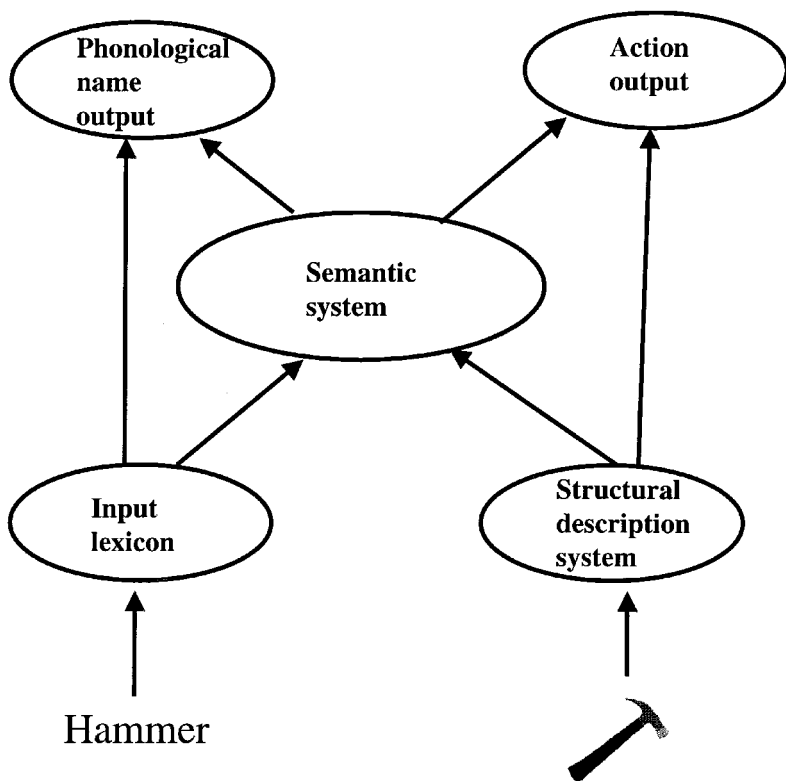


Figure 1. The “box and arrows” dual-route theory of action (taken from Riddoch et al., 1989).

semantic knowledge during the normal action process, and why this process may lead to a “blocking” of the semantically generated response in cases of visual apraxia (see Pilgrim & Humphreys, 1991; Riddoch et al., 1989). It is also unclear how interactions between “acting” and “naming” could take place. For instance, in cases of optic aphasia it has been noted that misnaming an object can sometimes induce action errors (where patients make an action that matches the name they produce; Lhermitte & Beauvois, 1973). However, it can also be noted that, quite often, such patients may only name an object after having retrieved the associated action (Riddoch & Humphreys, 1987). What is needed here is a more explicit account of the dynamic properties of semantic and visual routes to action that can account for the time course and interactions in naming and acting, but that is nevertheless able to capture the patterns of dissociation that occur. This is the aim of the present paper. We present the Action and Naming Model (NAM), which uses a “quasi-modular” connectionist framework to simulate action selection in normality and pathology (after brain lesions). NAM provides a dynamic account of action and name selection from objects and words. We use the model to simulate evidence on action slips in normal observers, and both optic aphasia and modality-specific apraxias (see earlier). Since NAM is applied to just action and name selection from known objects, it cannot currently simulate orientation decision tasks (cf., Tucker & Ellis, 1998) or tasks involving actions based on the locations of objects (cf., Riddoch et al., 1998). Nevertheless, we show that it can capture evidence on a visual route to action that is difficult for a traditional, semantic approach to account for. It can also simulate interactions between acting and naming, consistent with prior neuropsychological evidence. The model provides a framework for understanding how perception directly constrains action to objects.

The architecture of NAM is presented in Figure 2. The mathematical functions determining the operation of the model are given in Appendix A. The model can accept two forms of input, corresponding respectively to written words and objects, and it can generate two forms of output corresponding to actions and phonological names. The model is based around a quasi-modular set of representations for the different inputs and outputs and for three main forms of stored knowledge—a visual lexicon and a structural description system, concerned respectively with the visual properties of words and objects, and semantic knowledge. Semantic knowledge is further segregated according to whether units represent superordinate information (e.g., tool, kitchen utensil, clothing) or item-specific information (hammer, spoon, tie). The input and output representations can be thought to stand for lexica for objects, words, and actions. The action lexicon contains stored representations of classes of action (e.g., twist, pour, drink), which may be specified in relatively abstract terms and even independently of the particular effectors used for the action. The operation of the model is based on an interactive activation and competition network

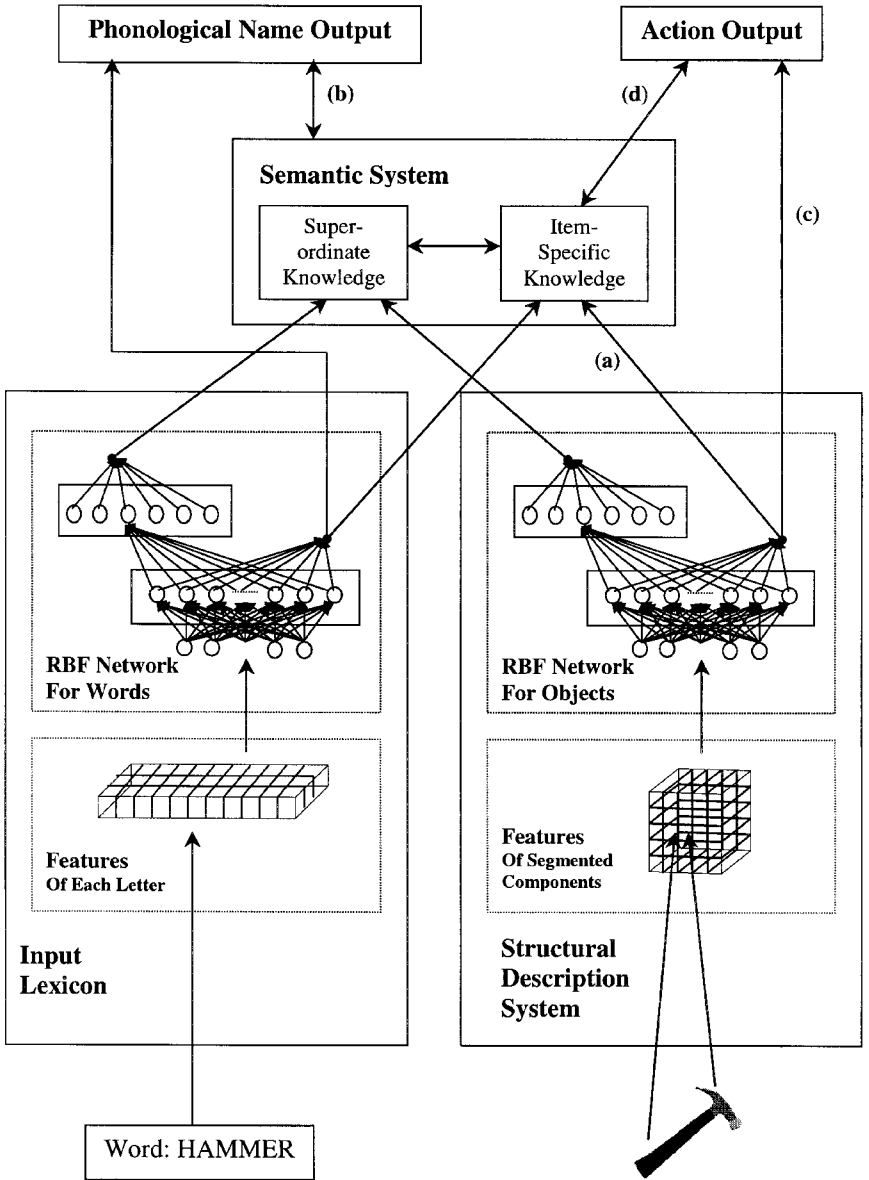


Figure 2. The architecture of NAM. (a) corresponds to the pathway from the structural description system to the item-specific semantic knowledge; (b) corresponds to the pathway from the semantic system to the phonological name output; (c) corresponds to the pathway from the structural description system to the action output; (d) corresponds to the pathway from the item-specific semantic knowledge to the action output. These pathways were lesioned to produce the neuropsychological disorders discussed in the text.

(Humphreys et al., 1995; Rumelhart & McClelland, 1982). Activation is transmitted continuously between units at the different levels, with units being competitive with one another within a given level. The time taken for convergence to be reached for a given response is taken as the RT for the model.

NAM primarily uses local representations, though words and objects produce a distributed pattern of activation across the local units in the input lexicon and in the structural description system, respectively—the source of feature similarity effects in naming and action selection for objects. Local representations are useful for understanding the functional operation of models, and they may fractionate in an appropriate way when lesioned (see Young & Burton, 1999). Furthermore, the interactive activation framework used in NAM gives rise to emergent interactions between separate processing routes, which help to capture some of the behaviours observed in optic aphasia and disorders such as visual apraxia. These emergent interactions extend our understanding beyond that offered by traditional “box and arrow” models, and they lead to effects that cannot be predicted purely on the basis of the architecture of the model.

Input descriptions

For objects, input to the structural description system is based on the activation of input units that capture plausible visual features within each nameable part of an object, with the parts arranged in the appropriate spatial location with respect to one another. There was one feature unit for each of the following features: The number of straight lines in the part, the number of curved lines, the number of vertices, and the length and width of the part. The feature units took a particular value, depending on the number and type of features present (see Table 1 for one example—a hammer). The feature units were also replicated across a two-dimensional visual field of size 5×5 , creating a three-dimensional feature matrix ($5x$ values $\times 5y$ values $\times 5$ feature types). Objects were positioned so that what was judged their main functional component was placed at the centre of the field (location 3,3), and other components activated units in appropriate locations with respect to this main component. Figure 3

TABLE 1
Measurements for the feature vector of the object “hammer”

<i>Features/components</i>	<i>Head of a hammer</i>	<i>Handle of a hammer</i>
Width	16	3
Length	3	9
Number of vertices	12	11
Number of curved lines	3	6
Number of straight lines	7	3

gives an example of a hammer, which was segmented into two parts (the handle and the head), with the head being judged the main functional part (placed at the centre of the field). The representation captured in the matrix represents the spatial relations between the parts of objects implicitly, according to which relative locations are activated. The representation is “centred” on the main component part of the object, as would be the case if the description were activated after attention had been drawn to this main component. One realization of

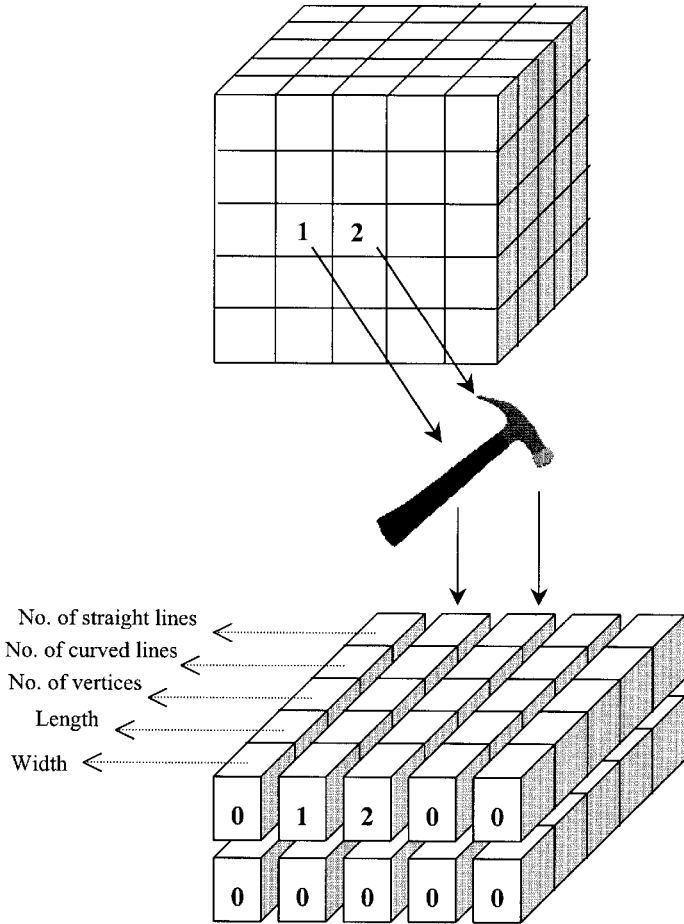


Figure 3. Example of the input coded for objects in NAM. Objects were coded in terms of the values of each part along five feature dimensions: Width, length, number of vertices, number of curved lines, and number of straight lines. Descriptions were centred on the major functional part in each object (which was assigned to the central sets of units in the input matrix).

this is the SAIM model of visual selection (Selective Attention for Identification Model) proposed by Humphreys and Heinke (1998; Heinke & Humphreys, in press), which fixes an attentional window onto a central part of an object. Stored representations are then activated according to the positions of object features with respect to this part. The unit that responds to the object parts being in the appropriate spatial locations corresponds to the structural description of an object.

