Journal of Visual Languages and Computing (1990) **1**, 77–95

Naming Pictures*

STEPHEN M. KOSSLYN AND CHRISTOPHER F. CHABRIS

Harvard University, 1236 William James Hall, 33 Kirkland Street, Cambridge, MA 02138, U.S.A.

Pictures are inherently ambiguous but people categorize and name pictured objects with remarkable consistency. However, the time to assign a name to a picture depends on a large number of variables, ranging from the quality of the picture itself to the level of hierarchy and frequency of the name. We review the empirical results in the psychological literature on how people name pictured objects, summarizing the major variables that affect the name assigned and the time spent assigning it. The underlying regularities in these data are explained by properties of three mechanisms used in picture naming: bottom-up perceptual encoding; hierarchical associative memory; and top-down knowledge-based search. The properties ascribed to these mechanisms are hypothesized on the basis of computational analyses and considerations of characteristics of the neural systems underlying vision.

Introduction

PICTURES, ALTHOUGH SELDOM WORTH as many as a thousand words, can often serve as labels, status reports, or even instructions. There are many advantages to using pictures instead of words for these purposes; they are named faster than sequences of words are read [1], they are language-independent, and even illiterates can interpret them. However, pictures are not without drawbacks, especially their inherent ambiguity; indeed, in principle there are an infinite number of interpretations of any picture [2]. Of course, most interpretations of any given picture are wildly implausible, but even so, pictures typically can be assigned more than one name, and some pictures are more difficult to identify than others. To overcome these potential problems, designers of visual representations and languages must understand how the human brain processes pictures and the factors that influence its performance. In this article we review the psychological literature on picture naming, which began with the pioneering work of Cattell [3, 4], and discuss the factors that determine what name will be assigned to a picture and how easily a pictured object can be named in different ways. We then describe three mechanisms used in naming pictures, and show how principles that describe their processing can account for these findings.

Before going further, we must distinguish *recognition*, *identification*, and *naming*. After recognizing an observed picture, one has matched it to a stored representation of the pictured object or scene, and knows only that it is familiar. This matching process is exclusively visual, and does not result in a name or any other information. In contrast, after identifying an object, one has access to the entire range of information associated with it, including the sounds it makes, its texture, the categories to which it belongs, some specific exemplars of its type, and so on. When

---

* The authors contributed equally to this article. This work was supported by grant 88-0012 from the Air Force Office of Scientific Research, which was supplemented by funds from the Office of Naval Research.

1045-926X/90/010077 + 19 $03.00/0

one has identified an object, one has access to a multimodal set of information. Indeed, input from multiple modalities can access these memories; for example, one can identify a cat by its meow or its rub against the shins. The brain performs the unimodal processing required for recognition before the multimodal processing that constitutes identification [e.g. 5], matching visual memories before sending them to an amodal associative memory [6]. Kosslyn [7] argues that this makes sense from a computational point of view, because the problem of using inputs from multiple modalities is simplified if prior modality-specific processing produces amodal representations.

What we mean by picture naming, then, is merely a component of the identification stage, the assignment of a word (or set of words) to the picture. Since a name is one 'end product' of visual information processing, the study of picture naming enables us to draw inferences about the processing involved in both recognizing and identifying objects. Thus, we must consider factors that affect matching visual representations as well as those that affect associative memory access. We begin with factors that affect how easily a shape can be recognized, and proceed to factors that affect using this information to access the name in associative memory.

Factors that Affect Picture Naming

Tables 1, 2 and 3 summarize the findings of every article we could locate that examined picture naming or name verification. We did not consider studies of interference effects, priming effects, or special populations (such as brain-damaged patients). Table 1 lists factors that affect the difficulty of encoding and recognizing a picture, Table 2 lists factors that affect the name that is assigned to a picture, and Table 3 lists factors that affect the difficulty of accessing and/or producing the assigned name. We discuss them in turn below.

Encoding and Recognizing the Picture

Not surprisingly, the time taken to name a picture in part depends on qualities of the picture itself. The following variables have been found to affect naming times:

- *Degraded contours* Subjects require more time to name drawings when the contours have been degraded. Eliminating part of the contour is particularly disruptive when vertices are removed [8]; although removing portions of segments between vertices impairs naming, the effect is smaller than when vertices are removed.
- *Missing parts* Subjects require more time to name objects when parts have been removed [8, 9].
- *Disrupting parts* Breaking parts in the middle does not retard naming times compared to disconnecting parts at their natural boundaries. For example, cutting the blades of a pair of scissors in two at arbitrary points retards times to the same degree as disconnecting the blades from the handles. In either case, subjects name divided objects more slowly than intact objects [9].
- *Spatial relations among parts* The effect of dividing an object into parts is amplified if the parts are scrambled. Subjects have great difficulty naming objects when the spatial relations among their parts are disrupted; this effect is much larger than the effect of breaking objects into parts versus leaving them intact [9].

Table 1. Factors that affect the difficulty of encoding and recognizing a picture.

Condition	Task/results	Reference
Degraded contours:	*Producing name*:	Biederman [8], Figure 20, combined across three exposure durations (means read off graphs)
Vertices: 25% removed	773ms	
45%	812ms	
65%	910ms	
Midsegments: 25%	762ms	
45%	782ms	
65%	867ms	
Missing parts:	*Producing name*:	Biederman [8], Figure 15, nine-component objects only (means read off graph); replicated by Cave and Kosslyn [9]
Parts missing: 0	700ms	
3	775ms	
5	820ms	
6	1015ms	
Disrupting parts:	*Producing name*:	Cave and Kosslyn [9], Experiment 2 (nonsignificant difference)
Natural boundaries	1110ms	
Broken in middle	1138ms	
Spatial relations among parts:	*Producing name*:	Cave and Kosslyn [9], Experiment 2
Natural	1210ms	
Unnatural	1258ms	
Angle (from upright):	*Producing name*:	Jolicoeur [10], Experiment 1, first block trials only (means read off graphs)
0°	830ms	
60°	880ms	
120°	945ms	
180°	950ms	
240°	910ms	
300°	930ms	
Perspective Canonicalness:	*Producing name*:	Palmer *et al.* [13], Experiment 2, unprimed condition only (means read off graphs); subjects had list of names to use in advance
Low	840ms	
Medium-low	790ms	
Medium-high	745ms	
High	730ms	
Foreshortened views:	*Producing name*:	Humphrey and Jolicoeur [14], Experiment 1; long times may be due to relatively small size of pictures
45° to lateral	1167ms	
80° to lateral	1460ms	
Similarity of category members:	*Producing name*:	Humphreys *et al.* [16], Experiment 1
Similar (high)	942ms	
Distinct (low)	852ms	
Similarity between categories:	*Classification*:	Snodgrass and McCullough [18], Experiment 1
Similar (fruits/vegetables)	808ms	
Dissimiliar (animals/fruits)	628ms	
Stimulus modality	*Producing name*:	Biederman and Ju [19], Experiment 1, averaged over exposure durations (nonsignificant difference)
Line drawing	907ms	
Colour photograph	918ms	

