Holistic and Analytic Representations

of Ignored and Attended Objects

Thesis presented by

Volker Thoma

Department of Psychology

Goldsmiths University of London

November 2002

In Partial Fulfillment for the Degree of Doctor of Philosophy

Acknowledgement

I am indebted to Jules Davidoff who closely supervised my thesis and never let me get distracted. His direction, patience, and impatience were invaluable for this research project and the learning process associated with it.

I am also grateful to John Hummel for welcoming me in his lab and introducing me to his model of object recognition. His enthusiasm for research was and is contagious.

I further thank the cognitive research group at Goldsmiths' for their comments and even more so for their encouragement, in particular Alan Richardson-Klavehn, Linda Pring, Tim Valentine, and Steve Darling.

Research needs infrastructure, and I was lucky to find the technical support I had at Goldsmiths', in particular provided by Maurice Douglas, Liz Elliott, Ian Hannent, Rob Davis, and Richard Billingham.

And of course, I thank my participants for their time.

Finally, I thank my friends Dani, Ulrich, Jonathan and Jane. Without them, I would have been lost.

Abstract

Attended images prime both themselves and their left-right reflections, whereas ignored images prime themselves but not their reflections (Stankiewicz, Hummel, & Cooper, 1998). These and other effects are predicted by the hybrid theory of object recognition (Hummel & Stankiewicz, 1996a) that the human visual system represents ignored images holistically (i.e., view-based), and attended images both holistically and analytically (i.e., part-based). In nine experiments using a naming task the predictions of the model were tested with split, plane-rotated and depth-rotated views of common objects.

Consistent with the prediction of the hybrid theory, Experiments 1 and 3 demonstrated that split images primed their intact and split counterparts when they were attended but not when they were ignored, whereas intact images primed themselves whether they were attended or not. Experiment 2 demonstrated that a substantial component of the observed priming for attended split images was specifically visual. In Experiment 5, attended images primed themselves and their plane-rotated versions (90°) whereas ignored images only primed themselves but not their rotated versions. Experiment 6 tested whether rotated objects with a definite upright orientation prime themselves in the same view. Substantial priming was observed for attended and ignored objects when shown in their upright view. However, rotated objects with a definite upright orientation primed themselves only when attended but not when ignored. This result indicates that ignored images make contact with stored representations.

Experiment 7 replicated the findings of Stankiewicz et al. for mirror images but with grey-level rendered 3D images. Experiment 8 tested priming for these objects using orientations in which parts change from study to test view. As before, there was substantial priming in all but the ignored-rotated condition. However, there was a greater reduction in priming for attended rotated objects than for ignored rotated objects. This result indicates that the representations mediating recognition of attended images are specifically sensitive to part changes. In Experiment 9, objects were rotated in depth such that equivalent parts were visible in both views. As in Experiment 7, the priming effects of view and attention were additive.

These data provide strong evidence that one function of visual attention is to permit the generation of analytic (i.e., part-based) representations of object shape. At the same time these results show that object recognition is also mediated by additional holistic representations.

Contents

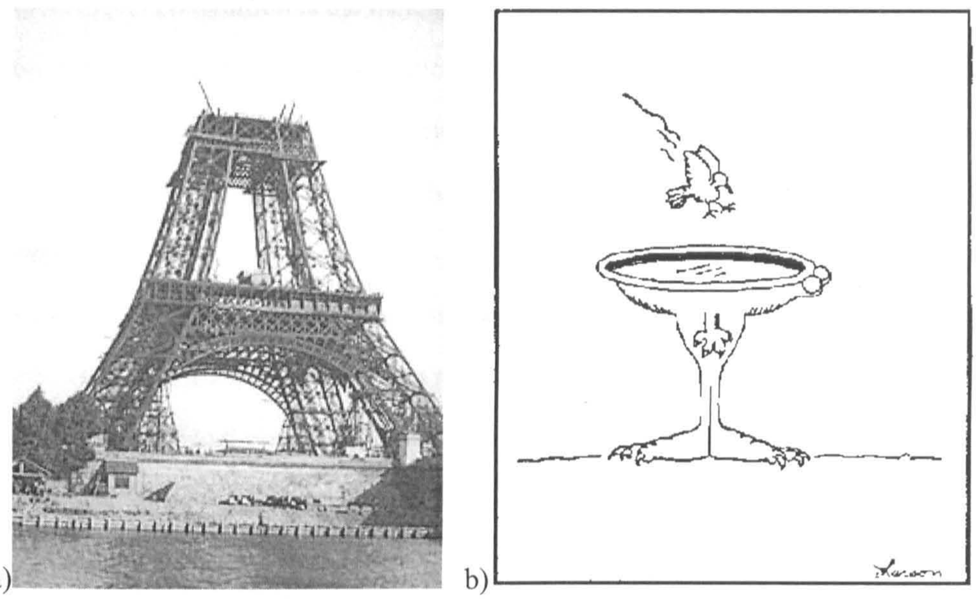
List of Figures

List of Tables

1. Chapter 1: Introduction

1.1 The Basic Problem and Scope of the Thesis

When we see an object such as a car the basic information we receive is the light reflected from it, which is absorbed by cells at the back of the eyes, the retina. But of course seeing does not automatically mean recognising or knowing what an object is, because recognition sometimes fails or requires more effort (see Figure 1).



a) b)

Figure 1: Even familiar objects are not always immediately recognised (a, construction of the Eiffel tower), or they are not what they appear to be (b, Gary Larson cartoon).

Figure 1a demonstrates that a very familiar object can be recognised even if it is seen only partly, but this partial representation is at first surprising (because most of us have never seen it that way). In Figure 1b a completely new object is classified as predator, but again only after an unusual effortful (although still swift) recognition process. Part of the fun in the cartoon lies in the fact that most observers are probably fooled (like the bird) by first seeing a water fountain instead of a monster because of its overall shape. Yet soon after noticing some unexpected differences concerning smaller parts of the "water fountain" the re-recognition process reveals the animal. Very simplified, the investigation presented in this thesis is about how we recognise objects both in an automatic effortless process as familiar whole shapes but also need attention to analyse their part structure.

In general, objects can be recognised not only by their shape, but also based on other visual cues, such as colour, texture, characteristic motion, context information, and also expectation. This research focuses on the recognition of isolated objects, using shape information alone and it will consider primarily object identification rather than classification. Questions concerning low-level vision processes will be neglected for the most part. The focus of this study is the single object and how its shape is recognised depending on attention and changes in orientation and configuration.

Visual object recognition is the matching of retinal images to representations of objects stored in memory. These representations are thought of as reconstructions of the information in the input array. However, at least two basic properties of object recognition make it difficult to understand the nature of the representations and the matching process. The first property is object constancy across variations in viewpoint. Although the retinal image of a particular object surely varies with changes in orientation, scale, illumination, pose or location in the visual field, observers readily recognise that object in most circumstances. Therefore, the visual system must have ways of generalising across variations of the 2D input from the retina. The second property is object constancy across variations in shape; not only do we generalise across different images of the same object, but also across different instances of a class of objects (Rosch, Mervis, Gray, & Boyes-Braem, 1976). For example, even though there is not much visual similarity between a nineteenth century Bell-telephone and a modern cordless handset, both are readily identified as "phone".

There is an ongoing debate in the literature of how the visual system achieves both viewpoint and shape invariance in object recognition. Most theories of shape recognition propose a set of representations of objects stored in long-term memory. These object models are canonical representations of the object's shape, describing some invariant features of the object. When we see a shape, it is converted into the same format in which the long-term memory representations are stored. Subsequent recognition is the successful match between input and stored representation. Theories differ according to the nature of the representation's format, the number of representations for a single object, and which class of objects will be mapped onto a single representation. This often means that a theory can be powerful to explain one form of object constancy but not the other.

In the first part of Chapter 1, the neural basis of object recognition will be described briefly, as it is the basis for many object recognition theories and models. A brief review of the properties of object recognition is followed by a description of major object recognition theories and how they can account for the two forms of object constancies. Chapter 2 focuses on the hybrid model of object recognition (Hummel, 2001) and its predictions. In the main part of the thesis these predictions are put to test in a series of experiments. The results and their implications for object recognition will be evaluated and discussed in the final part of the dissertation.

1.2 Object Recognition in the Brain

1.2.1 Visual Pathways

Light that is reflected from objects enters the eye via the cornea and the lens, which help to focus the image on the retina, a layer of cells at the back of the eye. Via their axons, visual information leaves the eye by way of the optic nerve. There is a partial crossing of axons at the optic chiasm. Here, half of the fibers from each eye cross such that each visual half-field is represented in the opposite hemisphere of the brain. After the chiasm, the axons wrap around the midbrain to reach the lateral geniculate nucleus (LGN), where almost all the axons must synapse. An important aspect of organisation in the LGN concerns the segregation of information received from the ganglion cells of the retina: The outer four layers of the LGN are composed of small cells, and correspondingly, receive inputs from the small ganglion cells of the retina. These so-called parvocellular layers code information about colour and form. The magnocellular layers 5 and 6 of the LGN, on the other hand, are composed of large cells and receive their input from large ganglion cells, coding information about luminance contrast, orientation, and coarse form. This segregation is believed to be the basis of independent visual pathways (e.g., Schiller, Logothetis, & Charles, 1990) from the retina to the cortex that are specified for detailed shape and colour (parvocellular), and motion and depth (magnocellular).