Features are mapped through to the structural description system by means of a radial basis function (RBF) network. Units in the structural description system are local, and each corresponds to one object (32 units in total). However, the profile of activation in the structural description system, given by the RBF network, is distributed across the objects, and this provides a similarity measure across the object set. Each unit in the structural description system is positively connected to units that are features of the particular object. However, given that features appear in more than one object, several structural descriptions can be activated by any given object input. The activation profile in the structural description system, then, indicates the degree of match between all known objects and the input object. Units in all other layers of the model are also local, each corresponding to one object. This allows for tractable analysis of the model (see Young & Burton, 1999).

At the level of the input lexicon for words, there is also a distributed activation of input across the local units corresponding to each word, generated in a similar way to the structural description for objects (see Table 2 for the example of a hammer). In this case, just three visual features were used to represent the visual features making up each letter. These were: The number of straight lines, the number of curved lines, and the number of vertices. On a more abstract level, the input could also be taken to be the auditory description for words, with the features corresponding to something like the phonemic properties of words. An example is presented in Figure 4. Each feature unit took a particular value according to the features presented, and this was coded in a two-dimensional vector, where the letters within a word were positioned from left to right in a feature matrix. As with the structural description for objects, the vector coded in the input lexicon was mapped through to the semantic layer by means of a radial basis function (RBF) network.

TABLE 2
Measurements for the feature vector of the word "hammer"

<i>Features/components</i>	<i>H</i>	<i>A</i>	<i>M</i>	<i>M</i>	<i>E</i>	<i>R</i>
Number of vertices	2	3	3	3	3	2
Number of curved lines	0	0	0	0	0	1
Number of straight lines	3	3	4	4	4	2

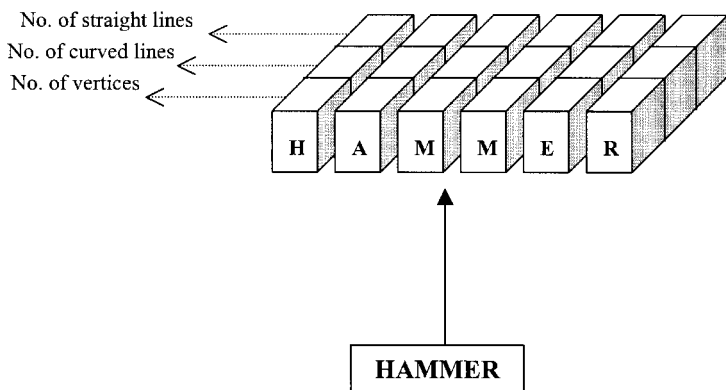


Figure 4. Example of the input coded for words in NAM. Words were coded in terms of the values of each part along three feature dimensions: Number of vertices, number of curved lines, and number of straight lines. Descriptions were centred on the letters within a word, assigned from left to right in the two-dimensional input matrix.

The semantic system and mapping to output

Both words and objects, as inputs, can be mapped through to the semantic system. In addition, each input type has a direct route to particular output units (to phonological names, for words as input, and to actions, for objects as input). The mapping is in all cases one-to-one, from each input system respectively to the semantic system and to the two different output representations (phonological names and actions). However, words can only be mapped to actions through the semantic system, and objects can only be mapped to names through this same system. This provides a dual-route architecture, for both words and objects (for naming in the one case, and for action in the other). Units in the input lexicon (for words) and the structural description system (for objects) are mapped through independently to both superordinate and item-specific semantic representations, and units at the semantic level are only competitive within the respective sub-systems (for superordinate and item-specific knowledge). There were six superordinate units (corresponding to accessories, tools, toys/sports, household objects, food, and office items), and 32 item-specific semantic units. Units at the semantic level draw together activation from stimuli in different modalities (words and objects), and they also capture categorical relations between visually different stimuli (via the superordinate units). At the output level, names can also correspond to superordinate or to item-specific representations, but these were all competitive with one another since only one name should be selected for a given input (38 phonological name units). A control unit was used to modulate the transmission of activation from the semantic to the name output units, so that naming could be biased either to

superordinates or to item-specific names (see Cohen, Dunbar, & McClelland, 1990, for an example of a similar approach in modelling the Stroop effect). Action units were item-specific (32 units), since a common action cannot be made to all objects belonging within a common category. The parameters for the model are provided in Appendix C, and the full list of objects in Appendix D.

The dual semantic and non-semantic routes from orthography to phonology are motivated by both normal experimental and neuropsychological evidence—including findings of phonological interference on semantic classification decisions to words (e.g., see Van Orden, 1987) and double dissociations between surface dyslexia, on the one hand, and phonological and deep dyslexia on the other (e.g., see Patterson & Morton, 1985). We have discussed the evidence for dual semantic and non-semantic routes to action from objects, and NAM stands as an existence proof of whether such an approach can accommodate the data on action and naming responses to objects and words.

EXPERIMENT 1

Basic times for action, naming, and categorization

Before going on to simulate evidence that favours the idea of a direct route from vision to action selection, we sought to apply NAM to basic data on naming, categorization, and action from words and objects. There is a substantial literature demonstrating that words can be named faster than objects, whereas, in contrast, objects seem to gain faster access to semantic knowledge, as measured in categorization times (e.g., see Glaser, 1994; Potter & Faulconer, 1975). Few studies have measured naming and semantic categorization times in relation to the time to access action knowledge. However, Chainay and Humphreys (in press) compared naming, superordinate classification, and action decisions to words and pictures of objects (action decisions required subjects to decide whether a given object would be used to make a “twisting” or a “pouring” action). They too found that words were named faster than objects. There was some advantage for objects over words when stimuli had to be assigned to categories for “use indoors” and “use outdoors”. The advantage for objects over words was increased further, though, for action decisions. These fast decisions for actions are again consistent with either a direct route to action for objects or with a semantic system organized so that visual information more heavily connects to action than does word input. Indeed there was little advantage for action decisions over superordinate classification for words, so the fast action decisions for objects cannot simply be because these decisions can be based on less semantic knowledge. NAM needs to be able to capture this qualitative pattern of effects to provide a plausible account of object and word processing.

Method

There were 32 objects used in the simulation (see Appendix D for a full list). The items corresponded to common everyday objects, each of which can be associated with a specific manual response (hammer, fork, pen), and each of which could be coded in terms of the set of visual features allocated to the model. There were thus 32 input units in each of the structural description and input word representations, 32 units for item-specific semantic knowledge, 32 item-specific name units, and 32 action units. The 32 objects divided into 6 categories, with different numbers of members. There were six superordinate units at the semantic level, and the same number of superordinate name units. The threshold for activation in the (name and action) output units was set to 0.9. Naming and action decision times were based on the number of cycles for units to reach threshold at the phonological name and action output layers. Figures 5 and 6 illustrate the mean activation at each network cycle over the 32 visual objects and words respectively, specifying the threshold and the mean reaction times (RTs) for naming and action decisions for objects and words. The times

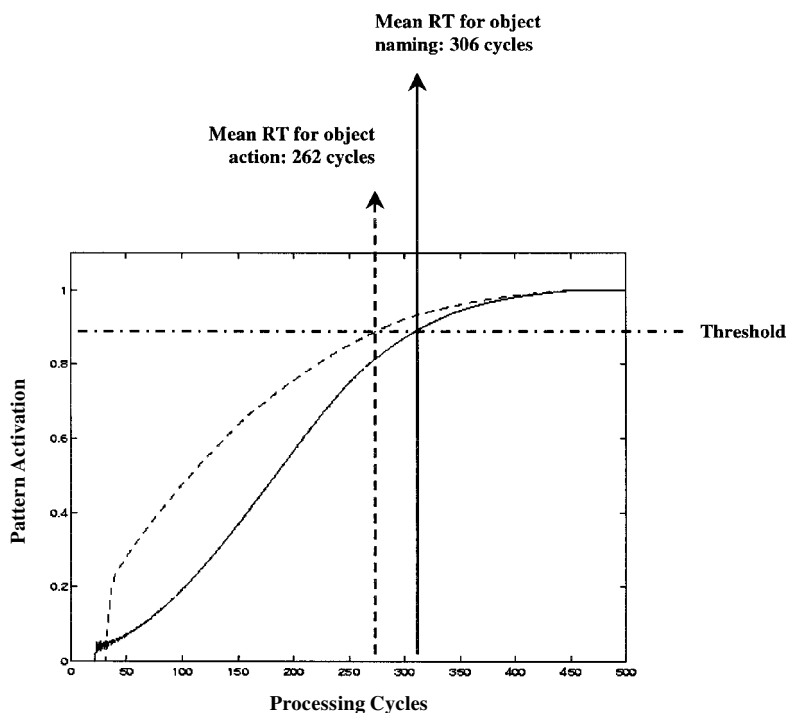


Figure 5. The mean activations across 32 visual objects for naming and action as a function of network cycles: The dotted line is for action and the solid for naming.

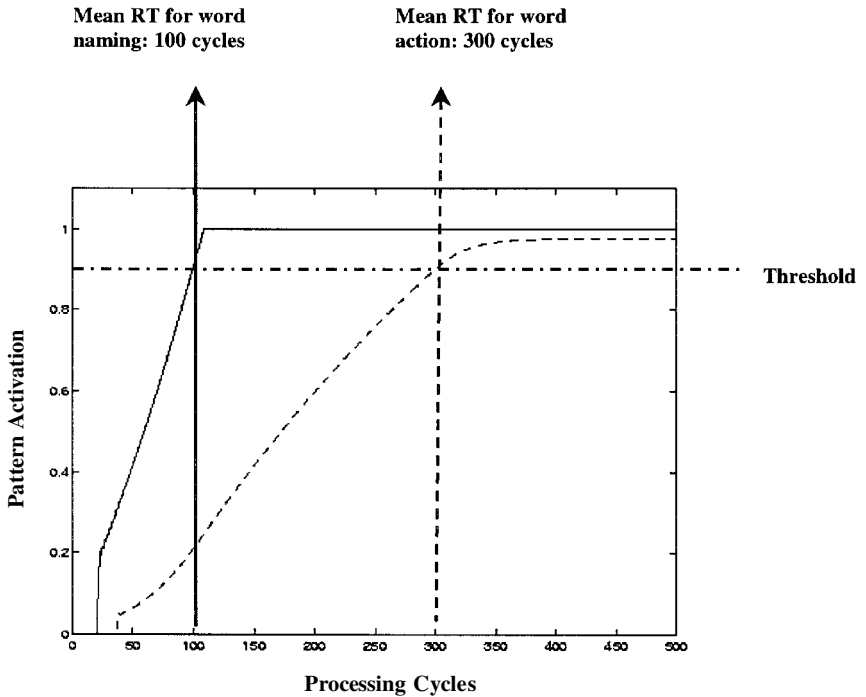


Figure 6. The mean activations across 32 words for naming and action as a function of network cycles: The dotted line is for action and the solid line for naming.

for superordinate classification were derived from the units corresponding to the superordinate phonological names, whereas naming times for individual objects were based on the item-specific units at the phonological name level. Word inputs were always for specific items (hammer, spoon) rather than for superordinate terms (tool, kitchen utensil), to prevent fast superordinate classification responses being based on the direct route to phonological names for words.