- *Angle* Pictured objects require more time to name when whey are tilted from the standard upright. Indeed, there is a monotonic increase in naming time with the angular deviation of objects from the standard upright orientation, although there can be a dip in the curve at 180 degrees [10]. (This may be a special case, however, because a picture 'rotated' 180 degrees from upright has simply been flipped along the horizontal and vertical axes.) The general increase in time with angle is much less than that observed in typical 'mental rotation' experiments in which mental images of objects are manipulated [11]. Furthermore, the effect was significant only the first few times the pictures were seen, diminishing greatly with practice. This practice effect does not transfer to new objects; that is, even after subjects learn to name a set of objects in unfamiliar orientations, they cannot easily name novel objects in unfamiliar orientations [cf. 12].

- *Perspective* Pictures are named most quickly when seen from a 'canonical' point of view, which tends to be from a point above and to one side of the front of the object [13]. Similarly, subjects name pictures more slowly when they are more foreshortened; when more extreme perspective effects alter the shape envelope, more time is required to produce the name [14]. However, practice in naming an object from a canonical perspective transfers somewhat to naming the same object from other perspectives [15].
- *Similarity within and between categories* Objects in pictures are categorized more quickly when their shapes are distinct from those of other objects in the category [16, 17]. When subjects must assign a picture to one of two categories given to them in advance, they are faster when there is little visual similarity between the categories, as an animal/fruit judgment compared to a fruit/vegetable judgment [18].
- *Stimulus modality* Subjects take the same amount of time to name an object in a colour photograph and in a line drawing, even for objects whose colour is partially diagnostic of their identity [19].

Assigning a Name

Several factors influence what name will be assigned to a picture.

- *Typicality* More typical examples of a category are named as members of the category more often (on 69·0% *vs.* 41·0% of trials [20]) and more quickly [21] than less typical members. For example, a picture of a canary is named 'bird' more frequently and rapidly than is a picture of a swan.
- *Level of hierarchy* Subjects tend to name an object at the 'basic level' of the object's taxonomic classification system. The basic level has been characterized in a number of ways [22]. For present purposes, the most important criterion is shape overlap: The basic level is the most inclusive level at which the members of a category have very similar shapes. For example, a navel orange would be named 'orange' (the basic level) rather than 'navel orange' (the subordinate level) or 'fruit' (the superordinate level). All navel oranges have very similar shapes, but the shapes of oranges in general overlap only slightly less than those of navel oranges; in contrast, shapes of members of the category fruit (which includes bananas, grapes, kiwis, etc.) do not overlap very much. Thus, the basic level is orange.

In one experiment, the basic level was used on 78·3% of the correct trials when subjects were asked simply to provide the first name for a stimulus that came to mind, compared to 21·6% for subordinate and 0·1% for superordinate levels [20]. As is evident in Table 2, when subjects are given a name and then asked whether it is appropriate for a picture, they generally verify names fastest when the names are at the basic level [22, 23]. However, the differences in time to verify different kinds of names diminish and eventually vanish with practice; subjects eventually become equally adept at naming pictured objects at superordinate, basic, and subordinate levels [23].

- *Interactions between typicality and level of hierarchy* As is evident in Table 2, the name people assign to a picture depends on both its typicality and its level of hierarchy. In the experiment discussed above [20], subjects chose the basic level

Table 2. Factors that affect which name is assigned to a picture.

Condition	Task/results	Reference
Typicality within category:	*Producing name*:	Smith *et al.* [21], Experiment 2; replicated by Jolicoeur *et al.* [20], Experiment 3
Typical	840ms	
Atypical	858ms	
Level of hierarchy:	*Verifying name*:	Rosch *et al.* [22], Experiment 7, true trials only; replicated by Murphy and Brownell [23], Experiment 1
Superordinate	590ms	
Basic	535ms	
Subordinate	659ms	
Practice (*block* × *level*):	*Verifying name*:	Murphy and Brownell [23], Experiment 1, true trials only, name appeared for 2500ms before being joined by picture on screen
Block 1: Superordinate	833ms	
Basic	709ms	
Subordinate	803ms	
Block 2: Superordinate	760ms	
Basic	677ms	
Subordinate	709ms	
Block 3: Superordinate	750ms	
Basic	648ms	
Subordinate	593ms	
Level × *typicality*:	*Verifying name*:	Murphy and Brownell [23], Experiment 1, true trials only; typicality was for the basic level of the superordinate level
Superordinate: Typical	779ms	
Atypical	784ms	
Basic: Typical	613ms	
Atypical	744ms	
Subordinate: Typical	675ms	
Atypical	729ms	
	Producing name:	Jolicoeur *et al.* [20], Experiment 4; first four entries replicated by Smith *et al.* [21] with subjects given list of names in advance
Superordinate: Typical	1120ms	
Atypical	1290ms	
Basic: Basic typical	890ms	
Basic atypical	885ms	
Subordinate typical	880ms	
Subordinate atypical	990ms	
Subordinate: Typical	990ms	
Atypical	900ms	

86·7% of the time, *vs.* 13·1% for the subordinate level, when the object was typical; in contrast, they chose the basic level 69·0% of the time, *vs.* 31·0% for the subordinate level, when the object was atypical [20]. Although the experiment did not address this specific issue, it appears that the more atypical an object is for its category, the more likely it is to be named at a level subordinate to the basic level.

In addition, if an object is a typical exemplar of its basic-level category, it is assigned a basic-level name more quickly than are atypical objects in the category. However, if the object is an atypical exemplar, it is named fastest not at the basic level, but at a level subordinate to it. For example, a penguin is named 'penguin' more quickly than it is named 'bird' when subjects are given alternative labels in advance to use to name pictures [20].