The division of visual information (motion, colour and form) is maintained (at least to some degree) in separate layers in the LGN, V1 (the first cortical area to receive visual information concerning shape information), and after passing through V2, in separate areas of associative cortex. The parietal visual cortical areas are concerned with motion of objects, navigation through the world, and spatial reasoning. Temporal visual areas such as V4 and infero-temporal cortex (IT) are involved with the complex perception of patterns and forms as recognisable objects (Logothetis & Sheinberg, 1996). This organisation is important for theories of object recognition as it implies that shape representation may be dealt with separately from other aspects of the object (such as location in space).

The division of labour in the higher areas of the visual system has been studied by Ungerleider and Mishkin (1982). These researchers trained monkeys with lesions in the parietal or in the temporal cortex to associate food with the shape of an object or the object's location. Monkeys with parietal lesions were able to recognise objects' shapes but they performed badly on a spatial location task. The group of monkeys with temporal lesions showed the reversed pattern of performance. According to Ungerleider and Mishkin, higher levels of visual computation seem to divide the labour of object recognition into two tasks associated with different pathways in the brain (see Figure 2). The so-called *what system,* associated with the infero-temporal lobe, is concerned with identification and categorisation of objects. The *where system,* associated with the parietal lobe, is concerned with the location of objects and surfaces in the visual field. Single cell recordings support this interpretation. Temporal cells with receptive fields that include the fovea and its surroundings are sensitive to specific shapes like faces or hands (Desimone, 1991). Parietal cells were found to respond to spatial properties of visual stimuli (Andersen, Essick, & Siegel, 1985). Goodale and Milner (1992) reinterpreted the role of these two streams. Rather than receiving different visual information, the ventral stream (associated with IT) is responsible for object identification processes, whereas the dorsal stream (projecting to the posterior parietal cortex) mediates the sensorimotor transformations to guide actions.

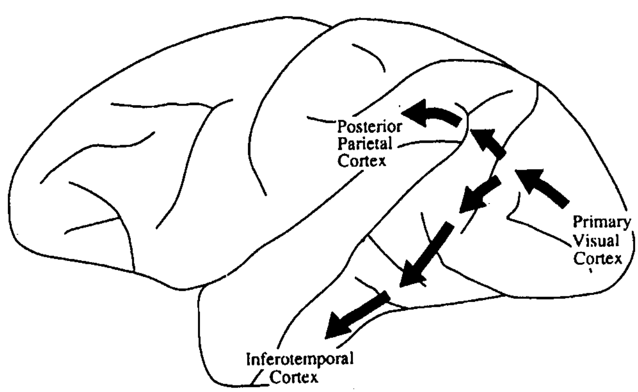


Figure 2. Ungerleider and Mishkin's (1982) conception of the "what" (projecting to the infero-temporal cortex) and the "where" path (projecting to the posterior parietal cortex) in the primate cortex (adapted from Goodale, 1995)

1.2.2 Representation of Shape in V1 and Beyond

Whereas the cells in the retina and LGN have a centre-surround organisation (responding maximally to spots of light or darkness) the so-called simple cells in V1 respond to variation in luminance at a particular orientation (e.g., to a horizontal but not a vertical bar). Another type of cell, so-called end-stopped cells, respond maximally to an oriented stimulus if it terminates within the receptive field of that cell (Hubel & Wiesel, 1959, 1965); this property makes them candidates for detecting contours that end at corners (vertices). According to Biederman (1995), activation of simple and end-stopped cells in V1 are the initial cortical representation of shape. Indeed, the differential activity of these cells would allow the discrimination of shapes. However, the same shape at a different location will activate different V1 cells, as would changes in orientation or size. Because of these properties a representation of shape is needed that is beyond that of V1.

Neurophysiological studies with monkeys as well as neuropsychological work shows that IT plays a major role in object recognition (Biederman, Gerhardstein, Cooper, & Nelson, 1997; Logothetis & Sheinberg, 1996). The so-called ventral pathway V1→V2→V4→IT is generally assumed to subserve object recognition. In physiological studies, the properties of visual neurons along this pathway show a hierarchy of complexity in response to visual stimuli. Their properties change in the course of processing from the retina, passing through the LGN and V1 and on to the ventral pathway structures in the inferior temporal lobe (Felleman & Van Essen, 1991). Neurons in the early stages of this hierarchy have small receptive fields and they are tuned to simple visual properties such as colour, orientation, and motion. Although the information about a single object is distributed across different neurons, these are all relatively close together, allowing the features of the object to be loosely "bundled" (Wolfe & Cave, 1999). Neurons at subsequent stages in the hierarchy have large receptive fields, responding to more complex, multipart stimuli (Tanaka, 1993). A consequence of this hierarchical organisation is that single neurons at upper stages respond to the combination of components that form a complex object or they respond to the combinations of features and properties that belong to a single object.

Lesion studies on animals also strongly indicated that IT serves an important role in the recognition of object shape. Holmes and Gross (1984a; 1984b) showed that bilateral removal of IT in monkeys results in deficits in learning object discriminations when the stimuli differ in shape. In addition, relearning of object discriminations that were trained before the lesions was severely impaired. At the same time, IT lesioned animals were not worse than controls in discriminating stimuli that were of the same shape but differed in size or in their 2D orientation. Thus, IT seems specifically sensitive to shape discrimination.

Although single cell recordings and lesioning studies clearly indicate a specific role of IT in object recognition, data from imaging studies are not always quite so clear. In a meta-analysis, Farah and Aguirre (1999) compiled data from 17 positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies intended to establish the brain areas associated with object recognition. Only those studies were considered in which a recognition condition was contrasted with a non-recognition baseline condition. Farah and Aguirre summarised the local maxima of activation reported in these studies by using a standardised coordinate system of the brain. A total of 84 coordinates were derived from 20 tasks from 17 different studies (see Figure 3).

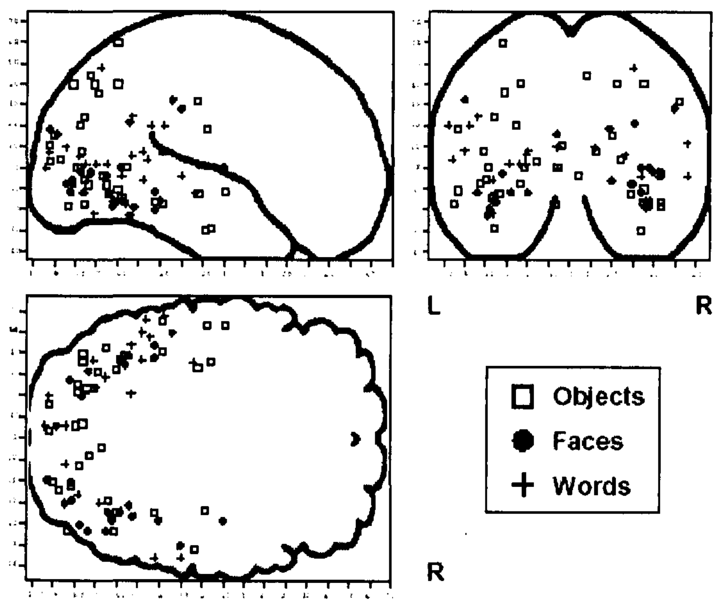


Figure 3: Local maxima placed in standard space derived from studies of visual recognition and projected on sagittal, coronal, and axial views (adapted from Farah & Aguirre, 1999).

The distribution of local activation maxima for the subsets of the three classes of visual stimuli - objects, faces, and words - showed no focal clusterings but subtended almost equally across the area covered by all the visual stimuli together. There were also no clear hemispheric differences for any of the stimulus subsets. Similarly, including distinctions such as active versus passive recognition tasks or line drawings versus grey-scale images reduced the scatter only marginally.

The lack of clear loci of visual processing led Farrah and Aguirre (1999) to the rather general conclusion that visual recognition activates posterior brain areas. However, the studies reviewed by Farah and Aguirre (1999) had the important limitation that a recognition condition was contrasted with a nonrecognition baseline condition. A different and more successful approach to studying visual recognition with neuroimaging is the subtraction of one recognition condition from another. For example Kanwisher and her colleagues (Kanwisher, Chun, McDermott, & Ledden, 1996) used fMRI to compare regional brain activity when participants viewed photographs of faces and of objects. An objects-minus-faces subtraction revealed areas more responsive to objects than faces whereas the reverse subtraction revealed a region that responded more to faces than other objects. Both types of stimuli activated inferior temporo-occipital regions, but face-specific activation was confined to part of the right fusiform gyrus. In a subsequent PET study, Kanwisher and her associates (Kanwisher, Woods, Iacoboni, & Mazziotta, 1997) located an area in human extrastriate cortex that seemed involved in shape recognition. Observers viewed line drawings of 3-dimensional objects either in an intact configuration or as scrambled drawings with no clear shape interpretation. Only when the shapes were intact, regional blood flow increased in a bilateral extrastriate area near the border between the occipital and temporal lobes and a smaller area in the right fusiform gyrus. Interestingly, responses were seen for both novel and familiar objects indicating that this area is involved in the bottom-up (i.e., memory-independent) processing of visual shape.