Results and discussion

The mean RTs for naming, superordinate classification, and action decisions for NAM are given in Figure 7(a). Figure 7(b) presents data from human subjects collected by Chainay and Humphreys (in press). Error proportions are given in Figures 7(c) (NAM) and 7(d) (humans). For NAM the error proportions are based on the residual, summed activity in output units for non-targets (those units not corresponding to the item presented on a trial), when the

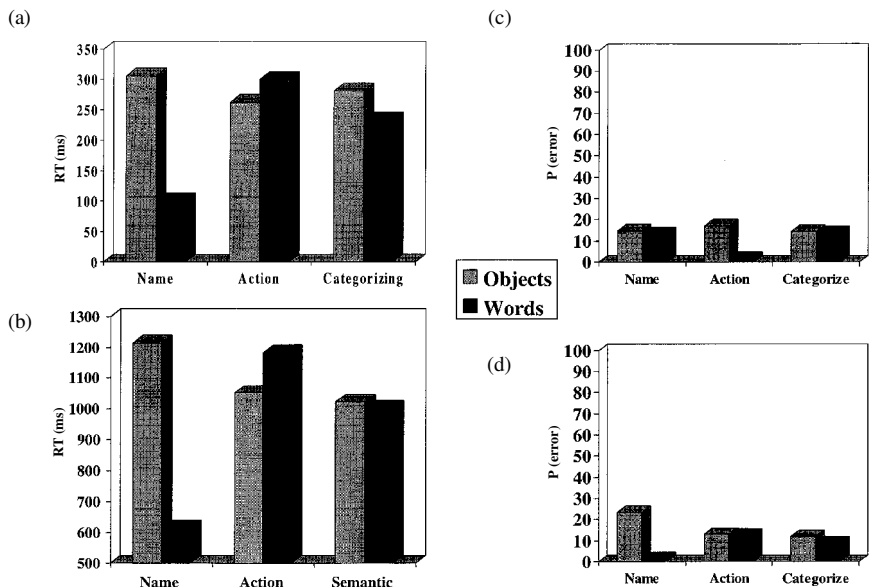


Figure 7. (a) Simulation of RTs (in network iterations) for naming, action selection, and categorization tasks. (b) Data from human participants performing naming, action decision, and categorization tasks (from Chainay & Humphreys, in press). (c) Residual error probabilities in NAM. (d) Error probabilities in human participants.

target's threshold is passed. This provides an estimate of the likelihood of response errors occurring in the model.¹

The RT data from NAM were analysed in a repeated measures ANOVA, with two factors: Input type (object, word) and response (individual name, superordinate classification and action decision). Objects were treated as a random factor. This showed significant effects of input type, $F(1, 31) = 38.78$, $p < .001$ and response, $F(2, 30) = 96.48$, $p < .001$. The interaction was highly reliable, $F(2, 30) = 229.86$, $p < .001$.

Naming times to words were considerably faster than naming times to objects, consistent with the literature on word and object naming (e.g., Glaser,

¹This would be the case if the model operated stochastically rather than in a deterministic fashion. In its normal state, the model was deterministic. However, with some degree of internal noise, activation in non-target units that is on average less than that in target units, would nevertheless sometimes give rise to errors if set higher than the target unit (by noise) on a given trial. The probability that this occurs is predicted by the average activation levels in the units, even when the model is run in deterministic mode.

1994; Potter & Faulconer, 1975). In NAM this word naming advantage comes about because words can be mapped directly through to individual phonological names, whereas objects can only be mapped through to names via sets of intermediary semantic units (i.e., words alone have a direct route to phonology). For words, actions are accessed slowly, now because activation must be transmitted through the semantic system. This means not only that a further set of connections must be passed through (when compared with the direct naming route), but also that there is added competition in selection. In particular, activation of superordinate units at the semantic level generates activation in a number of action units (for all objects within a given category), creating competition for action selection. Indeed, actions for words are selected somewhat more slowly than superordinate categorization responses, even though superordinate responses to word inputs must also be recovered through the semantic system, $t(31) = 85.38$, $p < .001$. For superordinate responses, however, activation from both superordinate and item-specific semantic units converges on a common response, ensuring little competition in superordinate name selection. In contrast, as we have noted, there is some competition for action selection.

For objects, actions are accessed faster than both superordinate and item-specific names, with the differences being greatest relative to item-specific names, $t(31) = 8.61$, $p < .001$; the advantage for action selection over superordinate classification was not reliable, $t(31) = -1.3$, $p > .05$. Actions are also selected faster from objects than from words, $t(31) = -2.49$, $p < .025$. The advantage in action selection for objects over words (due to the direct route from objects to action) is somewhat less than the advantage in naming for words over objects (due to the direct route from spelling to sound). This asymmetry occurs because action selection from objects requires more time in order to overcome competition created by the distributed representation of objects at the structural description level, relative to the competition created when input words are mapped through to their names. Interestingly, RTs for action selection and superordinate classification also go in opposite directions for objects and words. Objects are more advantaged over words for action than for superordinate classification, a result that matches our own psychological data (Chainay & Humphreys, in press). Indeed, in our simulations, words were categorized faster on average than objects, $t(31) = 3.48$, $p < .01$. For NAM, the differences in superordinate classification time reflect the contrast in the distributed representations between words and objects, and the fact that the categories here were comprised of items with heterogeneous perceptual structures. This heterogeneity, at a category level, means that overlap in structural descriptions is somewhat disruptive. In our own study examining human superordinate classification, when subjects were asked to decide whether or not a stimulus was a kitchen utensil, we have also found that RTs were faster to words as input (Chainay & Humphreys, in press; Figure 7b). It is possible that

the shared structural properties of objects cue several categories, slowing classification at a relatively fine-grained level. Note that studies that have shown a categorization advantage for objects over words have tended to high-level classifications, such as living vs non-living, when correlated visual features can be used to facilitate object classification (e.g., Job, Rumiati, & Lotto, 1992; Potter & Faulconer, 1975).

Overall, then, NAM can generate data that approximate existing findings on human naming, action decision, and classification, albeit with highly simplified input.

EXPERIMENT 2

Naming and acting to deadlines

When human subjects have to make appropriate actions, or name objects, to unusually fast deadlines, they generate different error profiles for each task (Rumiati & Humphreys, 1998). For actions, proportionately higher numbers of visual errors are made than semantic and semantic/visual errors; for naming, this pattern is reversed. The relatively high numbers of visual errors in action are difficult to explain if actions are derived solely through a semantic system used for both naming and action. We attempted to simulate this result in NAM, by setting response deadlines that were faster than the thresholds normally used for convergence. We contrasted action errors to deadline with objects and words as input, since the proportion of visual errors is higher for objects, and, for words (unlike for objects), more semantic plus semantic/visual errors than visual errors are made to words (Rumiati & Humphreys, 1998). This differential error pattern fits with a direct route to action for objects but not words.

Method

For the simulations of naming and action slips, a response deadline of 100 network cycles was chosen; under these conditions the residual error variance in non-target units increased considerably (see later). The naming control unit was set so that item-specific naming responses were favoured. In all other respects, the method followed that in Experiment 1.

For each object taken as a target, the other objects in the set were classified as being visually related, semantically related, both semantically and visually related or not related, using an error classification scheme used to label errors generated in neuropsychological studies of object naming (Hodges, Salmon, & Butters, 1991). Visual errors were defined as being responses corresponding to objects that shared visual properties with the target but they belonged to a different category and they were not associated with the target. Semantic errors were responses corresponding to objects that do not share visual properties with targets, but which do belong to the same category and/or are

associated with targets. Semantic/visual errors were responses corresponding to objects that are both semantically and visually related to targets; unrelated errors were for objects that were neither visually nor semantically related to targets. For a target “knife”, a visual error would be “pen”, a semantic error would be “jug”, a semantic/visual error would be “fork”, and an unrelated error would be “spinning top”. Two independent judges labelled each of the possible item-specific outputs as being a visual, semantic, semantic/visual, or unrelated error, for each target. Then, for each target input, the activation in outputs units corresponding to visual, semantic, semantic/visual, and unrelated errors was summed, to provide a measure of the probability that each error type could arise under the deadline condition. Judge 1 classified 176 comparisons of target input and non-target outputs as visually related, 140 as semantic, and 36 as semantic/visual. The equivalent scores for judge 2 were 172, 138, and 38 (out of 32 input \times 31 output comparisons). The same classifications were made for 89% of the comparisons by the two judges (880 out of 992 combinations). The data were analysed taking “judge” as a factor (to assess effects of the small differences in error classification across judges).

Objects and words were presented to the model. Outputs were derived from item-specific representations for names, for actions for objects, and for actions for words.

Results and discussion

The probabilities of visual and semantic (semantic + visual) errors are presented in Figure 8.

The data were analysed by taking the activation levels across the phonological name and action units for objects, and summing those for non-target units that were assigned to common error types by each judge (specifically for visual, semantic, and semantic/visual errors). These activation levels were subjected to a three-way analysis of variance with the within-item factors being task (naming objects, actions to objects, and actions to words), error type (visual vs summed semantic and semantic/visual errors²), and judge (activation assigned according to the error classifications performed by each judge). This analysis revealed a reliable main effect of task, $F(2, 30) = 14.05, p < .001$, and a Task \times Error interaction, $F(2, 30) = 8.66, p < .01$. There was no effect of judge, and no interactions with this factor. Accordingly, the data shown in Figure 8 are averaged across the two judges.

²In analysing their data, Rumiati and Humphreys (1998) summed semantic and semantic/visual errors as a conservative way of testing whether “pure” visual errors were more frequent than all possible semantic errors (noting that semantic/visual errors could be “pure” visual errors too).

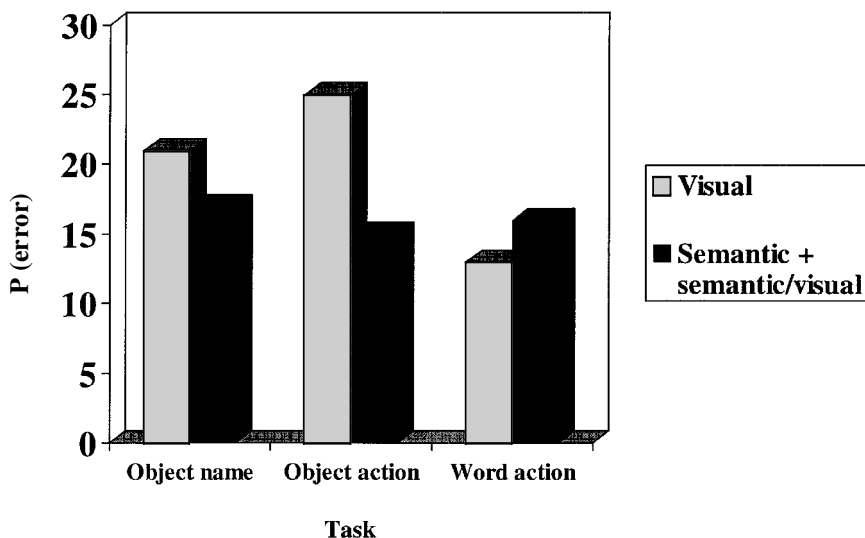


Figure 8. Simulation of responses to a fast deadline in NAM. Residual error probabilities are separated for “visual” and “semantic” (semantic + semantic/visual) error types.

The data are broadly consistent with the patterns of errors reported by Rumiati and Humphreys (1998). For action, there was considerably higher chance of making a visual action error than an action error that was semantically related to the target (the combined semantic and semantic/visual error rate). This increase in visual relative to semantic errors was less marked for naming responses (hence the interaction between task and error type). Unlike the human data, visual errors remained more prevalent, however, even in naming. This may in part reflect the constraints of the limited set of items we used, which curtail the errors across broad sets of category members (since each category was limited in size). More important, is the qualitative shift in the probabilities of the two types of error, when action and naming were compared. In addition, like Rumiati and Humphreys, we find a pattern in which more semantic plus semantic/visual errors than visual errors arose for action selection to words. Thus the high prevalence of visual errors in the same task with objects is not simply caused by interactions in the action selection system. Rather the contrast is due to the different ways in which actions are selected from objects when compared with both name selection from objects and action selection from words, with a direct visual route contributing to the proportionately high numbers of visual errors found when actions are selected to objects. For NAM, the distributed representation, generated across the structural description units, exerts a strong influence on action selection, and

the spread of activation across actions to objects that have similar structures leads to the occurrence of visual errors in action. For object naming, the distributed activity at the structural description level is cleaned up to some degree by the local semantic units, prior to high levels of activation being passed on to the phonological name system. This reduces the relative frequency of visual errors. Likewise, for action selection from words, the occurrence of visual errors in action diminishes as the distributed activity at the input lexicon is "filtered" through the semantic system prior to being passed to the action output system.

One disparity with the human data is that Rumati and Humphreys (1998) reported essentially no visual errors when subjects made gestures to deadline with words. There are some visual errors found in NAM, even with words, because the convergence procedure allows some small levels of activation to remain in units of all nontarget items; this generates a residual probability for each error type irrespective of the stimulus presented. Over and above this, though, NAM produces stimulus-specific changes in error probabilities that mimic those found when humans respond to fast deadlines.

EXPERIMENT 3

Optic aphasia

In the subsequent experiments with NAM, we evaluated whether the model could simulate patterns of performance in naming and action selection found in neuropsychological patients. Experiment 3 tested whether NAM could capture important aspects of the neuropsychological syndrome of optic aphasia. As we noted in the introductory sections, optic aphasic patients can make relatively precise actions to objects despite showing impaired object naming (e.g., Lhermitte & Beauvois, 1973). In some cases where it has been tested, visual access to detailed semantic knowledge about objects has also been shown to be impaired (e.g., Hillis & Caramazza, 1995; Riddoch & Humphreys, 1987), though the patients may still access general (superordinate) knowledge. We tested whether these deficits associated arose when the model suffered damage to the route mapping visual input descriptions of objects into the semantic system (and in particular, when damage affected mapping into item-specific semantic knowledge). In some cases of optic aphasia, the naming responses of patients seems contingent on their generating an appropriate action, though on other occasions the generation of an incorrect name can elicit an erroneous gesture (based on the name rather than the object; see Lhermitte & Beauvois, 1973). This suggests that the performance of these patients can be influenced by complex, dynamic interactions between the processes involved in name and action retrieval. We were interested in whether instances of these dynamic interactions would be apparent in NAM's performance.