The time to name an object at its basic or a subordinate level does not depend on how typical the object is of its superordinate category. For example, a watermelon is not a typical fruit, but it will nevertheless be named 'watermelon' as quickly as an apple will be named 'apple'. In contrast, subjects are almost always generally slowest to label an object at a superordinate level (when they are supplied with a list of names in advance and asked to use only members of the list to name pictured objects), compared to the basic or a subordinate level, but objects that are typical of a

Table 3. Factors that affect the difficulty of accessing or producing picture names.

Condition	Task/results	Reference
Name frequency:	*Producing name*:	Oldfield and Wingfield [24], categories of frequency recreated from item data; replicated by Bartram [15]
High	654ms	
Medium	737ms	
Low	1113ms	
Frequency × similarity:	*Producing name*:	Humphreys *et al.* [16], Experiment 1
High: Similar	969ms	
Distinct	800ms	
Low: Similar	916ms	
Distinct	905ms	
Familiarity of object:	*Producing name*:	Jolicoeur [10], Experiment 2, first block trials only
Familiar	1098ms	
Unfamiliar	1298ms	
	Verifying name:	Wingfield [26], Experiment 2, yes trials only, names presented auditorily before appearance of picture
Common	495ms	
Rare	508ms	

superordinate category are assigned the superordinate name faster than those that are atypical [20, 23].

In another task subjects are given a word followed by a picture, and are asked whether the word is a correct name for the picture. In this case, names of basic-level categories are compared to pictures of typical members more quickly than to pictures of atypical members. When the subjects are asked to verify a name superordinate to the basic level, the typicality of the picture at the basic level does not affect times [23]. In this case a name remained in view for 2500 ms before a picture was presented. The responses in this experiment are among the fastest in Table 2, suggesting that subjects can use the time before the picture appears to access useful information in advance.

Accessing the Name

The third important class of variables affects the time to produce the name itself.

- *Name frequency* Subjects name pictures more quickly when the name itself has a higher frequency of occurrence in the language [24, 15]. The effect of name frequency is larger if the object belongs to a heterogeneous category; if objects in a category have dissimilar shapes, times not only are generally faster (as noted above), but are especially fast for high-frequency names [16].
- *Age-of-acquisition of name* Subjects label pictures more quickly with names that are learned earlier in life. Indeed, this variable is correlated higher with naming times than is word frequency [25].
- *Familiarity* Subjects name familiar objects more quickly than unfamiliar ones [10, 26]. This effect is often small and tends to be evident only when subjects name a picture for the first time [10].
- *Nondefault naming* We have already discussed the frequency with which names at different levels of hierarchy are used to label an object. Subjects require more time to name an object if required to produce a name superordinate or subordinate to the default name [20].

Mechanisms of Picture Naming

The regularities in the data reviewed above suggest that we must consider properties of three mechanisms in order to understand how names are assigned to pictures. In order to understand naming we must first consider some essential aspects of how pictures are perceptually encoded and recognized. Such modality-specific processing is a prelude to naming itself. Next, at the heart of the process of assigning the name is an associative memory, in which input makes contact with stored information (including names). Finally, in some cases one will need to collect additional information to assign a name, which often involves top-down processing. Figure 1 outlines a theory of the information flow between these mechanisms in the brain; the detailed motivation for this particular decomposition is presented elsewhere [27].

Input to Associative Memory

Several aspects of the human visual system affect how objects are named and how easy it is to name them.

- *Object properties* vs. *spatial properties* One of the more striking discoveries in cognitive neuroscience is that object properties (such as shape and colour) are processed in different cortical systems from spatial properties (such as location and size). Object properties are analysed primarily in the temporal lobes, whereas spatial properties are analysed primarily in the parietal lobes (for reviews see [28–30]). These inferences are based on the effects of different lesions of the brain, results of recording from single neurons in the different areas, and anatomical studies that have documented distinct pathways leading from primary visual cortex in the occipital lobe to the other lobes (for a brief review, see [27]).
- *Attention window* There is more information entering the visual system than can be fully processed in real time. Attention, the selective aspect of perception, screens out all but a specific portion of the input, which is passed along for further processing. There is ample evidence that one can attend to only a single contiguous region of space at a time (e.g. [31–34]). Furthermore, there is a trade-off between

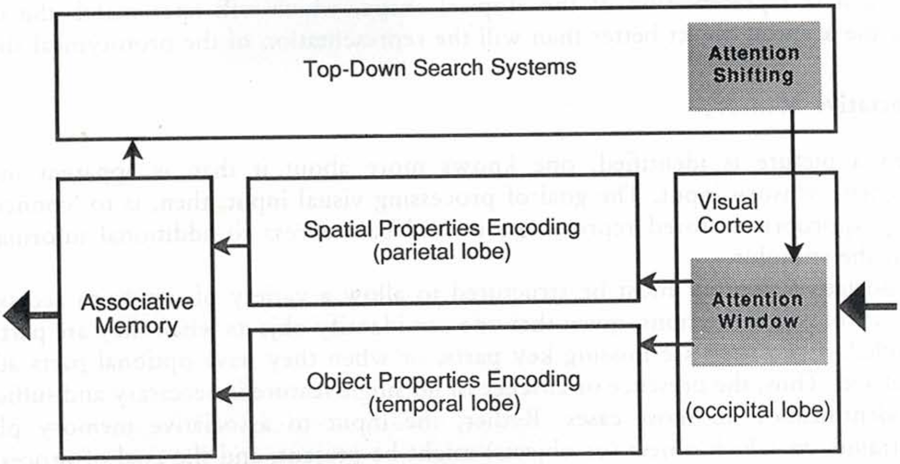

Figure 1. The major systems of high-level visual processing (adapted from [27]).

scope and resolution in attention: in general, the larger the visual angle attended to, the poorer the perceived resolution (e.g. [35–39]). This property follows if there are only a fixed number of output lines from the attention window, and they can be allocated over a small region, resulting in high resolution, or over a large region, resulting in low resolution.

Thus, if a picture is too large (subtending more than about two degrees of visual angle), one may not be able to attend to its entire shape with high resolution. In this case, multiple eye fixations will be used to examine the picture. This is important because different systems encode the location of each fixation and the shape found at the location, and hence this information must be integrated downstream in processing.