In a recent review of PET and fMRI studies, Cabeza and Nyberg (2000) found a more distinct pattern of activation maxima concerning object recognition than Farah and Aguirre (1999). Cabeza and Nyberg's meta-analysis supports the notion that the ventral occipito-temporal pathway is associated with object-information whereas the dorsal occipito-parietal pathway is responsible for spatial information (Ungerleider & Mishkin, 1982). The important role of IT for recognition of object shape has also been corroborated by studies on patients with anterior inferior temporal lobectomies (Biederman et al., 1997) who were severely impaired in object recognition. PET studies have, similarly, found that the posterior infero-temporal lobe is implicated in object naming (Moore & Price, 1999; Price, Moore, Humphreys, Frackowiak, & Friston, 1996).

Thus, there is strong converging evidence that IT seems to be specifically important for object recognition. However, it is not yet clear in how far neurons in IT respond to view changes and other transformations for a given stimulus. Although IT seems in general rather insensitive to variations in position and size, there is evidence for both view-dependent and view-invariant neurons from single-cell studies in monkeys (Logothetis, Pauls, & Poggio, 1995; Lueschow, Miller, & Desimone, 1994).

1.2.3 Image Transformations in the Ventral Pathway

A large region of the ventral occipito-temporal cortex appears to be involved in the analysis of object structure independent of the image cues (e.g., line drawings vs. grey-level images, shading, texture) defining the object's shape (Kourtzi & Kanwisher, 2000). Kourtzi and Kanwisher (2001) studied the human lateral occipital complex (LOC) which is thought to be involved in object recognition (e.g., Grill-Spector et al., 1998). Their interest was in whether the LOC represents low-level image features or perceived object shape. They used an fMRI adaptation paradigm in which the response to pairs of successively presented stimuli is usually lower when they are identical than when they are different. Adaptation in the LOC was found when the perceived shape was identical and its contours differed but not when the contours were identical and its perceived shape was changed. According to Kourtzi and Kanwisher, these results indicate that the LOC represents not just simple image features, but higher level shape information.

Earlier, Grill-Spector and colleagues (Grill-Spector et al., 1999) studied the invariant properties of human cortical neurons in the LOC by presenting the same object in different viewing conditions. The LOC exhibited stronger fMR adaptation to changes in size and position (more invariance) compared to changes in illumination and viewpoint. The authors propose that there are two subdivisions within LOC into posterior and anterior regions, with the posterior being more sensitive to object transformations. This is also consistent with lesion studies showing that damage to areas V4 and posterior IT (Schiller, 1995; Weiskrantz, 1990) affects the ability to compensate for object transformations such as size (Ungerleider, Ganz, & Pribram, 1977), orientation and illumination (Weiskrantz & Saunders, 1984) rather than the ability to recognise nontransformed shapes. In contrast, lesions to anterior IT caused a general deterioration in recognition capacity (Weiskrantz, 1990) independent of object transformations.

Gerlach and colleagues (Gerlach et al., 2002) conducted PET activation studies to investigate the brain areas involved in integrating contours into "wholistic" shape representations. Observers had to make global shape judgements during PET scans on three stimuli types: outline drawings, their contour-deleted (fragmented) versions and unrecognizable fragmented versions (non-collinear forms). The results showed that holistic integration of contours was associated with bilateral activation of the ventral parts of the occipital lobes as well as posterior parts of the temporal lobes. Activation in the fusiform and inferior temporal gyri was higher for recognisable stimuli (outline drawings and fragmented drawings) than for unrecognisable stimuli (non-collinear drawings). Gerlach et al. (2002) concluded that contours are integrated in the ventral part of the occipital lobe in a bottom-up fashion, whereas activation in the posterior parts of the fusiform and inferior temporal gyri indicate the involvement of stored structural representations.

In summary, animal research, neuropsychological investigations and brain imaging studies point to a crucial role of the infero-temporal cortex for visual object recognition in vision (for a review, see Logothetis & Sheinberg, 1996). Whereas neurons in V1 have small receptive fields and respond to simple bar-like stimuli neurons along the ventral stream show increasing receptive field sizes, as well as an increasing preference to complex stimuli (Kobatake & Tanaka, 1994). Furthermore, there is evidence for neuronal differences in the ventral pathway (in particular within the LOC) associated with the transformation of an object such as changes in scale and position, orientation, and fragmentation.

1.3 Properties of Object Recognition

1.3.1 View Invariance and View Dependency

A single 3D object can be encountered from a number of viewpoints each producing a potentially unique 2D projection. Naturally, the questions arise how dependent is visual object recognition on changes in viewpoint, and are there types of viewpoint change that affect recognition more than others? In a classic study by Shepard and Metzler (1971), observers had to simultaneously match novel, three-dimensional objects made up of blocks that were arranged at right angles. For match trials there was a linear relation between the speed of response and the angle of rotation separating the two views of an identical object, which was found for both plane and depth-rotations. It was argued that this monotonic relationship was best explained by an analogue transformation similar to mental rotation of the representation of an object.

Studies using more familiar and naturalistic objects have obtained results that confirmed the findings on view-dependence. For example, Bartram (1976) sequentially presented line drawings and photographs of familiar objects to his observers who were required to match them. In mismatch trials, objects from two different object categories were presented. Bartram's results show that identical view matches were more rapidly performed than matching different views of the same objects which was more rapid than matching different exemplars of the same name. Ellis and his colleagues (Ellis & Allport, 1986; Ellis, Allport, Humphreys, & Collis, 1989) supported these findings.

Strong evidence for view-dependent recognition effects across several tasks were subsequently obtained in studies using line drawings of familiar objects rotated in picture plane (e.g., McMullen & Jolicoeur, 1990, 1992; Murray, 1995a; Murray, 1995b, 1998, 1999; Murray, Jolicoeur, McMullen, & Ingleton, 1993) and in depth (e.g., Hayward, 1998; Lawson & Humphreys, 1996, 1998, 1999; Lawson, Humphreys, & Jolicoeur, 2000; Lawson, Humphreys, & Watson, 1994). View-point dependent effects have also been obtained in other laboratories with novel objects rotated in the picture plane (e.g., Tarr & Pinker, 1989, 1990) and in depth (e.g., Bulthoff & Edelman, 1992; Hayward & Tarr, 1997; Hayward & Williams, 2000; Tarr, 1995; Willems & Wagemans, 2001).

In addition to performance costs for view changes found in object recognition, there is another line of evidence for view-specificity in object recognition. Neuropsychological research by Warrington and her colleagues (Warrington & James, 1988; Warrington & Taylor, 1973, 1978) showed that patients with damage to their right parietal brain areas have particular problems in recognising objects from unusual (accidental or unconventional) views. The implications were examined in detail by Palmer, Rosch, and Chase (1981) who showed that some views of an object might be processed more efficiently than other views. For instance, "good" or "typical" views might optimally reveal important diagnostic features of the object that can activate a stored, abstract representation faster. In contrast, "bad" or "accidental" views may be ambiguous in terms of the interpretation of their three-dimensional structure. Palmer et al. (1981) defined the former type of view as "canonical" which was measured with a) ratings of the goodness of a view; b) judging of perspective when objects have to be imagined; c) and the best view from which a previously imagined object should be photographed. Palmer and his colleagues showed that objects presented in these canonical views were named faster than when shown in other views. Their conclusion was that in canonical views the critical information about an object is presented to an optimum within an image, and that matching a view with an object-centred representation is influenced by how efficiently this information can be extracted from a view. More recently, Blanz, Tarr, and Bulthoff (1999) also found a high degree of consistency of preferred views for a given object across participants. Blanz et al. used tasks similar to those of Palmer et al. (1981) in combination with modern computer graphics manipulation methods. The evidence for preferred or canonical views was also indirectly supported by other researchers (Humphrey & Jolicoeur, 1993; Lawson & Humphreys, 1998, 1999) who showed that foreshortened views - views in which the axis of elongation was almost parallel to the line of sight and some parts of an object were occluded - were generally harder to recognise than more typical non-foreshortened views (at least on initial presentations).

Despite the strong evidence for view-specificity in human object recognition, Biederman and his colleagues have obtained view-invariant effects after image transformations (e.g., Biederman & Cooper, 1992; Biederman & Gerhardstein, 1993). Biederman and his associates typically employed long-term priming tasks to investigate object constancy. Participants named a set of prime pictures of objects in one block of trials and then named the same set of objects as targets in a test block, in which some target pictures were manipulated from their initial presentation. Naming latencies in the test block were in general faster compared to those in the study block but the magnitude of these priming effects was often equally large for images that were transformed from the initial presentation as for identical images. According to these studies, object recognition is invariant with the location of the image in the visual field (Biederman & Cooper, 1991a), the size of the image

(Biederman & Cooper, 1992), left-right (i.e., mirror) reflection (Biederman & Cooper, 1991a), and some rotations in depth (Biederman & Gerhardstein, 1993, 1995).