Method

Lesions were simulated by adding randomly distributed noise to the pathway from the structural description system to item-specific semantic knowledge and by reducing the magnitude of the input values being transmitted at this level to 60% of the values used in the unlesioned model (see Appendix B for details). The patients studied by Hillis and Caramazza (1995) and Riddoch and Humphreys (1987) were both able to make accurate discrimination judgements between real objects and nonsense objects constructed by combining the parts of different objects. These authors suggested that the deficits in the patients occurred after intact access to stored knowledge of object form had taken place (enabling the patients to discriminate familiar, real from unfamiliar, novel objects). Hence our lesioning operated post access to the structural description units. Noise values varied between 0 and 1.5, and these values were applied to the weights shown in Figure 2, location (a). Thirty simulations were conducted, with the noise values varying across simulations but with it being constant across objects, for one given lesion example. Previous attempts to simulate neuropsychological data in connectionist models have demonstrated variable results, according to the precise lesion conducted (see Hinton & Shallice, 1991, for one example). By carrying out 30 simulations, we ensured that any results were generally true for lesions applied to this location in the model, irrespective of the particular values of noise added on one occasion. The different lesion replications can be considered to correspond to tests with different patients with damage to the same neural locus. Performance was assessed using four tasks: Object naming, object action, object categorization, and word action. Word naming and categorization were not included, as these tasks have not been critical to the neuropsychological studies.

Results and discussion

In Figures 9 and 10 we present the mean RTs and error rates for name and action selection, and for selection of the appropriate superordinate category name, with objects as input, and also the RTs and error rates for action selection to words.³ RTs overall were slowed by lesioning, relative to the unlesioned version of the model (Figure 9). This slowing was most pronounced for object naming relative to action selection from objects. For example there was an interaction between model version (intact vs lesioned) and task (object naming, object action, object categorization, word action), in a comparison of the RTs in Experiment 1 and the RTs for each object averaged across the 30 exemplars

³Word naming was preserved under all of the present lesion conditions, since all lesions left intact the direct route from word inputs to phonological name outputs.

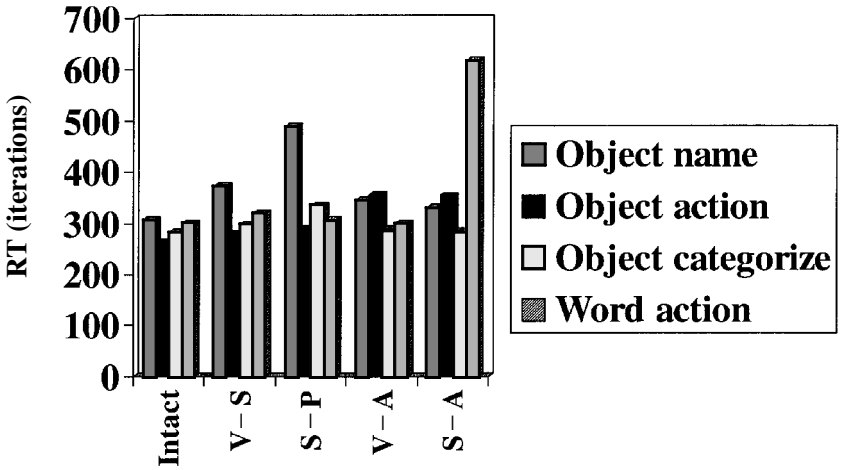


Figure 9. Simulation of RTs for object naming, object action decisions, object categorization, and word action decisions following different lesions to NAM. V-S corresponds to a lesion affecting the mapping from the structural description system to the semantic system. S-P corresponds to a lesion affecting mapping from the semantic system to the phonological name system. V-A corresponds to a lesion to the connections from the structural description system to the action lexicon. S-A corresponds to a lesion from the semantic system to the action lexicon.

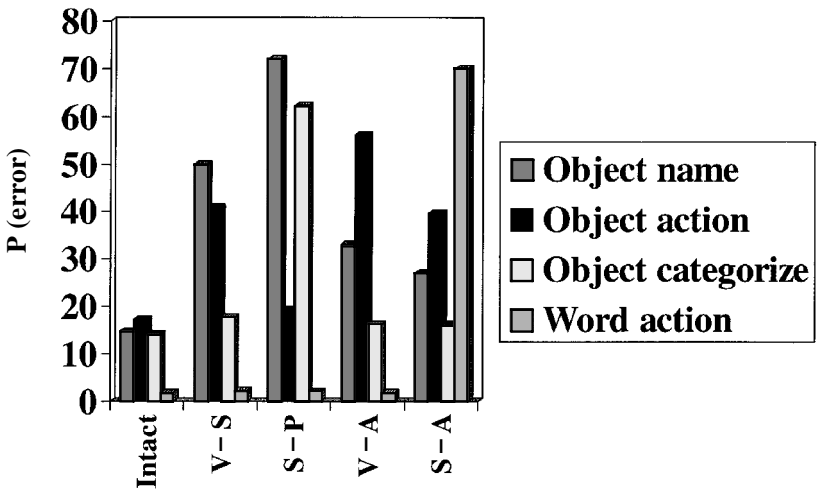


Figure 10. Simulation of the residual error probabilities for object naming, object action decision, object categorization, and word action decisions after lesioning. V-S corresponds to a lesion affecting the mapping from the structural description system to the semantic system. S-P corresponds to a lesion affecting mapping from the semantic system to the phonological name system. V-A corresponds to a lesion to the connections from the structural description system to the action lexicon. S-A corresponds to a lesion from the semantic system to the action lexicon.

with varying degrees of lesioning in Experiment 3, $F(3, 93) = 22.35, p < .001$, for the interaction between task and lesioning. Errors generally followed a similar pattern. The error rate was raised for object naming when compared with the other tasks, $F(3, 93) = 125.80, p < .001$, for the interaction of task and lesioning. Errors for superordinate classification were based on the activation levels in non-target category names. These remained low, despite the increase in errors for selecting the appropriate item-specific name. To relate to optic aphasia, the comparison between name selection and action selection to objects is most critical. RTs for object naming were slower than for action selection to objects, $t(31) = 10.56, p < .001$, and the error rate was raised for name selection, $t(31) = 12.60, p < .001$.

It is also of some interest that, relative to the intact model, RTs and errors were increased not only for object naming but also for action selection to objects. In the other tasks, though, RTs and errors were not affected (for the lesioned vs the unlesioned version of the model). This affect on actions to objects arises in NAM because activation from the semantic as well as the direct visual route converges on the stage of action selection. Disruption to semantic access from objects impairs the contribution of the semantic route to action. However, this in turn has a linked disruptive effect on visually mediated selection of action (via the direct visual route). The speed of action selection in optic aphasia has not been examined, so it is difficult to judge whether this account fits with the neuropsychological data. What is the case is that the accuracy to gesture from vision is relatively preserved when compared with object naming, as simulated here on the accuracy data (Figure 10). Nevertheless, gestures in such patients are often not perfect (Hillis & Caramazza, 1995; Riddoch & Humphreys, 1987). Partial preservation is predicted by NAM, with the magnitude of any deficits dependent on the precision of the required gestures. Note that, in the neuropsychological studies, less precise responses can be required for gesturing than for naming (e.g., axe and hammer would be given the same gesture but not the same name; see Funnell, 1987), and so gesturing may be scored more liberally. NAM simulates performance on tasks that require item-specific gesture responses.

We also examined whether correct naming in the model was sometimes contingent on the appropriate action being selected, and, contrariwise, whether retrieval of the wrong name sometimes led to the wrong action being selected. Figure 11 gives an example in which correct naming appeared to be contingent on retrieval of the appropriate example. Here the structure description for the object "pen" was presented. Activity in the action lexicon was rapidly generated and the appropriate output unit was activated above threshold (dotted line). Activity in the phonological output lexicon was generated more slowly, and initially the unit most activated did not correspond to the target ("pen") (solid lines). However, once the action unit for "pen" reached threshold, so support was provided (via the semantic system, see Figure 2) for the correct name,

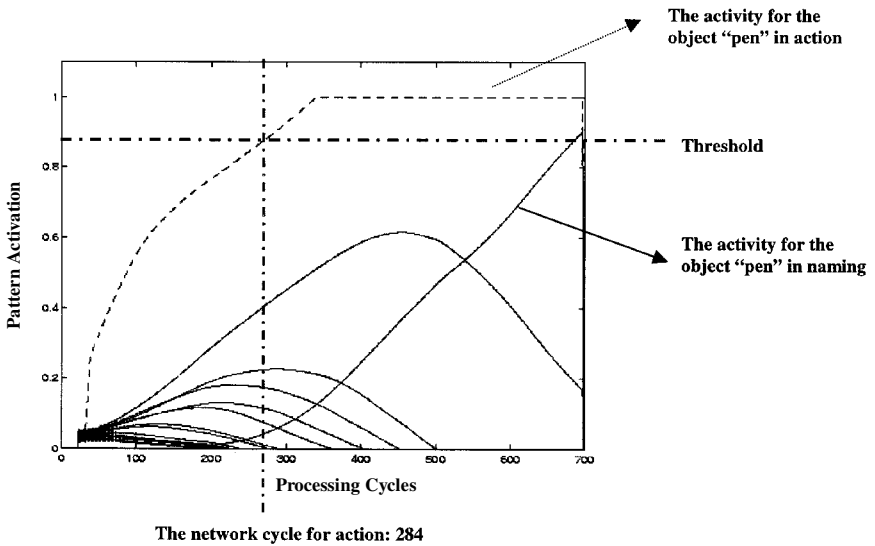


Figure 11. The effect of action retrieval on naming. The dotted line corresponds to the activation of the object “pen” in the action output lexicon. The solid lines indicate activation in the phonological output lexicon.

enabling the name unit for “pen” to gain in activation so that it eventually reached threshold. Action retrieval facilitated naming.

On some occasions, the opposite also occurred. In the example in Figure 12, the structural description for the object “ring” was presented. Although there was greater activation in the action output lexicon (dotted lines) than in the phonological lexicon (solid line), there was competition for action selection between the responses appropriate for “ring” and “bracelet”. This delayed action selection and enabled a name unit to reach the threshold first (for the incorrect name, “bracelet”). The selected name then provided support for a matching action (through the semantic system), driving the action for “bracelet” to threshold. Here an incorrect action was produced, due to retrieval of the inappropriate name.

Figures 11 and 12 illustrate that action and name selection interact in NAM, with the direction of the effects dependent on whether action or name information is dominant first. Generally action retrieval is faster (Figure 9), so that, on average, actions support naming.

In addition to showing more errors in naming than in action selection, optic aphasic patients also make a variety of errors in object naming, including both visual and semantic (and semantic/visual) errors (see Hillis & Caramazza, 1995; Lhermitte & Beauvois, 1983; Riddoch & Humphreys, 1987). The

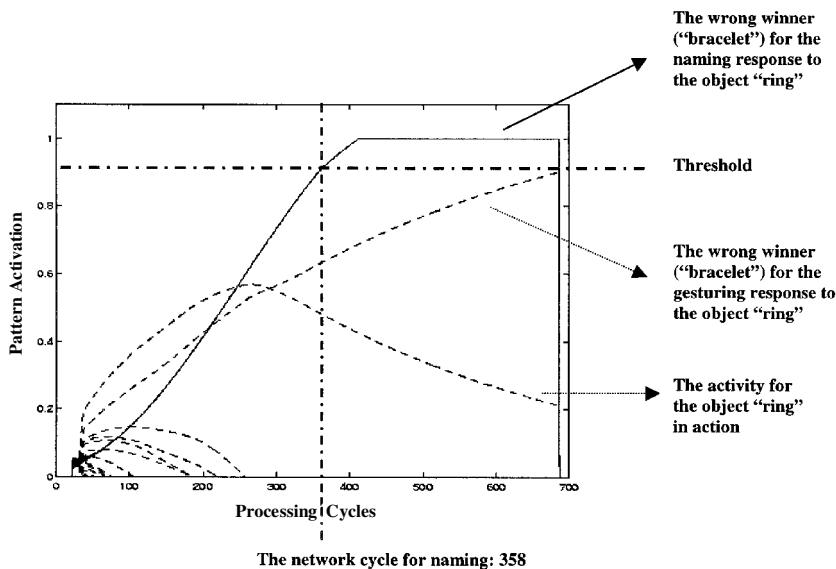


Figure 12. The effect of (incorrect) naming on gesturing. The solid line is the activation of the name "bracelet", though the object "ring" was presented to NAM. The dotted lines indicate activation in the action output lexicon.

proportions of these contrasting error types can differ across patients (see Davidoff & de Bleser, 1993; see also Riddoch, 1999, for further discussion). We assessed error types by breaking down the error scores in the model according to whether the non-targets had visual or semantic + semantic/visual relationships with target objects, using the scoring criteria applied by judge 1 (Experiment 2).⁴ The percentages of the error scores that were due to visual or semantic + semantic/visual errors are given in Figure 13, for object naming and action retrieval. Both visual and semantic errors (semantic + semantic/visual) occurred. Nevertheless, relative to normal performance under response deadline conditions (Figure 8), there was a disproportionate increase in semantic relative to visual errors, for both naming and action retrieval.