- *Constraint satisfaction in visual memory* Pictures can be identified when arbitrary portions are deleted. This was one of the phenomena that led Lowe to argue that certain aspects of shape are relatively invariant over various projections of the object [40]. These 'nonaccidental properties' include colinearity, parallel lines, points of intersection, symmetries, and so on. (These ideas have been developed further by Biederman [8].) Given such attributes, one can recognize objects by applying the 'viewpoint consistency constraint,' which states that stored representations of an object should be accessed only when the observed nonaccidental properties are consistent with seeing that object from a single point of view [41]. Although the individual nonaccidental properties only weakly constrain the recognition process, the requirement that the entire set be mutually consistent with seeing an object from a single point of view is not weak. Indeed, the constraints posed by the entire set of nonaccidental properties often may overdetermine recognition; many different subsets of the nonaccidental properties may suffice to recognize an object. Thus, deleting portions of an object will slow down recognition but not prevent it entirely.

By the same token, the object-properties encoding system can generalize over slight variations in shape. Thus, a range of different shapes can produce the same output. This generalization property is useful because members of a category often differ in minor ways. If an object's shape is very different from that of the typical members of its category (e.g. a penguin *vs.* a typical bird), however, it will not be easily subsumed by the representation of the typical shape. In such cases, we expect the system to store a new representation of the atypical shape, which will later match the input from the atypical object better than will the representation of the prototypical shape.

Associative Memory

When a picture is identified, one knows more about it than is apparent in the immediate sensory input. The goal of processing visual input, then, is to 'connect' it to the appropriate stored representations and gain access to additional information about the stimulus.

Associative memory must be structured to allow a variety of inputs to access the same set of representations, given that one can identify objects when they are partially occluded, when they are missing key parts, or when they have optional parts added or deleted. Thus, the presence or absence of no single feature is necessary and sufficient for identification in most cases. Rather, the input to associative memory places constraints on which object (or objects) might be present, and the goal of processing within associative memory is to find the representation that best satisfies those constraints. The system attempts to 'decide' what object is present by converging on the stored representation that is most consistent with the inputs to associative memory. This kind of constraint-satisfaction process can be accomplished efficiently in parallel distributed processing systems (e.g. [42, 43]) The system must seek confirmatory evidence, in the form of input that satisfies constraints, and at least in some cases may also have to disconfirm hypotheses by finding properties that are inconsistent with one or more of the competing stored representations.

- *Constraint satisfaction in associative memory* A simple implementation of such a constraint-satisfaction system for visual object recognition was proposed by Feldman [44]. Each object known to the system is represented as a node in a network of mutually inhibitory processing units. Each such 'object unit' is excited by a set of 'feature units' representing the properties or visual features associated with the object, so that the object unit becomes more activated when more of its associated properties are present. As a unit becomes more activated, its inhibition of other object units increases, yielding a 'winner take all' mechanism that results in only one object unit eventually being fully active. This system of local computation with connections among competing representations can perform the entire decision-making job without an intelligent agent overseeing the process.

Wittgenstein noted that categories seldom have sharply delineated boundaries, defined by sets of necessary and sufficient features; rather, category members are more often related by 'family resemblance' with each possessing some number of an overall set of features that can apply to the category [45]. Therefore, many exemplars of a class of objects have additional optional parts or are missing certain parts; for instance, chairs may or may not have arms and dogs may or may not have tails. Since the overall shape of the object depends on the parts that are present, proper generalizations from examplar to category cannot be made simply by generalizing to similar shapes (i.e. by treating as equivalent all shapes that fall within some range of variation from a prototype shape).

One way to solve this generalization problem is to assign an object to a category when a subset of that category's associated properties have been matched, allowing different subsets to be used in different situations. Each property of the input becomes a piece of evidence, with object identification occurring once enough evidence has accumulated, regardless of which specific pieces of evidence have been found. For example, the presence of at least three of the properties of a typical chair might be necessary to identify an input object as a chair; the sets {back, seat, legs}, {arms, seat, legs}, and {arms, cushion, back} would each provide sufficient evidence for a 'chair' classification. Unfortunately, the solution cannot be this simple because different properties are more or less diagnostic of an object's identity. For example, because a stool also has a seat and legs, finding a seat and a back would be much more important in identifying a chair than finding a seat and legs. Hence, we need a more sophisticated mechanism to solve the generalization problem.

Proper generalization can be made if we add to the connection between each feature unit and object unit a 'distinctiveness weight', which represents how distinctive the particular feature is, and therefore how diagnostic the feature is in identifying the object. Every time an input property matches a stored property, the weight associated with that property is added to total accumulating for each object. An object is considered identified when its accumulated weight exceeds a threshold, regardless of which features were activated and contributed their distinctiveness weights. Thus, a variety of different property combinations can provide sufficient total weight to identify a particular object. For example, suppose that the 'chair' unit has different weights for these feature units: 'back' (very important, very high weight), 'seat' (important, high weight), 'legs' (relatively important, medium weight), 'arms' (not important, low weight). The system could identify a wide variety of chairs if the threshold were set to be exceeded by the activation of 'back' and one of the other properties, or perhaps by the activation of all the properties but 'seat' (cf. [46]).

The correct weight for each property-unit-to-object-unit connection can be set by increasing it (but not necessarily by a uniform amount) every time the property is used to diagnose the presence of the object. Many learning algorithms explored for parallel distributed processing networks, including versions of the simple Hebb rule [47], can accomplish this task easily. Since an object's distinctive properties are defined as those that distinguish it from other similar objects, they will be used disproportionately often to identify the object, and hence the weights between their units and the object unit will become relatively large. (To avoid the potential problem of all weights continually increasing, we can assign a total weight limit to each object unit and require that the sum of the weights of the connections from all property units to the object unit not exceed that limit; the limit might be determined by the availability of fixed quantities of neural resources. In learning, then, weight would be redistributed among connections rather than simply added to all.)

Finally, note that the amount of evidence from the input required for object identification can differ in various situations. For example, a farmer gathering his cows at dusk can identify them from no more sensory input than passing shadows seen out of the corner of his eye. We can account for such abilities by allowing object unit thresholds in associative memory to be altered by context, with different amounts of information treated as criterial in different circumstances.

We can also understand another ability if thresholds can be adjusted by context. One can name objects at a requested level of hierarchy, e.g. one can name an individual from a picture of his face in one context, and name the object as 'face' in another. If thresholds can be dynamically altered, one need only raise them for certain object units to inhibit specific stimulus classifications (such as 'face' and 'man') in specific contexts. Therefore, we posit that associative memory incorporates a mechanism to inhibit activation of certain types of representations, depending on the task at hand. It follows that there must be some way to index those classes of representations. This index could be explicit, such as a property for each representation indicating its level of specificity, implicit, such a count of its superordinates in the hierarchy, or both.