There are controversial interpretations concerning the nature of view-dependent and view-independent performance in object recognition. For example, view-dependent effects found with depth-rotated objects may arise because different parts were visible across different views of an object. In cases where the part structure does not change recognition performance may be the same across different orientations. Biederman and Gerhardstein (1993) showed that in priming tasks view-dependent effects are pronounced if different parts are visible between study and test view. Similarly, the effects of view-dependency depend on the nature of the task employed to measure recognition performance (Biederman & Cooper, 1992). These issues will be addressed in later sections.

In summary, most recognition tasks produce decrements in performance after a change in viewing conditions. Concerning naming tasks, object recognition seems invariant to mirror-reflections (Biederman & Cooper, 1991a; Fiser & Biederman, 2001; Stankiewicz et al., 1998), changes in size (Biederman & Cooper, 1992; Fiser & Biederman, 1995, 2001; Stankiewicz & Hummel, 2002), translation across the visual field (Biederman & Cooper, 1991a; Fiser & Biederman, 2001; Stankiewicz & Hummel, 2002) and some rotations in depth (Biederman & Gerhardstein, 1993; Tarr, 1995; Tarr & Bulthoff, 1995). At the same time, object recognition is sensitive to rotations about the line of sight (Jolicoeur, 1985; Jolicoeur & Milliken, 1989; Tarr & Pinker, 1989, 1990, 1991) and most rotations in depth (for reviews, see Lawson, 1999; Logothetis & Sheinberg, 1996).

1.3.2 Object Invariance Across Variations in Shape

Studies on object recognition often show that effects of view dependence quickly diminish over time. For example, Posner and his colleagues (Posner, 1969; Posner & Keele, 1967) used letters that were either matched simultaneously or sequentially. The results showed an advantage for the matching of identical letters in comparison with the matching of letters with the same name but differing case. However, the advantage for sequential matching of identical relative to nonidentical stimuli was found only with short interstimulus intervals (ISI). Posner concluded that these results indicate that two representations may work when matching of stimuli was required: a visual representation mediating the rapid matching of identical images, and another name representation that allows generalisation across the same letters differing in case. These results were confirmed by other researchers (Ellis & Allport, 1986; Lawson & Humphreys, 1996) and taken as evidence for a more abstract representation that generalises over variations in shape.

According to Hummel (1997), there are at least two evident manifestations of the capacity of the visual system to generalise over variation in an object's 3D shape. First, humans can recognise a never seen before object as a member of a class, such as an exotic fish. Second, observers generalise immediately across particular instances of an object for example, when people spontaneously name a picture of an object such as a collie as simply a "dog"; this level of categorisation is termed "basic level" categorisation (Rosch et al., 1976) or "entry-level" (Jolicouer, Gluck, & Kosslyn, 1981). In contrast, naming a picture of a chair as "rocking chair" is defined as subordinate level categorisation. Subordinate classification has been found to take longer than basic level naming (Rosch et al., 1976), and superordinate classification takes even longer (Jolicouer et al., 1981).

Rosch et al. (1976) argued that visual classifications are initially made at a "basic" level because basic level names are more available than subordinate-level names. Rosch et al. showed that basic level terms are the first to be acquired in a child's vocabulary, have fewer syllables, and are used much more frequently to refer to objects than the subordinate-level terms. Moreover, basic-level concepts seem to enjoy a visual advantage over superordinate level concepts. Rosch et al. demonstrated that members of a basic-level class, such as "car", tended to have more similar shapes than members of a superordinate-level class, such as vehicles. By superimposing silhouettes of exemplars from different levels they found that basic-level composite image remained more identifiable than a composite of superordinate-level exemplars. There was also an advantage of basic-level over subordinate-level naming, although the difference in shape variability was not as large as between basic and superordinate level. According to Rosch et al. (1976), perceptual information required for basic-level classifications may simply be more discriminable or salient than the information required for subordinate-level classifications. Basic level categorisation may be the default recognition level because it represents an optimal trade-off between a sufficient amount of information and ease of distinction (Biederman, Subramaniam, Bar, Kalocsai, & Fiser, 1999; Biederman, Subramaniam, Kalocsai, & Bar, 1999; Humphreys, Price, & Riddoch, 1999).

The ability to generalise over variations in exact 3D shape seems to be resilient even for patients with problems in object recognition. For example, Davidoff and Warrington's (1999) patient RK suffered from a form of object agnosia that rendered him incapable of identifying objects with disconnected parts or objects shown in unusual views. However, despite being unable to detect part changes in intact images he was still normal at naming those same objects in familiar views.

In summary, the visual system generalises readily over variations in shape. Objects are usually recognised at a basic level rather than as a specific instance or exemplar. Thus, theories of human object recognition not only have to account for changes in viewpoint for a particular object, but also for how perceptual representations mediate entry-level classification.

1.4 Object Recognition Theories

1.4.1 The Debate on Formats of Object Representations

As the review above shows, the human capacity for visual object recognition is characterised by a number of properties that are jointly very challenging to explain. Visual representation of shape is invariant with (i.e., insensitive to) some, but not all, variations in viewpoint. At the same time, object recognition is remarkably robust to variations in shape. Object recognition theories place different stress on these properties. In particular, there is an ongoing debate between the two main groups of theorists about how the visual system achieves view and object constancy. Several issues of object recognition are controversially discussed by vision theorists such as the nature of the perceptual description, the nature of the stored description, and the nature of the matching process (Tarr, 1995; Willems & Wagemans, 2001). The alternative views for these three issues are presented in Table 1, and will be considered in depth in the subsequent sections.

These important issues are often summarised in the psychological literature into two groups of theories, sometimes categorised as viewpoint-dependent (or viewer-centred) and viewpoint-invariant (or object-centred) theories of object recognition (Tarr, 1995). This terminology is somewhat misleading, as Biederman (Biederman, 2000; Biederman, Subramaniam, Bar et al., 1999) pointed out, because all theories of object recognition include a stage in which the representation resembles the 2D input (e.g., in V1, see section 1.2.2), so any theory must entail some view-dependent recognition. However, the first group of theories (which will be referred to as image-based or view-based) claims that representations mediating object recognition are based on views or metric templates. These theories posit for example, primitive-to-reference point, coordinate relations (e.g., Bulthoff & Edelman, 1992 ; Lowe, 1987; Olshausen, Anderson, & Van Essen, 1993; Poggio & Edelman, 1990; Tarr, 1995; Tarr & Pinker, 1989; Ullman, 1989; 1998; Ullman & Basri, 1991) in which the exact distance of each primitive from a fixed reference point (or set of fixed reference points) is represented. The second group (which will be referred to as part-based) assumes that the visual system extracts a more abstract representation from the 2D image, such as parts and their relations, a type of representation often termed a structural description (e.g., Biederman, 1987; Dickinson, Pentland, & Rosenfeld, 1993; Hummel & Biederman, 1992; Marr, 1982; Marr & Nishihara, 1978) in which each primitive is related to other primitives in the representation using directional categorical descriptors (e.g., "above", "below", "side-of"). In the next section, both major clusters of theories will be discussed.

Theory	Perceptual description	Stored description	Matching process
"Part-based" Theories			
Marr & Nishihara (1978) Marr (1982)	3D object model	3D object model	Direct mapping
Biederman (1987) Hummel & Biederman (1992)	Geon structural description	Geon structural description	Direct mapping
"View-based" Theories			
Tarr & Pinker (1989) Tarr (1995)	2D	2D	Mental transformation
Ullman (1989)	2D	3D	Alignment
Poggio & Edelman (1990)	2D	Multiple 2D	View interpolation
Ullman & Basri (1991) Ullman (1998)	2D	Multiple 2D	Linear combination

Table 1: Overview of current theories of object recognition (adapted from Willems & Wagemans, 2001).

1.4.2 View-based Models

1.4.2.1 Introduction

Theories of human object recognition that stress the importance of the familiarity with certain views of an object are often subsumed as template matching theories. A template is a representation that resembles a specific view of an object (although it is often assumed that some pre-processing compensates for variations in location, illumination, etc.). The central idea is that a long-term memory representation is analogous to the pattern of retinal stimulation projected by that shape (Selfridge & Neisser, 1960). In principle, the input array can be superimposed on all the templates stored in memory. Recognition is achieved when the closest match between input array and memory-template is found. These models are often termed view-based or image-based theories because they propose that object representations are based on familiar 2D views (Gauthier & Tarr, 1997; Tarr, 1995).

To achieve viewpoint invariant recognition, a system based on template models must either store a very large number of views for each known object or store a small number of views (or 3D models) for each object and match them against incoming images by means of some transformation or normalisation process. Successful recognition occurs when a template is found that fits the image within a tolerable range of error. Template models have three main properties (Hummel & Biederman, 1992): (1) The match between the input image and a template depends on the extent to which points in the image spatially correspond to points in the template; (2) a template represents an entire view of an object, which means that spatial relations among points (or features) in the object are coded implicitly as spatial relations among points in the template; and (c) viewpoint invariance is achieved by either matching the input to a single template or to different templates corresponding to the same object label.