It is interesting that semantic + semantic/visual errors tended to increase disproportionately after a lesion that disrupts input into the semantic system. In a prior simulation of optic aphasia, Plaut and Shallice (1993) similarly found that a pattern of semantic and semantic/visual errors emerged even when there was a pre-semantic lesion. For both their model and ours, such errors occur

⁴Note that there were very few differences between the judges, and judge did not interact with any of the factors influencing the performance of the lesion model.

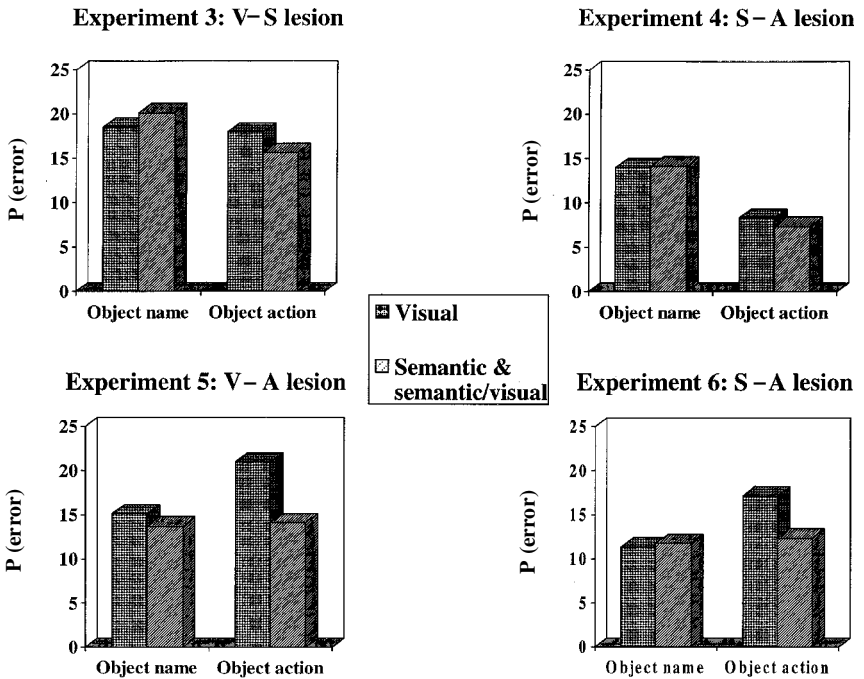


Figure 13. The proportions of the residual error that corresponds to visual and semantic (semantic + semantic/visual) errors after lesioning NAM.

because visual input vectors were directed by the lesion into the wrong part of semantic space. The simulations provide an existence proof that we should not conclude that a patient has a deficit at a semantic level simply due to relatively high proportions of semantic and semantic/visual errors being made.

In sum, a lesion that affects mapping from the structural description system to item-specific knowledge in the semantic system can generate patterns of deficit resembling those found in optic aphasia. Following the lesion, the overall performance for both naming and action to seen objects was disturbed. Nevertheless, NAM was more likely to make errors in selecting object names than in selecting actions, although the retrieval of super-ordinate names was relatively preserved. In addition, semantic + semantic/visual errors tended to increase relative to pure visual errors. The relative preservation of action retrieval over name retrieval occurs because, in NAM, action retrieval is supported by a direct route from visual to action. Semantic errors occur because the noise added to the mappings into semantics leads to stimuli being represented incorrectly within the semantic system.

EXPERIMENT 4

Anomia

In Experiment 4 we examined the effects of introducing a lesion to the mappings between the semantic system and phonological name output units. In the neuropsychological syndrome of anomia, patients can have preserved access to semantic knowledge about objects, and they can produce appropriate actions to objects, but they showed selectively impaired naming (e.g., Kay & Ellis, 1987). For NAM, this pattern should emerge after lesioning the “output route” from semantics to the phonological names of objects.

Method

Lesioning and testing was performed in the same way as for Experiment 3, except that noise was introduced to the mappings between the semantic system and the phonological name system (locus (b), Figure 2). We also reduced input from the semantic system to the phonological output system to 60% of its normal value.

Results and discussion

The mean RTs and error rates across items are illustrated in Figures 9 and 10. RTs were increased and error rates raised relative to the unlesioned version of the model, with this effect being particularly large for object naming (there was a significant interaction between task and lesioning for both RTs and errors, $F(3, 93) = 135.81$ and 1084.51 , both $p < .001$, for RTs and errors respectively). RTs were faster and error rates lower for action selection than for name selection; for RTs, $t(31) = 10.68$, $p < .001$, and for errors, $t(31) = 44.87$, $p < .001$. In contrast to when there was lesioning of input from objects to semantics, there was also some difficulty for the model in retrieving the superordinate names of the objects presented; thus, RTs were increased and error rates raised for superordinate as well as for item-specific naming, relative to action retrieval: for RTs, $t(31) = -2.37$, $p < .025$, and for errors, $t(31) = -27.59$, both $p < .001$. In anomia, patients may often provide high-level category information about objects (e.g., describing a dog as an animal), but to our knowledge studies have not tested whether such patients are impaired at retrieving “intermediate level” category information (e.g., discriminating between accessories, household objects, and stationery items), which is tested here. NAM predicts that such patients should encounter some difficulties with such intermediate level categories.

In contrast to Experiment 3, there was only a small increase in the error rate in the action retrieval tasks. Note that, here, both the semantic and the direct route to action were left undamaged, whereas there was damage to the semantic route to action in Experiment 3. The relative lack of an effect of the present lesion on action retrieval demonstrates that action retrieval for the

model is not affected by changes in the efficiency of name retrieval, when the processes that lead to action retrieval are undamaged.

Figure 13 gives the relative proportions of the errors that could be classified as either visually or semantically (semantic + semantic/visual) related to targets. There were proportionately fewer visual and semantic errors here than in Experiment 3, due to an increase in “other” (unrelated) errors. The pattern of visual to semantic errors, however, did not differ greatly. Studies of anomic patients with deficits at this late stage of name retrieval reveal that the patients typically make no response rather than either visual or semantic errors (Kay & Ellis, 1987).

EXPERIMENT 5

Visual apraxia

In Experiment 5 we present data on the effects of lesioning the direct route from structural descriptions of objects to action (locus (c), Figure 2). In contrast to Experiments 3 and 4, this should disrupt action rather than name selection, so generating a form of apraxia. In particular, the damage may generate a modality-specific visual apraxia, in which action selection to visually presented objects is worse than action selection to object names (see DeRenzi, Faglioni, & Sorgato, 1982; Pilgrim & Humphreys, 1991; Riddoch et al., 1989). Note that this pattern, if apparent, would stem from NAM’s adoption of an interactive framework, in which action retrieval is dependent on convergent activation from the visual and semantic routes. In this case, damage to the visual route could “block” retrieval via a semantic route. This would not follow from a more standard dual-route framework, where the visual and semantic routes would operate independently.

Method

The only difference relative to the prior lesion studies was that noise was added to the mappings between the structural description system and the action selection system, leaving other parts of the model intact. Input from the structural description system to the action selection system was also reduced to 60% of its normal value.

Results and discussion

The mean RTs and error rates are given in Figures 9 and 10. The data contrast with those reported in Experiments 3 and 4. Whereas name selection to objects was previously more likely to generate errors than action selection, the opposite results were found here. Comparing performance of the lesioned model to the intact version, there were interactions of Task \times Lesioning for both RTs and errors, $F(3, 93) = 30.04$ and 164.56 , both $p < .001$. Whereas, in the intact

version, action selection from objects tended to be faster than object categorization (Figure 9), here action selection from objects was slowed relative to object categorization, $t(31) = 3.91$, $p < .001$. Although RTs and errors increased for object naming as well as action retrieval, the increase was particularly large for action retrieval. Error rates were also increased for action selection to objects, relative to both object naming and object categorization, $t(31) = 17.41$ and 19.87 , both $p < .001$. There can be selective damage to the procedures more involved much greater in action retrieval than in name retrieval for objects. In addition, action selection to words remained relatively fast and accurate (Figure 10).

Figure 13 presents the proportions of visual and semantic (semantic + semantic/visual) errors made in object and action selection in this lesioned version of NAM. There tended to be proportionately fewer visual and semantic errors in object naming than in action retrieval. For action retrieval there were more visual than semantic errors, but this did not differ drastically from the effects of the other lesions.

NAM here mimics data showing that relatively good object recognition can occur in patients with visual apraxia (i.e., there should be good object categorization in patients who are impaired at retrieving actions for visually presented objects; Riddoch et al., 1989). In addition, the retrieval of action from words is relatively spared, in both patients and the model. The model is able to capture the qualitative patterns of spared and impaired processes in this syndrome. It is of some interest that a model such as NAM, with a dual-route architecture, produces a deficit in retrieving output even though access to that output can be achieved through an undamaged route—here the semantic route, which continued to operate relatively well (e.g., for action selection from words). For NAM, access to actions from the semantic route is disrupted by the damaged visual route when objects are presented. This blocking effect occurs because action selection in the model is contingent on the convergence of activity from both the visual and the semantic routes. Here the noise within the visual route was sufficiently high to override activation generated semantically, which would otherwise be sufficient for actions to be selected successfully. Indeed, due to input from the direct visual route (albeit noisy), there can be faster convergence of activation in the action selection system than in the semantic system, minimizing the impact of intact semantic information on action retrieval. In essence damage to the fast direct route “pushes” activation away from the correct unit in the action output lexicon, and this is not overcome by correct semantic input. One other point to note is that object naming was also disrupted to some degree by the lesion (albeit to a larger extent than action retrieval; see Figures 9 and 10). This occurred because early activation in the action output lexicon was fed back to influence semantic activation, which in turn affected object naming. NAM predicts that naming should not be completely intact in visual apraxia.

EXPERIMENT 6

Apraxia due to a damaged semantic route to action

Many patients with apraxia are more impaired when required to gesture to a name than when asked to gesture with the object present (see DeRenzi et al., 1982). This need not be a semantic problem, since object recognition can be intact. Within the framework of NAM, this pattern could follow from damage to a semantic route to action (affecting action selection to words) along with a preserved visual route. This was evaluated in Experiment 6.

Method

The mappings from the semantic system to the action selection system were lesioned (using the same lesioning procedure as before, with both noise added and input from the semantic to the action system reduced to 60% of normal). NAM was tested as in Experiments 3–5.

Results and discussion

Figures 9 and 10 reveal the mean RTs and error rates. Now a quite different pattern of results emerged compared with those found in Experiment 5. Responses were particularly slow and error prone for action selection to words. When compared with the unlesioned version of the model there were interactions between task and lesioning, for both RTs and errors, $F(3, 93) = 183.70$ and 498.74 , both $p < .001$. For action selection from words, RTs were slower and there were more errors than for action selection from objects, $t(31) = -11.13$ and -13.0 , both $p < .001$, for RTs and errors respectively. Action selection from words was also worse than object naming, $t(31) = -13.67$ and -18.32 , both $p < .001$, for RTs and errors respectively. Thus, there was selective disruption of action selection for words, even though actions could be selected to objects and objects could be named. This mimics the pattern found in many apraxic patients (DeRenzi et al., 1982).

Interestingly, as we observed in Experiment 3, damage to the semantic route to action again had some impact on action selection for objects, even though the direct visual route was preserved here. For example, the error rate in action selection to objects was raised when compared to when lesioning affecting the mappings from semantics to names (i.e., relative to when a comparable “output” lesion was made, in Experiment 4; see Figure 10). NAM makes the quantitative prediction that some deficit in action selection to objects should still be found in patients with all forms of apraxia, when compared with normal subjects.

The relative percentages of visual and semantic (semantic + semantic/visual) errors were relatively low, as found with the other lesion of output from

the semantic system (Experiment 4; see Figure 13). Damage to output from the semantic system can lead to unrelated errors rather than necessary increases in visual and semantically related errors.