• *Level of hierarchy in associative memory* One reason objects should be identified first at the basic level is because the process of doing so will activate the maximum number of properties that probably apply to the stimulus, giving immediate access to as much implicit information as possible. However, the basic level does not have this virtue for objects that are atypical of their categories. For example, initially categorizing a penguin as a bird would not be helpful, because it would activate many of the properties of birds that do not apply to penguins—they do not sing and do not fly. Furthermore, it would fail to activate many properties of penguins that do not generally apply to birds—they have adaptations for underwater swimming, and so on.

In such cases, when an object's overall shape is very different from that of the typical member of its category, we expect the input shape to match the representation of that particular shape better than the representation of the typical shape. Indeed, these outliers often also have unusual sizes and orientations, which also enable the spatial properties encoding system to provide distinctive information. For atypical objects, then, associative memory may receive different inputs from both the object and spatial properties encoding systems than it receives for typical objects.

Thus, people often name typical objects at a basic level, but name atypical ones at a level subordinate to the basic level [20]. Subjects tend to name an ostrich as an ostrich, but a sparrow as a bird, or a Porsche as a Porsche, but a Chevrolet as a car, and so on. The level of hierarchy at which a representation provides the maximum likely properties is called the *entry level* [20].

We expect the entry level representation to be the first to be activated because of two properties of associative memory. First, object units that have more associated properties that match those in the input, and fewer that do not match those in the input, will be better activated, and hence will more successfully inhibit representations of competing objects in associative memory. Thus, the system will be driven to activate representations associated with more of the properties in the input. This principle tends to drive the system to activate more specific, subordinate representations; this process is limited, however, by the specificity of the input reaching associative memory. Second, an intermediate (basic) level representation will be activated for typical objects because the output from visual memory is not sensitive to the relatively small variations among instances of typical objects—differences in shape are generalized away in the object properties encoding system and are not registered by the spatial properties encoding system. Thus, for typical objects, the initial input to associative memory will not contain enough information to identify very specific representations. Additional information must be collected for the system to encode subtle shape variations (e.g. by encoding individual parts at higher resolution), and hence more time will be required (as will be discussed shortly). Atypical objects, then, tend to be categorized initially at a more specific level because their object and spatial properties are sufficiently different from those of the typical members of their categories; thus, they activate different, more specific sets of feature units in associative memory.

Top-down Hypothesis Testing

If the overall shape of an object closely matches a pattern stored in visual memory, a single object representation in associative memory may be activated beyond its threshold. In this case, the object can be identified without consideration of its spatial properties. This exclusively bottom-up processing is likely to succeed only in certain circumstances, namely when an object has a very distinctive shape or when context biases the observer towards seeing a specific object. Even in such cases, it will not succeed if the object is so large (subtending more than about two degrees of visual angle) that it does not fall entirely on the retina's high-resolution fovea, and therefore requires multiple eye fixations to examine its entire shape. Whenever direct recognition is impossible, object properties and spatial properties will have to be integrated in associative memory, and attention shifted to new locations in space to focus on other parts of the object.

Two kinds of information can be used to direct attention. The first of these, properties of the stimulus, can automatically draw one's attention to a specific location. Sudden changes in the input, such as movement or intensity variations, are especially likely to have this effect [cf. 48]. Such bottom-up, stimulus-driven shifts of attention are especially apparent in the behavior of very young infants (e.g. [49]). The second, knowledge, belief, and expectation about yet-unobserved aspects of the stimulus, can direct attention to a specific location to search for anticipated properties. Such top-down influences on attention are apparent in the patterns of eye movements recorded during object identification tasks (e.g. [50]). People focus on the parts of pictures with the highest information content, systematically shifting among them [51]; for example, when examining a face, subjects target its eyes, nose, and mouth for the vast majority of their eye fixations.

Until one representation is activated above its threshold, additional object and spatial properties need to be encoded and combined in associative memory, so additional properties of the hypothesized object will be sought in the input. The object unit with the greatest initial activation represents this 'candidate' object, and the difference between its activation and those of its competitors dictates how easy it will be to detect. The distinctiveness weights for the object's properties can be used to guide additional information-gathering, with the search proceeding from the most distinctive unobserved property down the list until identification occurs or there are no more properties worth looking for. (A theory of how such processing may occur is developed elsewhere [27], so we will not reiterate the details here.)

Accounting for the Results

We now return to the data summarized in Tables 1, 2, and 3, and use the theoretical ideas discussed above to sketch an account for each class of findings.

Encoding and Recognizing the Picture

These manipulations affect naming primarily because they affect the recognition process in the object properties encoding system.

- *Degraded contours* Degrading the contours disrupts matching the input to visual memories; vertices typically are 'nonaccidental properties', and hence contribute especially strongly to the constraints that are used to recognize the input. Thus, removing them has deleterious effects (cf. [52]).
- *Missing parts* More time will be required to converge on a single object representation in associative memory when parts have been removed, because fewer feature units will be activated by the stimulus.
- *Disrupting parts* The viewpoint consistency constraint depends on a proper spatial arrangement among key features of shapes, and hence breaking up parts (for example, exploding a slice out of a pie chart) will impair recognition.
- *Spatial relations among parts* Scrambling the positions of parts disrupts the overall shape of an object, and hence the initial input will not match a stored representation in visual memory very well. Thus, top-down search mechanisms will be brought into play to collect additional information; however, when the locations of the parts have been scrambled, it will be difficult to locate distinctive parts using

stored information about their usual locations. Hence, the top-down search process will be impaired and more time will be required to encode the information necessary to identify the object.

- *Angle* Novel pictures generally require more time to name when the pictured objects are tilted from their standard upright position. The transient nature of this effect reflects the operation of the viewpoint consistency constraint. Once a shape representation has been encoded into visual memory, the input can be matched to it regardless of the object's orientation. However, when a picture is unfamiliar, it will not match a stored representation very well, and hence a high-confidence match cannot be made in visual memory. It is in this situation that top-down processing is used. Such processing will not be optimal when an object is in an unusual orientation because it will be difficult to locate distinctive parts. Hence, more time will be required to name novel objects in nonstandard orientations.