Before turning to specific theories it has to be pointed out why the most straightforward template approach to object recognition - storing a sufficient number of different views associated with each object - is not viable. In such a simple version of a template model, the currently viewed object would have to be matched with all the views stored in memory (Abu-Mostafa & Psaltis, 1987). Models of so-called associative memories have been proposed for implementing this 'direct' approach to recognition (Hopfield, 1982; Kohonen, 1978; Willshaw, Buneman, & Longuet-Higgins, 1969). Such direct comparisons to stored views may be useful to some extent, especially for the recognition of highly overlearned objects. However, this direct approach by itself is insufficient to explain object recognition in general. One reason is that the space of all possible views of all the objects to be recognised is likely to be prohibitively large (Ullman, 1998). A more fundamental reason is the problem of generalisation over changing viewing conditions. Consider a study seeking to quantitatively measure the variation between images due to changes in orientation and illumination (Adini, Moses, & Ullman, 1997). The goal was to compare images of different individuals with images of the same individual but under different viewing conditions. The results indicate that the differences induced by changes in the viewing conditions are large compared with the differences between different individuals. Therefore, direct image comparisons are not sufficient for recognition (Ullman, 1998). For reliable recognition, processes that can compensate for the effects of viewing conditions are required, some of which are described in the following sections. These include the multiple views account, Ullman's alignment of views, the metric template model of Lades et al. (1993), and the Poggio and Edelman (1990) view interpolation approach.

1.4.2.2 Multiple Views Account (Tarr, 1995; Tarr & Pinker, 1989)

One widely discussed variant on view-based theories uses multiple views to represent the three-dimensional structure of objects. While the direct approach to template matching would predict that a view-based system stores an object representation in many different viewpoints, a more comprehensive representation may be formed by linking a limited set of characteristic views of that object. For instance, in Figure 4 there are five characteristic views of an aeroplane that cover many of the observable part and surface configurations. According to the multiple views idea, recognition of an object in a particular view is achieved when the input image activates the view (or set of views) that corresponds to a familiar object. A certain input view may not match exactly with a particular stored view; therefore, a normalisation process may be employed to align the input view with the nearest stored view. This class of theories is therefore called multiple-views-plus-transformation theory of recognition (Tarr & Pinker, 1989).

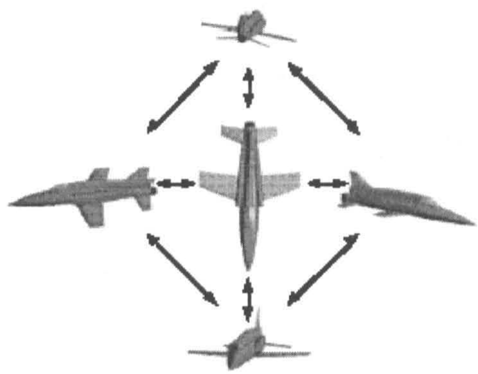


Figure 4: Visualisation of the multiple views concept (adapted from Tarr, 1995). The object jet fighter may be stored as a collection of typical or familiar views.

According to multiple views theory, objects are stored in a set of characteristic views. Many studies have shown orientation dependent recognition performance for objects rotated in the picture plane and for some rotations in depth (see section 1.3.1). Tarr and Pinker (1989) demonstrated that orientation dependent performance largely depends on the familiarity of views, a factor that may vary considerably in common objects. In a series of experiments, they employed line drawings of novel stimuli that were highly similar to one another. Each of these character-like objects had a clearly marked bottom and a vertical axis. First, participants practised to name the objects shown in a single viewpoint; later, the objects were presented in unfamiliar viewpoints. Multiple views theory predicts initial recognition costs after changes in orientation that are monotonically related to the degree of rotation from the training viewpoint. Furthermore, these mental rotation costs should diminish with some practice. Finally, the presentation of the objects at yet again different unfamiliar viewpoints should yield higher reaction times the more the stimuli are rotated away from the familiar (trained) viewpoints; this is exactly what was found by Tarr and Pinker (1989).

Tarr (1995) used connected-cube objects similar to stimuli employed by Shepard and Metzler (1971). These objects were novel to the participants and allowed the control of familiarity of some views by frequency of presentation. The objects were rotated in depth around all three axes. After extensive practice in naming the objects in familiar views, the same objects were presented to the participants in novel viewpoints as well as in trained viewpoints. The major findings were: 1) Recognition times and error rates generally increased with the angular distance from the nearest trained viewpoint; (2) With extensive training, performance became nearly equivalent at all familiar views. Practice effects, however, did not transfer to unfamiliar views; (3) The patterns of response times also implied that participants mentally rotated the stimuli along the shortest 3D path to a familiar viewpoint. Tarr (1995) claims that these results are consistent with earlier studies using 2D stimuli (Tarr & Pinker, 1989) and support a multiple views theory of object recognition. Multiple views theory predicts the observed response time patterns of an initial effect of viewpoint that diminishes with practice and returns for unfamiliar views. According to the authors, view-based theories of object recognition cannot explain this effect. Tarr and Pinker (1989) also tested the "rotation-for-handedness" hypothesis. This hypothesis has been put forward (Corballis, 1988; Hinton & Parsons, 1981) as a "special case" in which mental rotation was only required when handedness decision is involved in the recognition task. Tarr and Pinker (1989) found viewpoint-dependency for recognition as well as for handedness decision tasks. Viewpoint-dependency did not diminish when handedness was made explicitly irrelevant for the recognition task. Tarr also tested whether normalisation procedures are used only when participants have to perform a top-bottom discrimination. He found that response time patterns were equivalent irrespective of whether objects preserved their top - bottom position with respect to gravity or not.

In summary, the multiple-views-plus-transformation account is able to account for many viewpoint-dependent effects found in object recognition studies. A problem for the multiple-views-plus-transformation account is that the idea of an analogue process of mental rotation has been recently challenged. For example, Jolicoeur and his colleagues (Jolicoeur, Corballis, & Lawson, 1998) let participants name rotated objects or decide whether the objects would face left or right if they were upright. Latencies in the left-right task were influenced by a rotation after-effect or by the physical rotation of the object. However, perceived rotary motion did not influence naming time, which suggests that the identification of rotated objects does not involve mental rotation. Neuropsychological evidence also casts some doubt over the importance of mental rotation for achieving object constancy. Farah and Hammond (1988) report that their patient RT revealed poor performance on mental rotation tasks despite being able to identify a set of upside-down objects. The next sections describe models of object recognition based on views that circumvent the problems associated with mental rotation by using different matching mechanisms (see Table 1).

1.4.2.3 Alignment and Linear Combination of Views (Ullman, 1989, 1998)

Instead of assuming an analogue process such as mental transformation to normalise a view with stored object representations, some theorists have proposed that normalisation can be achieved by computing a view transformation. In a model by Ullman (1989) recognition is divided into two stages. First, a transformation in space is computed to bring the viewed object into alignment with candidate 3D object models. This can be achieved by minimal information such as the objects general orientation or a small number of corresponding feature points (termed alignment keys) in the view and the model. In the second stage the model selects the best match for the current view.

Ullman (1998) later proposed another variant of this model in which a 3D object is represented by the linear combination of 2D views of the object. The views comprising the representation of a single object are not merely a collection of independent 2D object views. In contrast, Ullman's approach uses a number of object views collectively in the recognition process by the mathematical concept of linear combinations. It allows all possible views of a rigid object that can undergo rotation in space, translation, and scaling, to be spanned by the linear combinations of three views of the object (Ullman & Basri, 1991).

However, there are at least four limitations concerning the use of linear combinations of views (Ullman, 1998). First, objects with smooth bounding contours, such as an egg or a ball, require more views because the object's shape is not generated by fixed contours on the object. Second, more corresponding features between views are needed if the projections are non-orthographic, that is in perspective. Third, three views are insufficient for representing an object from all orientations due to self-occlusion. For example, a different set of views will be required to represent the 'front' and the 'back' of a grand piano. Finally, the transformations that the object is allowed to undergo are restricted to rigid transformations.

1.4.2.4 Metric Templates (Lades et al., 1993)

Von der Malsburg and his colleagues (Lades et al., 1993) developed a biologically inspired image recognition model based on metric templates. The input layer of the model consists of an array of columns of units with properties that are modeled after those from V1 cells. A particular cell is tuned to variation in luminance at a particular orientation, scale (i.e. spatial frequency) and orientation. This tuning of a cell is modeled mathematically by a sinusoidal filter called a Gabor filter. A column of multiscale, multiorientation spatial (Gabor) kernels with local receptive fields centred on a particular point in the image input is termed a Gabor jet, roughly corresponding to a simple cell of a V1 hypercolumn. The pattern of activation of the 80 kernels (5 scales, 8 orientations, 2 phases - sine and cosine) in each of the jets is mapped onto the second layer. This second layer simply stores the pattern of activation over the kernels from a given image. By this method a large number of images can be stored to form a long term memory. A new image is matched against stored images by allowing the jets to independently change their positions to determine their own best fit. The similarity of a pair of images is a function of the similarity of the activation values of the Gabor filters for corresponding jets combined with the degree to which a given jet has to be displaced to find its best match in a new image. Therefore, the deformation of the lattice of original positions typically provides a visual measure of the similarity of two images: The greater the deformation of the lattice, the lower the similarity. This allows a matching of two input images with moderately different orientations or - in the case of faces - expressions.