GENERAL DISCUSSION

NAM is a dual-route model of action selection, incorporating a direct visual route to action along with an indirect semantic route. In its normal state, NAM is able to simulate patterns of performance found with normal human subjects. For the task of action selection, RTs are faster to objects than to words, and action selection for objects is faster than name selection and even classification of objects into superordinate groups (Experiment 1). Chainay and Humphreys (in press) reported similar effects when subjects had to decide whether objects were found in a kitchen, in the classification task. When an artificially fast response deadline was used, then NAM has an increased probability of making an error and, with action selection to objects, there are proportionately more errors that are visually related to targets than errors that are semantically related. With words, this trend for more visual than semantic errors was reversed (Experiment 2; see Figure 8). This matches the data reported by Rumiati and Humphreys (1998).

NAM also breaks down in a manner that simulates findings from the human neuropsychological literature. When noise was added to the mappings between the structural description system and the semantic system, performance was similar to that found in the syndrome of optic aphasia (Experiment 3). The model was impaired at object naming but superordinate classification and action selection were both relatively spared. In addition there was a relative increase in the proportion of semantic to visual errors made in object naming (Figure 13), consistent with the neuropsychological data. When the lesion affected the output mappings from the semantic system to the phonological output system, there was again difficulty in object naming, though this now affected the retrieval of superordinate names as well as the retrieval of specific names. There were also fewer errors on action selection than when the lesion affected input from objects into the semantic system (Experiment 4; see Figure 13). The pattern of impaired naming and preserved action matches that found in anomia (e.g., Kay & Ellis, 1987). The increased errors in action selection after a lesion into the semantic system (in Experiments 3 and 6) also illustrates that action selection in NAM is dependent on convergent input from the indirect (semantic) as well as the direct (visual) route. For instance, in Experiment 3 the semantic route to action was disrupted by the lesion affecting input into the semantic system, although the direct visual route was preserved. Nevertheless, errors were increased in action selection relative to when input into the semantic system was intact (though action selection errors remained less than object naming errors with the input lesion; see earlier).

In Experiments 5 and 6, lesions were applied specifically to the processes mediating action selection. Lesions to the direct visual pathway primarily affected action selection to visual objects; action selection to words was less affected (Experiment 5). In contrast, lesions affecting the mapping between semantics and the action lexicon impaired action selection to words more than to objects (Experiment 6). The opposite effects of these lesions is congruent with the double dissociation between apraxic patients who, in one case, are better at acting to words than to objects, and, in the other, are better at acting to objects than words (e.g., DeRenzi et al., 1982; Pilgrim & Humphreys, 1991; Riddoch et al., 1989).

The lesioning data provide an existence proof that a model with a dual-route architecture can generate the patterns of deficit present in the human neuropsychological literature. Particularly interesting is that finding that, despite the dual-route architecture, damage to one route affected action retrieval through the other. For NAM, the visual route to action may be thought relatively dominant, since actions are activated more rapidly via this route (extra connections must be passed through for objects to activate actions through the semantic system; see Figure 5).⁵ Despite this, damage to the semantic route to action (Experiments 3 and 6) disrupted visual action retrieval to some degree. In addition, damage to the visual route itself led to impairments in the retrieval of actions to objects through the semantic system, so that the deficits associated with visual apraxia were produced (Experiment 5). This pattern emerges in NAM because the model depends on convergence across routes for outputs to be selected. The model incorporates both a modular architecture and functional interactivity, and this contribution is necessary to account for the full set of results. This is clearly useful in order to capture neuropsychological data. The modular architecture allows processes to be lesioned selectively, so generating overall dissociations between different patients (or different lesion sites, in the model), whereas functional interactivity between routes leads to "blocking" effects even when one route is undamaged. If a modular, dual-route architecture was used without the kind of functional interactivity apparent in NAM, then disorders such as visual apraxia become difficult to explain.

The functional interactivity within NAM is also helpful in accounting for some of the qualitative patterns of performance found in disorders such as optic aphasia. For example, such patients can sometimes use gesturing to prompt name retrieval; on other occasions, however, they may pronounce an incorrect name and produce an action in accordance with that (Lhermitte & Beauvois,

⁵Indeed the data from Experiment 2, where more visual than semantic errors were produced when actions were made to a deadline, are a consequence of this.

1973; Riddoch & Humphreys, 1987). These interactions were evident too in NAM, where activation within one route could be influenced by the emergence of a winning item along the other route (even if the winner was incorrect, leading to erroneous responses action and name responses; see Figures 11 and 12). Whichever route is more influential depends upon the dynamics of activation; for objects, activation tended to rise fastest for action responses, making performance biases to favour action over naming (see Figure 5).

New predictions

As well as accounting for a broad set of data from both normal observers and neuropsychological patients, several new predictions can be derived from NAM's performance. These are: (1) In optic aphasia ($V \rightarrow S$ lesion) there should be some slowing and increase in errors when actions are made to objects, relative to normal subjects (Figures 11 and 12). This effect should vary according to the specificity of the required action; more errors are predicted when more specific responses must be made; (2) in anomia ($S \rightarrow P$ lesion) there may be an impairment in fine-grained object categorization, as well as in object naming; (3) in visual apraxia ($V \rightarrow A$ lesion), object naming is unlikely to be perfect; and (4) in other forms of apraxia ($S \rightarrow A$ lesion), there should be deficits on acting to objects as well as to words, and there may be a somewhat weaker impairment on object naming.

To the best of our knowledge, these predictions have not been evaluated in detail in the neuropsychological literature (e.g., due to generic rather than item-specific gestures being examined), and we look forward to the predictions being put to the test. Note also that, for the most part, the predictions concern associated deficits that may occur in patients in addition to their primary deficit (e.g., a deficit in object categorization as well as naming, in anomia). Associated deficits are often difficult to interpret in neuropsychology, since they can be caused by anatomical damage to separable neural subsystems (see Humphreys & Price, 2001, for one recent account). An explicit computational model, such as NAM, gives a principled framework for explaining such data patterns. Even in this framework, though, the degree of association will depend on the magnitude of the lesion. With smaller lesions, only the primary deficit will tend to occur (an object naming impairment, in anomia).

Extending the model to account for other data

In this paper NAM has been applied to account for RT patterns in normal action and name selection to objects and words, error patterns when normal subjects respond to a fast deadline, and patterns of impairment found in brain-lesioned patients. However, the model has not simulated the effects of action priming in tasks such as deciding whether objects are in an appropriate orientation (e.g.,

Tucker & Ellis, 1998, 2001; see also Craighero et al., 1998, 1999), or the visually driven utilization behaviours found in tasks where patients have to respond to the location of a stimulus whilst suppressing a more familiar response (e.g., see Riddoch et al., 1998, 2000). In order to capture these additional results, we would need to extend NAM to include not only action and name selection, but also information about the orientation and locations of objects (e.g., to enable responses to be made to the locations of objects). In principle it should not be too difficult to do this. For example, we could incorporate output units that could represent occupied locations in the environment, so that activation of these locations would dictate where an action of a particular class (determined by the action selection system) should be made. Utilization behaviours, of the type noted by Riddoch et al., would arise when the familiar action, linked with an object in one location, is activated more than an action to the location determined by the task set. Current work is underway to extend the model in this way.

Neural substrates of action retrieval

There is considerable neurophysiological and neuropsychological evidence implicating the role of the dorsal visual system in the control of prehensile actions to objects (e.g., see reviews by Jeannerod, 1997; Milner & Goodale, 1995). However, we have been concerned not with the guidance of prehensile actions but with the selection of which articulated action to perform on an object—to lift a jug to pour rather than to drink out of—once a reach and grasp have been effected. The data on the neural substrates of this process of action selection are less comprehensive than the data on prehensile actions, and there is evidence that action selection is mediated by ventral as well as dorsal visual streams. Studies using functional brain imaging, for example, have shown that there is activation of regions in ventral cortex, particularly the left medial temporal and the inferior frontal lobes, across a range of tasks that putatively involve action retrieval; these include: Naming tools, naming the action performed tools, identifying and imagining gestures with objects (e.g., Decety et al., 1997; Grabowski, Damasio, & Damasio, 1998; Grafton, Fadiga, Arbib, & Rizzolatti, 1997; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995). There is also, however, some indication of activation in areas linked to the dorsal visual stream, such as area MT, in similar tasks (Chao, Haxby, & Martin, 1999; Martin et al., 1995). However, these studies have typically not compared action retrieval across different modalities of input, to attempt to isolate a direct visual route from a supramodal semantic route to action. Thus the areas implicated could be important solely for action selection based on semantic knowledge.

One imaging study that has attempted to distinguish between direct and indirect routes to action was conducted by Phillips, Humphreys, Noppeney,

and Price (this issue). They had subjects make action decisions (would you make a pouring or twisting action with this object?) to either words or pictures of objects, comparing activation in these tasks with that found when physical size decisions were made to the same stimuli. Action and size decisions were also made to pictures of non-objects. They found that both the middle temporal and inferior frontal regions of the left hemisphere were activated in the action compared with the size decision tasks, but this was equally the case for objects and words. Thus, these regions seem to mediate action selection irrespective of the input modality, and they serve as good candidates for a semantic route to action. In contrast, regions of the left inferior, posterior temporal lobe were more activated to objects and non-objects relative to words, for action relative to size decisions. These regions are implicated in high-level visual processing of stimuli (e.g., Ungerleider & Haxby, 1994). The data suggest that there is increased high-level visual processing of objects and object parts in ventral visual areas in order to sustain action selection to objects. Activation in this region may subsequently project to more anterior brain regions to provide a direct visual route to action. It is interesting to note that, in NAM, the direct visual route is based on a distributed pattern of activation that reflects the parts that are common to several objects. This could lead to parts-based activation of actions to non-objects as well as to objects. The direct visual route may be said to “afford” action even when objects may not be recognized as known exemplars.

Phillips et al. (this issue) failed to find any evidence for a dorsal route to action selection. Nevertheless, there is neurophysiological evidence for cells in inferior parietal cortex firing when a particular action is performed with an object (Taira, Mine, Georgopoulos, Murata, & Sakata, 1991). In addition, Sakata and colleagues (1998) have reported that cells in this region are sensitive to the orientation of objects in three dimensions, suggesting that they can support articulated actions to objects and not just prehensile movements to particular locations. It remains possible, then, that a direct route to action is supported also through more dorsal visual areas. A question for future research to examine is whether dorsal and ventral areas differ in the types of action retrieval process they support (e.g., “afforded” actions to object parts vs retrieved actions to known objects).

Perception and action

NAM holds that perceptual processes directly constrain the selection of actions to objects. These processes can operate in parallel with action retrieval processes based on the activation of semantic knowledge about stimuli. The question of the relations between action and visual object perception is currently undergoing some debate (e.g., Rossetti, *in press*). There is considerable evidence indicating that prehensile actions can be dissociated from

conscious perception, coded within the ventral visual stream (e.g., Milner & Goodale, 1995; Pisella et al., 2000). Nevertheless, there are circumstances too in which prehensile actions are affected by perceptual information likely to be coded by ventral brain areas—such as actions determined by the colour of stimuli (Pisella et al., 2000). Stored knowledge can also influence simple reach and grasp movements (Jeannerod, Decety, & Michel, 1994; Riddoch et al., 1998). Several accounts can be put forward to explain both the dissociations and the interactions between perception and action. Thus, Rossetti has argued that the timing of the actions is crucial. Fast prehensile actions are based on a dorsal visual route, whereas slow actions may be governed by a ventral visual pathway. NAM does not simulate prehensile actions for reaching and grasping, but rather the selection of a category of action that may be performed once a grasp has been completed. The model assumes that the process of action selection is contingent on the activation of stored object knowledge, though this is distributed across exemplars and may not correspond to any one known object in particular. Now it is possible that information about the category of action (e.g., to drink or pour) is retrieved in parallel with the initiation of a grasp to an object, but operates over a longer time course and so is influenced by stored knowledge. The inter-relations between reaching and grasping on the one hand, and action selection on the other, have hitherto been little studied, but the development of an explicit model of the selection process invites an analysis of how selection, reaching, and grasping interface. For now, NAM shows how psychological and neuropsychological evidence can be captured if visual information is allowed not only to affect the reach-and-grasp component but also the action selection process.

REFERENCES

- Beauvois, M.F. (1982). Optic aphasia: A process of interaction between vision and language. *Philosophical Transactions of the Royal Society*, *B298*, 35–47.
- Biederman, I., & Cooper, E.E. (1991). Object recognition and laterality: Null effects. *Neuropsychologica*, *29*, 685–694.
- Caramazza, A., Hillis, A.E., Rapp, B.C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology*, *7*, 161–189.
- Chainay, H., & Humphreys, G.W. (in press). Privileged access to action for objects. *Psychonomic Bulletin & Review*.
- Chao, L.L., Haxby, J.V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*, 913–919.
- Cohen, J.D., Dunbar, K., & McClelland, J.L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.

- Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1998). Visuomotor priming. *Visual Cognition*, 5, 109–126.
- Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1999). Action for perception: A motor-visual attentional effect. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1673–1692.
- Davidoff, J., & De Bleser, R. (1993). Optic aphasia: A review of past studies and a reappraisal. *Aphasiology*, 7, 135–154.
- Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., & Procyk, E. (1997). Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain*, 120, 1763–1777.
- DeRenzi, E., & Faglioni, P. (1999). Apraxia. In G. Denes & L. Pizzamiglio (Eds.), *Handbook of clinical and experimental psychology* (pp. 421–440). Hove, UK: Psychology Press.
- DeRenzi, E., Faglioni, P., & Sorgato, P. (1982). Modality-specific and supramodal mechanisms of apraxia. *Brain*, 105, 301–312.
- Funnell, E. (1987). Object concepts and object names: Some deductions from acquired disorders of word processing. In G.W. Humphreys & M.J. Riddoch (Eds.), *Visual object processing: A cognitive neuropsychological approach* (pp. 233–264). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Glaser, W. (1994). Picture naming. *Cognition*, 42, 61–105.
- Grabowski, T.J., Damasio, H., & Damasio, A.R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *Neuroimage*, 7, 232–243.
- Grafton, S.T., Fadiga, L., Arbib, M.A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*, 6, 231–236.
- Heinke, D., & Humphreys, G.W. (in press). Attention, spatial representation and visual neglect: Simulating emergent attention and spatial memory in the Selective Attention for Identification Model (SAIM). *Psychological Review*.
- Hillis, A.E., & Caramazza, A. (1991). Constraining claims about theories of semantic memory: More on unitary versus multiple semantics. *Cognitive Neuropsychology*, 12, 175–186.
- Hillis, A., & Caramazza, A. (1995). Cognitive and neural mechanisms underlying visual and semantic processing: Implication from “optic aphasia”. *Journal of Cognitive Neuroscience*, 7, 457–478.
- Hinton, G., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74–96.
- Hodges, J.R., Salmon, D.P., & Butters, N. (1991). The nature of the naming deficit in Alzheimer's and Huntington's diseases. *Brain*, 114, 1547–1558.
- Hodges, J.R., Spatt, J., & Patterson, K. (1999). “What” and “how”: Evidence for the dissociation of object knowledge and mechanical problem-solving skills in the human brain. *Proceedings of the National Academy of Sciences*, 96, 9444–9448.
- Humphreys, G.W., Forde, E.M.E., & Francis, D. (2000). The sequential organization of actions. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes*. Cambridge, MA: MIT Press.
- Humphreys, G.W., & Heinke, D. (1998). Spatial representation and selection in the brain: Neuropsychological and computational constraints. *Visual Cognition*, 5, 9–47.
- Humphreys, G.W., Lamote, G., & Lloyd-Jones, T.J. (1995). An interactive activation approach to object processing: Effects of structural similarity, name and frequency and task in normality and pathology. *Memory*, 3, 535–586.
- Humphreys, G.W., & Price, C.J. (2001). Cognitive neuropsychology and functional brain imaging: Implications for functional and anatomical models of cognition. *Acta Psychologica*, 107, 119–153.
- Humphreys, G.W., Price, C.J., & Riddoch, M.J. (1999). From objects to names: A cognitive neuroscience approach. *Psychological Research*, 62, 118–130.
- Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford, UK: Blackwells.

- Jeannerod, M., Decety, J., & Michel, F. (1994). Impairment of grasping movements following a bilateral posterior parietal lesion. *Neuropsychologia*, *32*, 369–380.
- Job, R., Rumiati, R., & Loto, L. (1992). The picture superiority effect in categorization: Visual or semantic? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1019–1028.
- Kay, J., & Ellis, A.W. (1987). A cognitive neuropsychological case study of anomia: Implications for psychological models of word retrieval. *Brain*, *110*, 613–629.
- Lauro-Grotto, R., Piccini, C., & Shallice, T. (1997). Modality-specific operations in semantic dementia. *Cortex*, *33*, 593–622.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in language production. *Behavioral and Brain Sciences*, *22*, 1–38.
- Lhermitte, F. (1983). Utilisation behaviour and its relation to lesions of the frontal lobes. *Brain*, *106*, 237–255.
- Lhermitte, F., & Beauvois, M.F. (1973). A visual-speech disconnection syndrome: Report of a case with optic aphasia. *Brain*, *96*, 695–714.
- Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.
- Martin, A., Haxby, J.V., Lalonde, F.M., Wiggs, C.L., & Underleider, L.G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, *270*, 102–105.
- Milner, A.D., & Goodale, M. (1995). *The visual brain in action*. London: Academic Press.
- Mjolsness, E., & Garrett, C. (1990). Algebraic transformations of objective functions. *Neural Networks*, *3*, 651–669.
- Patterson, K., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K.E. Patterson, J.C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia* (pp. 333–360). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Phillips, J., Humphreys, G.W., Noppeney, S., & Price, C.J. (this issue). The neural substrates of action retrieval: An examination of semantic and visual routes to action. *Visual Cognition*, *9*(4/5), 662–684.
- Pilgrim, E., & Humphreys, G.W. (1991). Impairment of action to visual objects in a case of ideomotor apraxia. *Cognitive Neuropsychology*, *8*, 459–473.
- Pisella, L., Grea, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., Boisson, D., & Rossetti, Y. (2000). An “automatic pilot” for the hand in human posterior parietal cortex: Toward reinterpreting optic ataxia. *Nature Neuroscience*, *3*, 729–736.
- Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.
- Plaut, D.C., & Shallice, T. (1993). Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience*, *5*, 89–117.
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*, 263–266.
- Potter, M., & Faulconer, B.A. (1975). Time to understand pictures and words. *Nature*, *253*, 437–438.
- Reason, J.T. (1984). Lapses of attention in everyday life. In W. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 515–549). Orlando, FL: Academic Press.
- Riddoch, M.J. (1999). Optic aphasia. In G.W. Humphreys (Ed.), *Case studies in the neuropsychology of vision* (pp. 133–160). Hove, UK: Psychology Press.
- Riddoch, M.J., Edwards, M.G., Humphreys, G.W., West, R., & Heafield, T. (1998). Visual affordances direct action: Neuropsychological evidence from manual interference. *Cognitive Neuropsychology*, *15*, 645–684.
- Riddoch, M.J., & Humphreys, G.W. (1987). Visual object processing in a case of optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology*, *4*, 131–185.

- Riddoch, M.J., Humphreys, G.W., & Edwards, M.G. (2000). Visual affordance and object selection. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes*. Cambridge, MA: MIT Press.
- Riddoch, M.J., Humphreys, G.W., Heslop, J., & Castermans, E. (in press). Dissociations between object knowledge and everyday action. *Neurocase*.
- Riddoch, M.J., Humphreys, G.W., & Price, C.J. (1989). Routes to action: Evidence from apraxia. *Cognitive Neuropsychology*, 6, 437–454.
- Rosetti, Y. (in press). Several different “vision for action” systems: A guide to dissociating and integrating dorsal and ventral functions. In W. Prinz & B. Hommel (Eds.), *Attention and performance XIX: Common mechanisms in perception and action*. Oxford, UK: Oxford University Press.
- Rothi, L.J.G., Mack, L., & Heilman, K.M. (1986). Pantomime agnosia. *Journal of Neurology, Neurosurgery and Psychiatry*, 49, 451–454.
- Roy, E.A., & Square, P.A. (1985). Common considerations in the study of limb, verbal, and oral apraxia. In E.A. Roy (Ed.), *Neuropsychological studies of apraxia and related disorders* (pp. 111–162). Amsterdam: North-Holland.
- Rumelhart, D.E., & McClelland, J.L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60–94.
- Rumiati, R.I., & Humphreys, G.W. (1998). Recognition by action: Dissociating visual and semantic routes to actions in normal observers. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 631–647.
- Sakata, H., Taira, M., Kusunoki, M., Murata, M., Tanaka, Y., & Tsutsui, K. (1998). Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. *Philosophical Transactions of the Royal Society*, B353, 1363–1373.
- Sirigu, A., Duhamel, J.-R., & Poncet, M. (1991). The role of sensorimotor experience in object recognition. *Brain*, 114, 2555–2573.
- Taira, M., Mine, G., Georgopoulos, A.P., Murata, A., & Sakata, H. (1991). Parietal cortex neurons of the monkey related to the visual guidance of hand movements. *Experimental Brain Research*, 83, 29–36.
- Tikhonov, A.N., & Arsenin, V.Y. (1977). *Solutions of ill-posed problems*. Chichester, UK: J. Wiley & Sons.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 830–846.
- Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, 8, 769–800.
- Ungerleider, L.G., & Haxby, J.V. (1994). “What” and “where” in the human brain? *Current Opinion in Biology*, 4, 157–164.
- Van Orden, G.C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory and Cognition*, 15, 181–198.
- Vitkovitch, M., & Humphreys, G.W. (1991). Perseverative responding in speeded naming to pictures: It’s in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 664–680.
- Vitkovitch, M., Humphreys, G.W., & Lloyd-Jones, T.J. (1993). On naming a giraffe a zebra: Picture naming errors across different categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 243–259.
- Young, A.W., & Burton, A.M. (1999). Simulating face recognition: Implications for modelling cognition. *Cognitive Neuropsychology*, 16, 1–48.

APPENDICES

Appendix A. NAM

Radial basis function (RBF) network

The structural description system of NAM contained two radial basis function (RBF) networks. One RBF network transformed the input description of objects into a distributed representation of objects and object categories, and the other network did the same for input descriptions of words.

An RBF network is a two-layered network (Poggio & Edelman, 1990) whose hidden layer uses a Gaussian output function:

$$y_i^{h-rbf} = f(\|x - w_i^h\|) \quad (1)$$

$$\text{with } f(x) = e^{-\frac{x^2}{2\sigma^2}}$$

Where $\|x - w_i^h\|$ is the Euclidian distance between the input (x) and the weights of the i -th unit (w_i^h). $f(x)$ is the Gaussian output function with σ as standard deviation. The units in the output layer combine the activations from the units in the hidden layers by a weighted sum:

$$y_j^o-rbf = \sum_{i=1}^n w_{ji}^o \cdot y_i^{h-rbf} \quad (2)$$

Where w_{ji}^o are the n weights for the j -th output unit.

For objects, the inputs (x) into the RBF network were the feature values of the input object. The number of units in the hidden layer was equal to the number of objects known to NAM. The weights of each unit in the hidden layer (w_i^h) were set to the feature values of the known objects, so that each hidden unit was linked to exactly one object and each unit had the maximum value of one when the object it represents is the input. The resulting output activation at the hidden layer level was a distributed representation of the input object, which captured the similarity between this object and all other known objects. To ensure this distributed representation, the standard deviation was set to 13. This value represented a compromise between two constraints: A distributed representation (large σ) and a sufficient contrast between activations (smaller σ). The second constraint resulted from the fact that the "winner takes all" (WTA) layers, which receive the activations as input (see the following section), need a reasonable contrast between activations, in order to converge successfully into a state with a clear winner.

For words as input a second RBF network followed an analogous approach. Here, the inputs (x) were the feature values of the input word. The number of units in the hidden layer was equal to the number of words and the weights of each unit encoded the feature values of the known words. The resulting output activation at the hidden layer level was a distributed representation of the input word. For the hidden layer of this RBF network the standard deviation had to be chosen so that the likelihood of a visual error for words was small and so σ was set to 1.0. The relationship between visual errors and σ results from an important property of the WTA mechanism used here (see later). The speed of convergence for inhibited units depends on their input activations. If the input activations are high, the convergence is delayed, whereas for low input activations the speed of convergence is increased. To determine the visual error, the output activations of WTA layers at a defined point of time were used. The point of time could be defined by either NAM's reaction time or a set deadline (see main text for details). So, if output activations were still high at this point of time, a large visual error was recorded. A faster convergence thus led to a smaller visual error, whereas a slower convergence resulted in a larger visual error. As mentioned earlier, a small σ leads to small output activations feeding into WTA layers and vice versa. Consequently, a small σ led to a fast convergence and a low visual error.