The increased times with greater tilt can be understood if attention must be shifted across the visual input to find sought parts, and more time is required when more scanning is needed. The farther an object is tilted from the upright, the greater the distance will be from the expected to the actual location of a sought part. Since this process is circumvented only by encoding a new image into visual memory, there will be no transfer of practice to dissimilar new objects (which have not been encoded into visual memory).

- *Perspective* A canonical view of an object is named fastest for several reasons. It not only matches the stored visual memory best, but also is the view in which the most parts are visible. Thus, if top-down search is necessary, more parts will be found efficiently. In contrast, when an object is foreshortened, parts are obscured and it is more difficult to search top-down for distinctive parts. Practice with a canonical view of an object transfers to other views [15] because it creates a near-optimal shape representation. The canonical perspective depicts the most features, and hence will match a novel input on average better than any other representation.
- *Similarity of category members* Subjects name objects in pictures more quickly when the objects have shapes that are distinct from those of other objects in the category. This result can be understood simply because a distinctive shape is likely to be represented separately in visual memory; when a new shape is encountered, its representation will be stored if it is not very similar to any previously stored representation. Hence, a distinctive shape is likely to match a stored representation very well, and produce a high-confidence output from visual memory. On the other hand, if a shape is similar to others encoded into visual memory, it will match more than one stored representation reasonably well; thus the output from visual memory will be a low-confidence match, and additional information will need to be collected.
- *Similarity between categories* When subjects indicated which of two previously given categories applied to a picture [18], they could use top-down processing to search for distinctive parts because there were only a few alternatives. If the objects within each category are more similar to one another than to the objects in the other category, it will be easier to locate a stored representation (in associative memory) of a part that applies to the members of one category but not the other. In contrast, if the objects in both categories all have similar shapes, it may be necessary to search for several parts before identifying the stimulus. Hence, in this particular situation, subjects can choose faster between two alternative categories for an object if the dissimilarity in shape is greater between than within categories.

- *Stimulus modality* Line drawings simplify the problem of finding edges, and make explicit the nonaccidental properties from the start. Hence, we are not surprised that colour photographs are not named more quickly than line drawings. But why are they not named more slowly? Because photographs include additional properties (primarily colour and texture), the initial input to associative memory may contain more properties. These additional constraints on the appropriate representation may compensate for the additional time to extract the nonaccidental properties from photographs.

Assigning a Name

As discussed earlier, the name assigned to a picture depends primarily on how it is recognized (matched in visual memory), on the output from the dorsal spatial properties system, on the organization of associative memory, and on the possible need to collect additional properties after the initial ones (encoded during the first pass) have been entered into associative memory. (It also depends on context, but we will not discuss such effects further here.)

- *Typicality* Input shapes are matched to similar shapes stored in visual memory. If a previously seen picture is very distinctive, it is likely to be stored as a separate representation—because it is distinctive, when first seen it did not match any other stored representations very well, and hence a new representation was created. In contrast, if an object is typical of a category, it is likely not to be stored as a separate exemplar—its image already corresponds to at least one stored representation, and hence a new representation need not be created when it is seen.

Thus, a typical shape will tend to match a stored visual representation that will later be labeled by its category name. Atypical shapes, in contrast, may match a representation of an exemplar better than the representation of the typical shape. If the match is best to an exemplar (e.g. penguin), then additional processing will be required in associative memory to produce the name of the category (e.g. bird). Therefore, typical members of a category will be assigned the category name faster than atypical members.

- *Level of hierarchy* The constraint satisfaction mechanism in associative memory usually will select basic level representations, because those object units will be most activated by the input properties encoded on the first pass (these objects have the most properties that match the input and the fewest that are not initially encoded). If one spends more time studying the object, so that more properties are encoded, then a subordinate level term will best classify the input.

People require more time to name objects at a superordinate level. This result makes sense because the constraint-satisfaction mechanism in associative memory produces a single name for an object at the entry level. Thus, a name at a superordinate level must be inferred (by searching a hierarchical associative memory) after the object is identified, and this process requires additional time. For example, if a stimulus is initially identified as a dog, additional processing will be necessary to infer that it is also an animal.

The role of top-down processing explains the results of the verification task in which the names paired with objects were either above or below the object's assumed entry level name in the hierarchy [20]. For example, a picture of an apple could be paired with the name 'fruit' or 'delicious apple'. If the picture is named at the entry level (e.g. 'apple'), as expected, then its superordinate name can be verified merely by searching associative memory. However, the subordinate name can only be verified by collecting more information about the object. If, when asked whether an object is a delicious apple, one first spontaneously names it 'apple,' one must go back and look carefully at its aspect ratio and dimpled bottom (using knowledge about the shape of a delicious apple to guide the search). Only after such additional information is collected can one verify that it is indeed a delicious and not a MacIntosh or some other kind of apple.

If this is what happens, then presenting a random pattern immediately after the picture, which masks (wipes out) the iconic representation of the visual input, should have very different effects when different kinds of names are being verified. If the picture is masked when the name is a superordinate, this will not disrupt processing in associative memory. Once the entry level representation has been activated over threshold, superordinate category names can be accessed directly (if they are associated with the object) or inferred (by some other system accessing associative memory). However, if a picture is masked when the name is subordinate to the entry level, this should greatly disrupt the additional top-down encoding cycles. For example, if one names a canary as a bird, and has to decide whether it is a canary, one needs to check whether it is yellow, of a certain size, and so on. If the picture is masked, less information can be encoded and so this more specific categorization should be difficult. This prediction was confirmed [20]; masking a picture did indeed have very different effects depending on the to-be-verified name and the level of the entry name.

The dissipation of the basic-level advantage with practice [23] suggests that over time subjects become familiar with the properties that distinguish the pictures in the particular set used as stimuli in an experiment. Hence, they can use top–down search immediately to look for distinctive properties, obviating the need to wait for initial object and spatial properties to narrow down the search space. Such a strategy would disrupt the usual pattern of results because some objects have features that distinguish them at a subordinate level, whereas others do not.

- *Interactions between typicality and level of hierarchy* We have already discussed the concept of the entry level term, which accounts for why atypical objects are named fastest not at the basic level, but at a level subordinate to it. What about naming at a level superordinate to the entry level name? When an object must be named at a superordinate level, more time is taken when the object is atypical for that category. It is important to note that the same result is obtained when a word is provided as input instead of a picture [20]. This finding suggests that the effect has nothing to do with encoding or searching for visual properties, but rather, that the picture is identified at the entry level, and an inference must be drawn from there. If the object is atypical for members of that category, it is less likely to be directly associated with the category in memory (either by having the superordinate listed as a property of the object itself, or by having the object listed as an example of the superordinate), and hence more time will be required to derive its superordinate name.