The model is completely view-based; when different views of an object are encountered, the pattern of activation in the filters vary, and so may the layout of the Gabor jets. The model was able to generalise across view changes of about 30° in depth for face stimuli. Rotations beyond that extent, however, caused a serious decline in recognition memory. However, the model is insensitive to changes of position in the visual field and size, as the whole lattice can be repositioned or scaled. Although based on template representations, the Lades et al. model is considered to be a successful attempt to characterise the processes in early stages of visual processing, e.g., from V1 to V4, even by proponents of structural description models (Biederman, 2000).

1.4.2.5 Non-Linear View Interpolation (Poggio & Edelman, 1990)

A variant of the multiple views approach concerns the matching or transformation process. Instead of linear combination of views or an analogue process such as mental rotation, this approach claims that two or more views of an object may be matched by nonlinear view interpolation (Poggio & Edelman, 1990). In this model, generalisation from stored to novel views is achieved by a multivariate function interpolation in the space of all possible views. Poggio and Edelman postulate that for each object a smooth function exists that maps any perspective view into a standard view of the object. Furthermore, this multivariate function may be approximated from only a small number of views. A new view of a target object may be recognised by interpolation among its selected stored views that represent the object. The function used for interpolation is object specific - each object has a different function.

The problem of synthesising an approximation to a function from a limited number of views can be solved by learning input-output patterns from a set of examples. One solution to approximate smooth functions are radial basis functions (RBF) which consist of a basis function (e.g., a Gaussian) as well as coefficients and parameters whose values are found during a learning stage. The learning of weights is achieved by minimising a measure of the error between the network's prediction and the desired output for each of the examples. The function is based on a prototypical view which may be updated during learning.

Poggio and Edelman (1990) designed a neural network model similar to that of Lades et al. (1993) in that it basically matches the output of units modeling simple cells of a V1 hypercolumn with patterns of activation in an object memory. In addition to the two layers in the Lades et al. model, the Poggio and Edelman network includes a single hidden layer between input and output stage. The units in this hidden layer correspond to RBFs and allow the optimal generalisation across a large number of images. In simulations with novel objects Poggio and Edelman (1990) found that after training with 10 to 40 views the centres of the radial basis functions corresponded to views that were different from any of the trained views. They showed that with as few as two RBF units, these views of a single object may be recognised for rotations across the viewing sphere. Whereas the object layer in the Lades et al. model represent a particular view, the RBFs learn a series of views rather than just one particular view. The RBF can be conceptualised as a prototype for a range of views of an object. A number of RBFs relating to a single object are linked together to form an object prototype. A shortcoming of the model is that it has to be instructed which object is projecting a new image so that an existing RBF can be modified, linked with other RBFs, or a new RBF established.

In the Poggio and Edelman model generalisation from familiar to unfamiliar views is regarded as a problem of approximating a prototype in the space of all possible views (Bulthoff & Edelman, 1992). A recognition system based on this method should perform well for unfamiliar views that are close to stored familiar views, but should become increasingly worse with views that are far from familiar views. A recognition system based on linear combination of views (e.g., Ullman, 1989) should achieve equivalent high recognition performance on those views that fall within the space spanned by the stored collection of model views. In contrast, such a system should perform poorly on views that belong to an orthogonal space.

To test these predictions, Bulthoff and Edelman (1992) trained observers with views of novel objects (reminiscent of bent paperclips and amoebae) by showing them rotated in a motion sequence. In the test phase the objects were presented in new static views and participants had to indicate if the object was a target or not. The test views were either falling within the previous study views (intra), outside the studied views but on the same equator on the hypothetical viewing sphere (extra), or completely different view, however, on a meridian that passes through a trained view. As predicted by the view approximation scheme the performance in these three conditions declined progressively. The results confirmed the predictions by Bulthoff and Edelman (1992) and were taken as behavioural evidence of the non-linear view interpolation model and the view-based approach in general (Edelman, 1997; Tarr, 1999).

1.4.2.6 Evidence in Support of View-based Models

The most striking evidence in favour of view-based models comes from the ample demonstrations of view-dependent performance in object recognition with human observers. In most of these studies, response time and accuracy costs were found when common objects were rotated from familiar views or novel objects from trained views (Hayward & Tarr, 1997; Hayward & Williams, 2000; Jolicoeur, 1985, 1988; Lawson, 1999; Lawson & Humphreys, 1999; Tarr & Pinker, 1989). These effects of view-dependency have been observed employing various recognition and matching tasks and were taken as evidence that the representations mediating object recognition conformed to the view-based models described in the previous sections (Tarr & Bulthoff, 1995, 1998). Moreover, recognition performance was found to improve after training with additional object views (Bulthoff & Edelman, 1992; Tarr & Pinker, 1989). A similar finding was also observed in monkeys trained for object recognition (Logothetis, Pauls, Bulthoff, & Poggio, 1994). This is expected on the view-based approach, because it predicts that generalisation depends primarily on the availability of a sufficient number of representative views.

In addition to behavioural evidence, there is also neuroscientific evidence in support for view-dependent representations. IT cells studied by Tanaka (1996) appear to be governed by 2D similarity of the test view to the view preferred by these cells. Tanaka first showed a large collection of different 3D objects. When a cell responded maximally to a particular object the determining features of the stimuli could be studied. Units seem to be more responsive to particular 2D patterns rather than preferred 3D shapes, regardless of their orientation in space. Similarly, single cell recordings in the macaque superior temporal sulcus revealed that cells were only stimulated by certain perspectives of familiar faces (Perrett et al., 1991) which was taken as clear evidence for viewer centred representations. Another kind of cell responded almost uniformly over all views of a person's face but these object specific outputs were explained as the result of simple superposition of several viewer centred cells. A similar interpretation was applied to the increased tolerance to viewing direction found for broadly-tuned cells that receive converging input from a number of view-specific cells (Logothetis et al., 1995).

More physiological evidence consistent with the notion that the visual system relies on stored views and that there are mechanisms for generalisation to novel views comes from animal lesion studies. Damage to area V4 and posterior IT (Weiskrantz, 1990; Schiller, 1995) appears to particularly impair the ability to generalise over changes in size, orientation, or illumination, although the animal's ability to identify the original views usually is intact. In experiments by Weiskrantz (1990), monkeys were trained to recognise a set of test objects under a variety of particular viewing conditions. Whereas lesions to the anterior part of IT (AIT) caused a general deterioration in recognition capacity, lesioning the more posterior part of IT and some prestriate areas had a more specific effect on the ability to recognise the objects but only when the viewing conditions were changed. In general, these results conform with other physiological work showing that IT is at least to some degree sensitive to stored views (for a review, see Logothetis & Sheinberg, 1996).

1.4.3 Part-Based Theories

1.4.3.1 Introduction

View-based models account for object constancy by assuming that objects are represented in their metric properties, notably as a 2D view or collection of views. In contrast so-called 'object centered' or 'structural description' models propose that objects are represented as descriptions of the relational spatial arrangements of their parts in a three-dimensional coordinate system. Two of the most prominent and widely discussed object recognition theories that include shape representations in the form of a structural description are described in the following sections: Marr's computational theory, Biederman's (1987) recognition by components theory and its computational instantiation by Hummel and Biederman (1992).

1.4.3.2 Computational Theory of Object Recognition (Marr, 1982)

Marr (1982) developed a computational theory of vision in which information from image data is extracted almost completely through bottom-up processing. In his influential theory of object recognition he proposed that a series of representations provide increasingly detailed information about the visual environment. It can be summarised in three stages: the primal sketch, the $2^1/_2$-D sketch, and the 3D model representation.

In its most basic form, the retinal raw-image contains information about light-intensity changes in the visual scene. If represented as a grey-level representation of the retinal image these intensity changes reflect the perceived scene in a pixel-like image consisting of intensity values, which Marr termed the raw primal sketch. Computational filtering procedures yield the boundaries between two regions of different intensity as well as changes in the gradient of intensity. These intensity boundaries vary with the size of the filter and can be mapped as blobs, bars, edges and terminations. In a further step, the structure or organisation inherent in the still ambiguous raw primal sketch is extracted to form the full primal sketch, by grouping visual elements using constraints such as proximity, figural continuity and closure. In the next stage, an intermediate representation of the scene is constructed from the primal sketch. Discontinuities in surface orientation and depth are detected, and higher-level descriptions (e.g., of concave and convex junctions) are generated. The orientation of surfaces can be extracted but surfaces have not yet been grouped into objects. The resulting representation is called a $2^1/_2$-D sketch because it is not completely three-dimensional, but rather reconstructs the third dimension using information like shading, texture, and motion. The final stage allows the extraction of 3D descriptions from the $2^1/_2$-D sketch. This is achieved by dividing the image into parts according to extrema of curvature and then computing the parts' axes and filling in lines joining the endpoints of neighbouring axes. The angles between lines and axes are measured and the resulting description is compared with models stored in memory to choose the best match.

Marr and Nishihara (1978) proposed that objects derived from these parts are described with respect to a co-ordinate system that is centred on the object. The origin of this co-ordinate system lies on the object itself. One or more axes are aligned with standard parts of the object rather than with respect to the viewer-centred co-ordinate system of the $2^1/_2$-D sketch. The 3D representation of objects is viewpoint-invariant, because when the object as a whole is moved, the locations of the objects' parts change relative to the viewer, but they do not change with respect to the object itself.