In NAM the output units of the RBF networks encoded the superordinate representation of objects and words. The weights to the output units (w_{ji}^o) were set with the help of a widely used interpolation method (Tikhonov & Arsenin, 1977). Let us assume for now that we aimed at a local representation of

superordinate representations. For this assumption to hold, the weights of the output layer (w_{ji}^o) had to fulfill the following equation for an object k :

$$(0,0,\dots,1,0,\dots,0) = \sum_{i=1}^n w_{ji}^o \cdot f(\mathbf{||}x_k - w_i^h\mathbf{||})$$

where $(0,0,\dots,1,0,\dots,0)$ denotes the targeted superordinate representation of the k -th object, given the feature values (x_k) as input for the k -th object. In the case of n objects, n of these equations have to be fulfilled:

$$\begin{pmatrix} 1 & 0 & \mathbf{K} \\ 1 & 0 & \mathbf{K} \\ \vdots & \vdots & \vdots \\ 0 & \mathbf{K} & 1 \end{pmatrix} = \begin{pmatrix} w_{00}^o & w_{0i}^o & \mathbf{K} & w_{0n}^o \\ w_{10}^o & w_{1i}^o & \mathbf{K} & w_{1n}^o \\ \vdots & \vdots & \vdots & \vdots \\ w_{n0}^o & w_{ni}^o & \mathbf{K} & w_{nn}^o \end{pmatrix} \cdot \begin{pmatrix} f(\mathbf{||}x_0 - w_0^h\mathbf{||}) & f(\mathbf{||}x_1 - w_1^h\mathbf{||}) & \mathbf{K} & f(\mathbf{||}x_n - w_n^h\mathbf{||}) \\ f(\mathbf{||}x_0 - w_i^h\mathbf{||}) & f(\mathbf{||}x_1 - w_i^h\mathbf{||}) & \mathbf{K} & f(\mathbf{||}x_n - w_i^h\mathbf{||}) \\ \vdots & \vdots & \vdots & \vdots \\ f(\mathbf{||}x_0 - w_n^h\mathbf{||}) & f(\mathbf{||}x_1 - w_n^h\mathbf{||}) & \mathbf{K} & f(\mathbf{||}x_n - w_n^h\mathbf{||}) \end{pmatrix}$$

In order to shorten this expression the previous equation is rewritten as:

$$Y^{o_rbf} = W^o \cdot H$$

where H are the activations of the hidden layer for all the objects. W^o are the weights to the output layer and Y^{o_rbf} are the superordinate representations for all the objects. Now the process of finding the weights to the output layer (W^o) turns into a simple matrix inversion:

$$W^o = Y^{o_rbf} \cdot H^{-1}$$

Note that this approach is only applicable when the number of units in the hidden layer are the same as the number of training patterns. However, this exact interpolation leads to a strictly local representation of object categories. In order to reach a distributed representation, an additional parameter λ was introduced, turning the equation for the weights into:

$$W^o = Y^{o_rbf} \cdot (H + \lambda \cdot I)^{-1}$$

where I is the identity matrix. λ makes the interpolation less precise and lets the RBF network generate a distributed representation at the output layer. In order to maintain the fact that the correct category unit for each object generates the largest activation, λ should be smaller than one. Here it was set 0.2 for both RBF networks.

Interactive activation within NAM

The core of the interactive activation system is based on winner take all (WTA) layers, which interact with each other. There are four separate WTA networks for (1) item-specific semantic knowledge (*ITEM*), (2) superordinate semantic knowledge (*SUP*), (3) the action output system (*ACT*), (4) the phonological name output system (*OUT*). For the WTA layer a mechanism suggested by Mjolsnes and Garrett (1990) was used. The following equations denote this mechanism for an arbitrary layer LA .

$$x_y^{la}(t) = x^{la}(t-1) + \tau^{la} \cdot \left(-x_i^{la}(t-1) - a^{la} \cdot \left(\sum_{i=1}^{n^{la}} f(x_i^{la}(t-1)) - 1 \right) + I_i(t-1) \right) \tag{3}$$

with an approximation of a sigmoid function:

$$f(x) = \begin{cases} 0 & 4 \cdot m^{la} \cdot (x - s^{la}) + 0.5 < 0 \\ 4 \cdot m^{la} \cdot (x - s^{la}) + 0.5 & 0 \leq 4 \cdot m^{la} \cdot (x - s^{la}) + 0.5 \leq 0 \\ 1 & 4 \cdot m^{la} \cdot (x - s^{la}) + 0.5 > 0 \end{cases}$$

The output of the WTA was derived from: $y_i(t) = f(x_i(t))$.

In order to shorten the documentation of NAM a shorthand notation for the WTA layer LA is used here:

$$y^{la}(t) = LA(I(t))$$

Note that this notation implicitly comprises the parameters of this particular layer, where n^{la} corresponds to the number units in the WTA layer, a^{la} as the strength of inhibition τ^{la} the speed of activation changes in the layer, m^{la} the slope of the sigmoid function, and finally, s^{la} the “threshold” of the sigmoid function.

With a visually presented object, activation at the action output layer (ACT) was given by:

$$y^{act}(t) = ACT(a^{a1} \cdot y^{h_rbf} + y^{item}(t)) \tag{4}$$

Where $a^{a1} \cdot y^{h_rbf}$ is the direct path from the structural description system to the action output layer and y^{item} is the indirect pathway from the semantic system. For words as input the direct path was omitted:

$$y^{act}(t) = ACT(y^{item}(t)) \tag{5}$$

Activation of the superordinate units (SUP) was based on:

$$y^{cat}(t) = SUP(y^{input2}(t) + y^{o_sup}(t) + a^{ci} \cdot W^{sup_item} \cdot y^{item}(t)) \tag{6}$$

y^{input2} can be either y^{o_rbf} for visually presented objects of y^{c_word} for words as input. y^{o_sup} is the feedback from the phonological name output. $W^{sup_item} \cdot y^{item}$ denotes a weighted sum of the output activation from the item-specific knowledge layer. The weighted sum transforms the item-specific activation into superordinate activation by summing the activation of all items belonging to the same category. Furthermore the weights were set in a way that the sum of activities was normalized, by dividing them by the number of items in a category.

The activation in the item-specific knowledge layer ($ITEM$) was given by:

$$y^{item}(t) = ITEM(y^{input1} + y^{o_item}(t) + a^{ci} \cdot W^{sup_item^T} \cdot y^{sup}(t) + a^{a1} \cdot y^{act}(t)) \tag{7}$$

y^{input1} can be either y^{h_rbf} for objects or word for words. y^{o_item} is the feedback from the phonological name output layer. $W^{sup_item^T} \cdot y^{sup}(t)$ denotes the feedback from the superordinate knowledge layer. The weighted sum in this term results in support for all items which are members of the same category in the layer.

The phonological name output layer (OUT) received input from the category layer and the item-specific layer:

$$[y^{o_item}(t), y^{o_sup}(t)] = OUT([y^{item}(t) + y^{word} + c_1, y^{sup}(t) + c_2]) \tag{8}$$

The notation [. , .] states the fact that two sets of units are combined to be part of one WTA layer. y^{word} was set to zero, if objects were used. c_1 and c_2 were the two control units. Their values determined, if naming operated at a superordinate level ($c_1 = 0$ and $c_2 = 1$) or an item-specific level ($c_1 = 1$ and $c_2 = 0$). The control results from modulating the competition between units in the phonological name output layer. So setting c_1 to one results in mainly superordinate output units (y^{o_sup}) winning the competition, whereas setting c_2 to one gives the item-specific units (y^{o_item}) an advantage, forcing NAM to operate on an item-specific level in naming.

Appendix B: Lesioning

NAM was “lesioned” by inserting noise into a specific pathway. The noise was based on a uniformly distributed random variable η which produced values in the range of 0 to 1.

$$y^{lesion} = a^{lesion} \cdot y^{pathway} + (b^{lesion} - a^{lesion}) \cdot \eta \quad (9)$$

The equation of y^{lesion} was chosen in a way that the parameter b^{lesion} determined the possible maximum value of y^{lesion} . This is possible because the maximum activation of every $y^{pathway}$ is one. This limitation aims at maintaining a correct working of the WTAs. Furthermore, $a^{lesion1}$ weights the amplitude of noise against the amplitude of pathway activation ($y^{pathway}$), where a small a^{lesion} gives rise to a large influence of noise whereas a large a^{lesion} reduces the influence of noise. Throughout this paper the parameter of the lesion was kept constant with $b^{lesion} = 1.5$ and $a^{lesion} = 0.6$.

Four different pathways were lesioned. The first lesion affected the pathway between the structural description system and the item-specific semantic knowledge layer. Here, equation (7) turned into:

$$y^{lesion1} = a^{lesion} \cdot y^{input} + (b^{lesion} - a^{lesion}) \cdot \eta$$

$$y^{item}(t) = ITEM(y^{lesion1} + y^{o_item}(t) + a^{ci} \cdot w^{sup_item^T} \cdot y^{sup}(t) + a^{a1} \cdot y^{act}(t)) \quad (10)$$

The second lesion affected the pathway between the semantic and the phonological output layer and turned equation (8) into:

$$y^{lesion2}(t) = a^{lesion} \cdot [y^{item}(t) + y^{word} + c_1, y^{sup}(t) + c_2] + (b^{lesion} - a^{lesion}) \cdot \eta$$

$$[y^{o_item}(t), y^{o_sup}(t)] = OUT(y^{lesion2}(t)) \quad (11)$$

The third lesioned pathway was the direct pathway between the structural description system and the action output layer. With this lesion equation (4) turned into:

$$y^{lesion3} = a^{lesion} \cdot y^{h_rbf} + (b^{lesion} - a^{lesion}) \cdot \eta$$

$$y^{act}(t) = ACT(y^{lesion3} + a^{a1} \cdot y^{item}(t)) \quad (12)$$

Finally, the fourth and last lesion affected the pathway between the item-specific knowledge layer and the action output layer. The equations were:

$$y^{lesion4} = a^{lesion} \cdot y^{item}(t) + (b^{lesion} - a^{lesion}) \cdot \eta$$

$$y^{act}(t) = ACT(a^{a1} \cdot y^{h_rbf} + y^{lesion4}(t)) \quad (13)$$

Appendix C: Parameters

The choice of parameters is an important aspect of NAM. However, as with other models of this type, there is a concern that any data pattern could be captured, if there are sufficient parameters. In order to minimize this concern, we aimed to keep the parameters as parsimonious as possible by keeping them the same in every WTA and by maintaining the strength of interactions between the layers at a value of one. However, this was not always possible as the following discussion shows.

The parameters of all layers, apart from the speed of the action output layer and the inhibition of the superordinate layer, were the same: $m^{sup} = m^{item} = m^{out} = 0.5$, $a^{sup} = a^{out} = 8$, $s^{sup} = s^{item} = s^{out} = 0.95$, $\tau^{sup} = \tau^{item} = \tau^{out} = 0.0038$.

The speed of the action output layer was set $\tau^{act} = 0.025$. The speed was set to a value different from that used within the other layers in order to slow down the selection of a winning node. This was

necessary because the action output layer received two inputs which would otherwise speed up the selection process beyond that justified by an experimental data. The parameter for the interaction between the superordinate layer and the item-specific layer (a^{ci}) was increased from 1 to 2 to inflate the influence of the superordinate layer. This was necessary, because with $a^{ci} = 1.0$, simulations for word inputs showed a similar proportion of semantic errors and visual errors in action, contrary to the empirical data. With the increase of a^{ci} the difference between semantic and visual errors was increased and provided a better fit to experimental data.

The influence of the direct route on the action output layer (a^{a1}) was set to 1.5. With $a^{a1} = 1.0$, gesturing to word stimuli was faster than gesturing to objects. This contradicts the experimental data, where gesturing to objects is faster than gesturing to words (see the main text for details). With an increase of a^{a1} to 1.5, NAM's simulation results matched the experimental data, because the higher influence of the direct pathway speeds up the WTA of the action output layer. As a consequence of increasing a^{a1} and a^{ci} , the item-specific layer received higher activation than before. To ensure convergence of this layer into a state with only winner (see Mjolsness & Garrett, 1990), the inhibition (a^{item}) was increased to 13. This alteration influenced the dynamics of NAM only marginally. It should be noted, however, that once set, the parameters for NAM were maintained across simulations (e.g., for both the "normal" and lesioned versions of the model).

Appendix D

The list of objects used

- | | |
|----------------|------------------|
| 1. Aerosol can | 17. Nail |
| 2. Axe | 18. Nut |
| 3. Banana | 19. Opener |
| 4. Bat | 20. Peg |
| 5. Boomerang | 21. Pen |
| 6. Bottle | 22. Pliers |
| 7. Bracelet | 23. Ring |
| 8. Doorknob | 24. Ruler |
| 9. Drill | 25. Saw |
| 10. Fork | 26. Screw |
| 11. Hair clip | 27. Screwdriver |
| 12. Hammer | 28. Spinning top |
| 13. Ice-cream | 29. Spoon |
| 14. Jug | 30. Stapler |
| 15. Knife | 31. Toothbrush |
| 16. Lighter | 32. Tweezers |

Object categorizations for the 32 items used

1. Accessories: Aerosol-can, bracelet, hair-clip, ring, toothbrush, tweezers
2. Tool: Axe, doorknob, drill, hammer, nail, pliers, saw, screw, screwdriver
3. Play/Sports: Bat, boomerang, spinning top
4. Household: Bottle, fork, knife, jug, lighter, opener, peg, spoon
5. Food: Banana, Ice-cream, nut
6. Stationery: Pen, ruler, stapler