In another task, subjects are given a name and then a picture, and asked whether the name is appropriate. In this task, the typicality of the pictured object at the basic level does not affect the time to name it at a superodinate level [23]. At first glance, this result is surprising: If subjects first name an object at the basic level, and then infer the superordinate, typical objects should be named faster. However, this result is not so surprising in light of the finding that subjects tend to name typical objects at the basic level, and atypical ones at a subordinate level. A superordinate term may be associated directly with an entry level name regardless of whether the entry level is basic or subordinate; similarly, the entry level name itself may be directly associated with the superordinate term. Therefore, in either case (typical or atypical), only one inference will be required to access the superordinate name.

Accessing the Name

- *Name frequency* Subjects name objects more quickly when the label is a word that is used frequently in the language. We assume that this effect reflects the ease of accessing the representation of the name. When the word is used more often, it comes to have stronger weights on its connections to the object unit in a parallel distributed processing network. (Alternatively, one can think of this as moving the representation of the name higher on the property list associated with the object.) In addition, subjects use high-frequency names especially quickly when an object is in a category whose members have dissimilar shapes. These especially fast times reflect more than a faster match in visual memory for distinctive shapes and a more accessible name in associative memory. If these two variables were all that were at work, the effects of similarity and frequency would add, but they do not—subjects are especially fast when using high-frequency words to name objects with distinctive shapes [16]. We can understand this result if these distinctive shapes also have distinctive spatial properties (e.g. size). If so, then more of the properties initially entering associative memory will be distinctive (having high weights), and fewer properties will need to be considered before processing in associative memory converges on a single representation; consequently, naming times will be particularly fast.
- *Age-of-acquisition of name* The mechanism that accounts for the effect of word frequency will also account for the effect of learning a word early in life. If a word is entered into associative memory earlier in life, on average it will probably have been accessed more often than one entered later in life, not only because the word is particularly useful but also simply because the person has more opportunities to use it. Hence, it will have a stronger weight on its connection to the object representation (or be higher on a property list), and will be accessed more readily than words learned later in life.
- *Familiarity* Familiar objects are named more quickly than unfamiliar ones. At first glance, we could attempt to explain this finding by assuming that familiar objects have a better representation in visual memory and/or that their names are more accessible in associative memory. We rejected the second possibility for two reasons. First, more familiar objects are named fastest the first time a picture is named in an experiment, but not thereafter. This effect is analogous to that of orientation on naming time, and hence can be explained by appeal to the same principles. We explain the effects of orientation on naming time in part by noting that if the representations in visual memory are not specific to the object, a perfect match will not be made, and hence top-down search will be used. Once the subject sees particular stimuli used in

the task, either they can be explicitly represented in visual memory or distinctive features can be noted and used to direct top-down search. In either case, the effects of familiarity disappear with practice. Second, we argued that name frequency affects times because the representation of the name itself becomes more accessible with repeated use. This effect is more robust than the effect of familiarity, which would not be expected if the same mechanism caused both effects.

- *Nondefault naming* If the entry level name is not appropriate for the task at hand, then additional processing will be required. If the appropriate name is a superordinate to the entry level name, then it must be inferred by searching associative memory; if it is a subordinate, then additional information must be collected. Both of these cases have been discussed above, and such additional processing clearly will increase naming time.

Conclusions

Naming a picture is a complex activity, which is influenced by many factors. The effects of these factors can be understood by considering the properties of three classes of mechanisms, those used in bottom-up processing of the visual input, in organizing and accessing associative memory, and in applying stored knowledge to search top-down for distinctive properties of objects. Although we have had to make numerous assumptions about these mechanisms to account for the data, the assumptions are based on computational analyses and neural constraints [27] and, perhaps most important, suggest clear new lines of empirical research to detail the separate roles and interactions of the mechanisms.

Although the differences in times presented in Tables 1, 2 and 3 may not appear very large, the statistically reliable ones (discussed in the review) index differences in how much processing is necessary to identify a pictured object. When a viewer must work harder to identify an object, he or she cannot work as well at a concurrent task [53]. Since such concurrent tasks often lead the viewer to want to name a picture in the first place, it is important that a viewer be able to interpret a picture as accurately, easily, and rapidly as possible.

Thus, the results summarized here are important for human factors engineering. They indicate what factors lead a viewer to name a picture in a particular way, and what kinds of processing are necessary if a nondefault name is required. In short, if a picture is to be easily named when it is initially seen, it should be drawn with intact contours, have as many distinctive parts visible as possible, have parts in their typical locations, and be presented upright or in a canonical orientation. In addition, if the picture is used to label a general category, and hence the name of the category should be evoked by it, then the pictured object should have a shape similar to other members of its basic-level category; in contrast, if an exemplar name is required, the object should have a distinctive shape. The shape should be familiar, and the name of the object should occur often in the language.

The factors that influence the ease with which a viewer will connect a picture with an idea should be kept in mind when images for visual databases are drawn, graphical user interfaces are designed, and icons, online help, hypertext systems, and instruction manuals are created. In addition, the mechanisms we describe to account for the results may be applied in building artificial visual systems that mimic human performance. These mechanisms will prove useful not only because human beings will interact easily with systems that mirror their abilities, but also because the remarkable power of the human visual system is clearly worth emulating in its own right.