The objects' parts in Marr's model are described in terms of generalised cones (Binford, 1971). A generalised cone is a 3D volume defined by its two-dimensional cross-section which can be moved along the volume's axis. In this way, various volumes may be categorised according to the shape of the axis, the two-dimensional shape of the cross-section and its change in size. Marr and Nishihara (1978) proposed that the shape of an object is stored in a hierarchy of descriptions each representing parts of different sizes and each with its own co-ordinate system. A local co-ordinate system is centred on a part of the shape represented in the model, aligned with its axis of elongation, symmetry or rotation (if the parts are movable). Thus, for example, there would be a top-level model for a human shape with the co-ordinate system's centre on the torso. This centre would be the starting point for specifying the main parts of the model, their locations, angles and lengths. In the same way subordinate models for each part can be described e.g., for the legs, hands and so forth. This hierarchical structure solves the problem of stability that would arise with only a single object-centred description within a global co-ordinate system. The latter could not be matched successfully with an input in certain cases for example, when the object bends at its joints or when there are a number of similar objects. The description of the right arm of a human shape remains the same in a general co-ordinate system only if its position relative to the torso would not change. Scratching the left shoulder, however, would result in a different description within a general co-ordinate system.

To summarise, Marr proposed an object-centred co-ordinate system in which the object's parts are specified relative to the main axis of the object itself (Marr, 1982; Marr & Nishihara, 1978). Descriptions of the object's shape will be stable across different orientations because rotation of the object will be accompanied by rotation of the co-ordinate system. Thus, a single (canonical) hierarchical representation (3D model) of an object stored in memory is sufficient for identification of familiar objects seen from different viewpoints. According to Marr's approach, object recognition performance should in principle be equivalent across different perspectives.

1.4.3.3 Geon Structural Description Theory (Biederman, 1987)

Biederman (1987) proposed that representations of object shape are based on the recognition of components (RBC). Similar to Marr's model, RBC assumes an early stage in which edges are extracted from the image of an isolated object. Edges will exhibit non-accidental properties (NAPs; Lowe, 1985) which are qualitative 2D properties that remain stable across a range of rotations of the object in depth (except when the curvature coincides with the line of sight which would be an uncommon or "accidental" viewpoint). For example, a curved contour in an image of an object almost always indicates a curved three-dimensional curvature. Concave regions of contour and non-accidental properties allow the extraction of generic parts or "geons" (derived from "geometric ions"). Geons are simple volumetric primitives that can be distinguished from each other from almost any viewpoint, and they can be recognised even when the input image is affected by visual noise. NAPs can be expressed contrastively, such as whether an edge is straight or curved, approximately parallel or nonparallel, or which type of vertex is formed from the cotermination of edges. In this way, regularities in a lower or intermediate level of representation reflect four shape parameters for generalised cylinders that describe the edges, symmetry and size change of a shape's cross sections and its curvature of axis. NAPs are distinguished from metric properties, such as aspect ratio, degree of curvature or degree of angles between two parts.

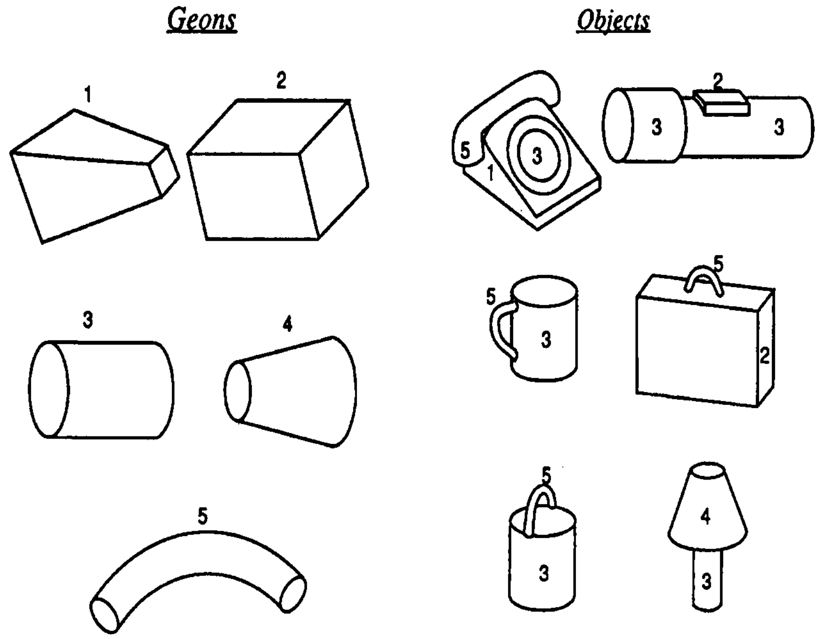


Figure 5. A set of geons and geon-structured objects (adapted from Biederman, 1995). The geons labelled 1 to 5 on the left can be found in the corresponding parts in the objects on the right.

Every object can be thought of as a combination of geons (see Figure 5). An object is thus determined by its structural description, that is the type, number and attributes of geons and their spatial relations to each other. Biederman (1995) claims that with only 36 types of geons there are potentially several million possible descriptions for object consisting of two or more geons. These objects can be rapidly discriminated for entry-level classes. Additional information such as colour will be processed in later stages of recognition.

Since geons can be recognised from a general viewpoint, the same holds for objects built of geons as long as certain criteria are met. Importantly, in contrast to multiple views models, Biederman proposed that viewpoint invariance can be achieved without familiarity or former practice. Biederman and Gerhardstein (1993) suggest three conditions necessary for viewpoint invariance. The first condition is that an object must have a geon description which can be readily extracted by their contours. Thus, wire frame and clay mass objects do not fulfil this condition, because these objects are not decomposable into parts or the parts are highly irregular. The second condition is that each stimulus in a given set must have a distinctive geon structural description (GSD). Thus, the stimuli used by Shepard and Metzler (1971) and Tarr (1995) cannot fulfil this condition, since they are built of only one type of component (or geon) - a cube. The third condition is that rotation must not reveal novel or occlude visible geons of an object (see Figure 6). For example, an extensive rotation of an object consisting of three geons may only leave two or even one geon visible, with the other parts occluded. This would result in an activation of different GSDs and viewpoint-invariance can not be expected.

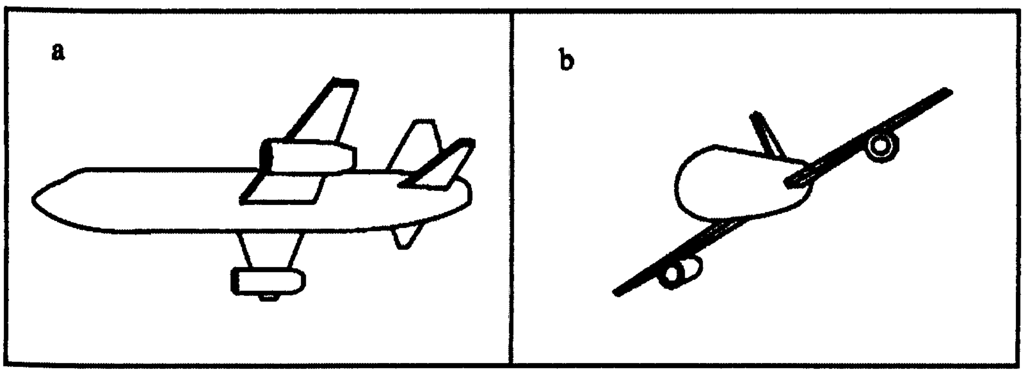


Figure 6: Examples of a change in viewpoint that produces a change of geon-structural description of an object. In the right panel (b) the elevators of the plane are completely occluded and the engines are foreshortened compared to panel (a) so that their cylindrical shape is difficult to infer (adapted from Biederman and Gerhardstein, 1993).

Biederman's RBC model promoted a computational instantiation by Hummel and Biederman (1992). They presented a neural network model JIM (John and Irv's model) which is able to activate a viewpoint invariant representation of a shape in form of a structural description specifying both the visual attributes of an object (e.g., edges, vertices, or parts) and the relations between them. It is described here in some detail as it is the basis of the hybrid model of object recognition that is the focus of this thesis.

Consider a familiar object such as a teacup. Its structural description would consist of two parts, the body (a straight cylinder) and the handle (a smaller curved cylinder), and one relation ("side-attached"). The structural description remains the same across many changes in viewpoint, such as when we see the mug at a different place in the visual field (translation), further away or closer (scale), or if we see its mirror image. Even for many rotations in depth the structural description is stable, as long as the two parts are visible. Such a representation must bind the shape attributes straight cylinder and curved cylinder with the relational attribute "side-attached", otherwise the structural description would be indistinguishable from a bucket, an object with very similar parts, but different relations ("on-top " or "below" of) between them. The challenge for structural description models is to find a solution to the binding problem, that is to establish how attribute conjunctions are represented.