References

1. M. C. Potter & B. A. Faulconer (1975) Time to understand pictures and words. *Nature* **253,** 427–438.
2. N. Goodman (1968) *Languages of Art.* Bobbs-Merill, Indianapolis.
3. J. M. Cattell (1886) The time it takes to see and name objects. *Mind* **11,** 63–65.
4. J. M. Cattell (1886) The time taken up by cerebral operations (in three parts), *Mind* **11,** 220–242, 377–387, 524–538.
5. M. Mishkin & T. Appenzeller (1987) The anatomy of memory. *Scientific American* **256,** 80–89.
6. L. R. Squire (1987) *Memory and Brain.* Oxford University Press, New York.
7. S. M. Kosslyn (in preparation) *Networks of High-Level Vision.* MIT Press, Cambridge, MA.
8. I. Biederman (1987) Recognition-by-components: A theory of human image understanding. *Psychological Review* **94,** 115–147.
9. C. B. Cave & S. M. Kosslyn (1990) The role of parts and spatial relations in picture naming, Harvard University manuscript.
10. P. Jolicoeur (1985) The time to name disoriented natural objects. *Memory and Cognition* **13,** 289–303.
11. L. A. Cooper & R. N. Shepard (1973) Chronometric studies of the rotation of mental images. In: *Visual Information Processing* (W. G. Chase, ed.), Academic Press, New York, pp. 75–176.
12. M. J. Tarr & S. Pinker (1989) Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology* **21,** 233–282.
13. S. Palmer, E. Rosch & P. Chase (1981) Canonical perspective and the perception of objects. In: *Attention and Performance IX* (J. Long & A. Baddeley, eds) Erlbaum, Hillsdale, NJ, pp. 135–151.
14. G. K. Humphrey & P. Jolicoeur (1988) Visual object identification: Some effects of image foreshortening and monocular depth cues. In: *Computational Processes in Human Vision* (Z. W. Pylyshyn, ed.) Ablex, Norwood, NJ, pp. 429–442.
15. D. Bartram (1974) The role of visual and semantic codes in picture naming. *Cognitive Psychology* **6,** 325–356.
16. G. W. Humphreys, M. J. Riddoch & P. T. Quinlan (1988) Cascade processes in picture identification. *Cognitive Neuropsychology* **5,** 67–104.
17. M. J. Riddoch & G. W. Humphreys (1987) Picture naming. In: *Visual Object Processing: A Cognitive Neuropsychological Approach* (G. W. Humphreys & M. J. Riddoch, eds) Erlbaum, Hillsdale, NJ, pp. 107–143.
18. J. G. Snodgrass & B. McCullough (1986) The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition* **12,** 147–154.
19. I. Biederman & G. Ju (1988) Surface *vs.* edge-based determinants of visual recognition. *Cognitive Psychology* **20,** 38–64.
20. P. Jolicoeur, M. Gluck & S. M. Kosslyn (1984) Pictures and names: Making the connection. *Cognitive Psychology* **16,** 243–275.
21. E. E. Smith, G. J. Balzano & J. Walker (1978) Nominal, perceptual, and semantic codes in picture categorization. In: *Semantic Factors in Cognition* (J. W. Cotton & R. L. Klatzky, eds) Erlbaum, Hillsdale, NJ, pp. 137–168.
22. E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson & P. Boyes-Braem (1976) Basic objects in natural categories. *Cognitive Psychology* **8,** 382–439.
23. G. L. Murphy & H. H. Brownell (1985) Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **11,** 70–84.

24. R. C. Oldfield & A. Wingfield (1965) Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* **17**, 273–281.
25. J. B. Caroll and M. N. White (1973) Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology* **25**, 85–95.
26. A. Wingfield (1968) Effects of frequency on identification and naming of objects. *The American Journal of Psychology* **81**, 226–234.
27. S. M. Kosslyn, R. A. Flynn, J. B. Amsterdam & G. Wang (in press) Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition.*
28. J. H. R. Manusell and W. T. Newsome (1987) Visual processing in monkey extrastriate cortex *Annual Review of Neuroscience* **10**, 363–401.
29. M. Mishkin, L. G. Ungerleider & K. A. Macko (1983) Object vision and spatial vision: Two cortical pathways *Trends in Neurosciences* **6**, 414–417.
30. L. G. Ungeleider & M. Mishkin (1982) Two cortical visual systems, In: *Analysis of Visual Behavior* (D. J. Ingle, M. A. Goodale & R. J. W. Mansfield, eds) MIT Press, Cambridge, MA, pp. 549–586.
31. C. J. Downing & S. Pinker (1985) The spatial structure of visual attention, In: *Attention and Performance XI* (M. I. Posner & O. S. I. Marin, eds), Erlbaum, Hillsdale, NJ, pp. 171–187.
32. D. LaBerge (1983) Spatial extent of attention to letters in words *Journal of Experimental Psychology: Human Perception and Performance* **9**, 371–379.
33. M. I. Posner, C. R. R. Snyder & B. J. Davidson, Attention and the detection of signals. *Journal of Experimental Psychology: General* **109**, 160–174.
34. A. M. Treisman & G. Gelade (1980) A feature-integration theory of attention *Cognitive Psychology* **12**, 97–136.
35. H. Egeth (1977) Attention and preattention. In: *The Psychology of Learning and Motivation 11* (G. H. Bower, ed.) Academic Press, New York, pp. 277–320.
36. C. W. Eriksen & J. D. St. James (1986) Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics* **40**, 225–240.
37. J. Jonides (1983) Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society* **21**, 247–250.
38. D. Marr (1982) *Vision* Freeman, San Francisco.
39. G. L. Shulman and J. Wilson (1987) Spatial frequency and selective attention to local and global information *Perception* **16**, 89–101.
40. D. G. Lowe (1987) Three-dimensional object recognition from single two-dimensional images *Artificial Intelligence* **31**, 355–395.
41. D. G. Lowe (1987) The viewpoint consistency constraint *International Journal of Computer Vision* **1**, 57–72.
42. D. E. Rumelhart & J. L. McClelland (1986) *Parallel Distribution Processing: Explorations in the Microstructure of Cognition (Volume 1: Foundations)* MIT Press, Cambridge, MA.
43. J. L. McClelland & D. E. Rumelhart (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Volume 2: Biological and Psychological Models)* MIT Press, Cambridge, MA.
44. J. A. Feldman (1985) Four frames suffice: A provisional model of vision and space *Behavioral and Brain Sciences* **8**, 265–289.
45. L. Wittgenstein (1953) *Philosophical Investigations* Macmillan, New York.
46. E. E. Smith & D. Medin (1983) *Categories and Concepts* Harvard University Press, Cambridge, MA.
47. D. O. Hebb (1949) *The Organization of Behavior* Wiley, New York.
48. U. Neisser (1969) *Cognitive Psychology* Appleton-Century-Crofts, New York.
49. T. Bower (1970) *Perceptual Development* Freeman, San Francisco.
50. A. L. Yarbus (1967) *Eye Movements and Vision* Plenum Press, New York.
51. G. R. Loftus (1972) Eye fixations and recognition memory for pictures *Cognitive Psychology* **3**, 525–551.
52. F. Attneave (1954) Some informational aspects of visual perception *Psychological Review* **61**, 183–193.
53. M. I. Posner (1978) *Chronometric Explorations of Mind* Oxford University Press, New York.