One solution to bind attributes (parts and spatial relations) in a structural description is to use separate units for each conjunction. Hypothetical units proposed in connectionist networks to represent the body of a cup as straight cylinder would differ if on one occasion the curved cylinder appeared "on-top" of the cone (such as when the cup is rotated 90º), seen alone, or in some other relation. Such a solution to binding parts and their spatial relation is called static binding (Hummel, 2001), that is separate units are pre-dedicated for each conjunction. Static binding assigns a unit to respond to cylinders "side-attached" to other parts, but another unit might respond to cylinders "on-top" of other parts, and so forth. Although static binding offers a solution to the binding problem, it has some serious shortcomings. The number of units required to pre-code all possible part–relation conjunctions would be prohibitively large, because it would grow exponentially with the number of relations. More importantly, static binding cannot convey the similarity structure between objects, because it does not allow the independent binding of attributes. Thus, for example, in static binding a cylinder "side-attached" to a cone is not more similar to a cylinder "on-top" of a cone than to a cube "on-top" of a cone because in these representations each part–relation binding is coded by a separate unit. The problem of keeping the similarity structure is a fundamental property of static binding (Hummel, 2001; Hummel & Biederman, 1992).

Hummel and Biederman (1992) pointed out that the shortcomings of static binding make it a less desirable solution for the problem of binding object attributes. Rather, representing structural descriptions in a connectionist architecture requires a mechanism for binding attributes dynamically - the binding of attributes must be temporary and explicit so that the same units can be used in multiple conjunctions. For example, one unit represents cylinders and another might represent the "side-attached" relation; a cylinder "side-attached" to another part would be represented by explicitly tagging these units as bound together (rather than implicit conjunctions as in static binding). The same units can be used in different conjunctions at other times because these tags are assigned in a dynamic fashion. As a result, the similarity between different objects with the same parts (represented by the same units) is preserved.

Hummel and Biederman (1992) presented a neural network model of object recognition (JIM; John and Irv's model) in which attributes of an object such as parts and part relations are dynamically and explicitly bound together. Dynamic binding is implemented by using the concept of temporal synchrony, in which conjunctions of attributes can be established by synchronising the outputs of the units (or cells) representing those attributes (Milner, 1974). Given the image of an object, the units representing one part (or geon) and its relation fire in synchrony with each other, but out of synchrony with units representing other parts and their relations. JIM consists of a seven-layer connectionist network that takes as input the image of a line drawing of an object and, as output, activates a unit representing the identity of the object. The same identity output unit will respond no matter where the object's image appears in the visual field, the size of the image, and the orientation in depth from which the object is depicted (provided no part is completely occluded).

The first layer (L1) in the model is an array of orientation-tuned cells with overlapping receptive fields. Its cells respond to image edges in terms of their location, orientation, curvature (straight vs. curved), and whether the edges terminate within the cell's receptive field (termination cells) or pass through (segment cells). The second layer (L2) contains cells that respond to vertices (such as Ls, arrows, forks, and tangent Ys), 2D axes of symmetry, and oriented blobs which give information about a geon's, size, centre of mass, and elongation. Together, cells in L1 and L2 describe geons by synchronising oscillations in their outputs. Units activated by contours and vertices of one geon are induced to fire in synchrony with each other, and units relating to a different geon fire also together in synchrony but out of phase with the former geon. How does the model achieve this temporal synchrony/asynchrony?

Hummel and Biederman (1992) propose a mechanism by which certain constraints force units belonging to the same geon to fire together. Two cells can synchronise their outputs by sending activation signals over a fast enabling link (FEL). A unit that receives input from an FEL will reset its refractory state below its threshold and will fire if it is in an active state. Because in the model signals are sent along FELs with almost infinite speed, the refractory states of the two units will covary and they will fire in synchrony. Two units should only be connected by an FEL if the features they represent are likely to belong to the same geon. This is established if certain conditions are fulfilled for example, if two units represent image edges of the same curvature, (approximately) the same orientation, and have overlapping receptive fields. Thus, FELs are designed to link units together representing features which are likely to cooccur in the real world, for instance image contours that coterminate in a L- or Y-vertex are likely to belong to the same geon.

The model's first two layers represent the local features of an image, which are parsed and grouped into temporal sets corresponding to geons. Layer 3 (L3) responds to the attributes of complete geons (such as shape of its major axis, shape of the cross section, parallel or nonparallel sides, etc) to describe geons such as cones, cylinders, and so forth. Thus, the model's first three layers segment an image into its constituent geons by parsing the local features of geons into separate groups. Layer 4 and Layer 5 derive relative size, relative location, and relative orientation from the attributes size, location, and fine orientation of layer 3. Layers 4 and 5 are used to describe the categorical relations among geons explicitly and bind them to the geon they describe. The relations must be invariant with geon identity and viewpoint meaning that a "below" unit will always fire when one geon is below another. Layers 3 and 5 combined produce a pattern of activation that describe a geon in terms of its shape, general orientation, location, size, and relative orientation. Several of these geon feature assemblies (GFAs) collected over different time slices for a given object result in a structural description in Layers 6 and 7.

Layer 6 receives the output of Layers 3 and 5 to form the structural description of an object. Each cell in L6 responds to a particular geon-relation conjunction. If an object is in JIM's memory, then each of its GFAs will activate a different cell in L6. The activation of L6 is summed up over time in L7, combining two or more assemblies into a representation of a complete object, which is invariant with translation, scale, and orientation in depth.

Hummel and Biederman tested their model in computer simulations demonstrating that, after training with one view of an object, the recognition performance was invariant with changes in depth-rotation, size, and location as well as after mirror reflections. Moreover, as in human observers, recognition performance dropped with rotations in the picture plane. Thus, the model's recognition performance reflects properties of human object recognition concerning these manipulations of changes in viewpoint.

1.4.3.4 Evidence for Part-Based Models

There are intuitive reasons to believe that human observers store object descriptions in terms of parts and categorical relations as Hummel and Biederman (1992) claim. People often list the parts of the object when they are asked to describe the characteristics of basic-level objects, and object's parts often provide much of the functionality of an object class (Rosch et al., 1976; Tversky & Hemenway, 1984). Tversky and Hemenway (1991) noted that familiar subordinate exemplar objects are also distinguished by their parts. These observations are backed up by more direct evidence in support of JIM. In particular, data have been provided that demonstrate the role of geons in defining parts.

Biederman (1987) showed that human recognition performance varied greatly depending on which - rather than how much - of the contours in a line-drawing were deleted. Contours of line drawings of common objects were deleted that either corresponded to geon vertices (unrecoverable contour deletion) or the same amount that was uncritical for geon extraction (recoverable contour deletion). As predicted, recognition performance was only mildly affected by the latter manipulation, whereas performance dropped significantly for unrecoverable contour deletion.

In a similar vein Biederman and Cooper (1991b) showed observers pairs of complementary contour-deleted line-drawings of common objects that would produce an intact original line drawing if superimposed. In these line-drawings every other vertex and line was deleted from each part, so that the geon structure could be recognised from either member of a pair. In the study phase participants named either the identical image or its complement version. There was no difference in priming between identical and complementary conditions, indicating that the activated representation was not based on lower level features (lines and vertices), but corresponds to a more abstract part-based representation of the object

(Biederman & Cooper, 1991b). In contrast, images formed by deleting half of the parts did not visually prime their complements. These results indicate that priming was mediated by representations based on parts rather than particular vertices and contours present in the original image.

Further evidence that parts may be coded as suggested by RBC theory has been presented by Stankiewicz (2002) who showed that observers process an geon's aspect ratio and primary-axis curvature independently. Boutsen and Marendaz (2001) also demonstrated the importance of axis information in determining part structure. They investigated the often reported effect of orientation search asymmetry in a visual search task (i.e., a faster detection of a tilted target among vertical distracters than the reverse) for the global orientation of 2D polygons with a salient, "principal" axis of symmetry. Their results show that search asymmetry depends on the orientation of the principal axis, rather than on the orientation of local contours, which means that the perception of the global orientation of simple shapes is mediated by descriptions in terms of axes of symmetry and elongation.

An important prediction of RBC is the role of non-accidental properties (NAPs) of 2D image contours for the extraction of a geon structural description. This was tested by showing observers different versions of a standard object: one version with a NAP change (e.g., wineglass vs. champagne glass), and another version in which the aspect ratio was changed (e.g., elongation of the wineglass), which was a slightly bigger metric change than the former manipulation (Cooper & Biederman, 1993). Consequently, the difference between the aspect ratio change and standard images was more readily detected than the difference between the NAP change and the standard image when participants performed a simultaneous physical identity matching task. However, in a sequential object matching task the NAP change resulted in far greater disruption than a change in a metric property. Thus, object memory is more sensitive to qualitative changes in the part structure than to perceptual changes.

Since the basic assumption of RBC is that the detection of individual geons is viewpoint-invariant due to non-accidental properties, Biederman and Gerhardstein (1993) tested their predictions about conditions of viewpoint-dependency with geons and geon structured objects. In one experiment, participants had to identify single target-geons among a set of distracter geons. No effect of changes in orientation was observed for the geons, confirming the prediction of view-invariance. In another experiment, line drawings of common geon-built objects were used that met the conditions described above (see section 1.4.3.3). Participants named a set of objects in a study phase. In the test phase, participants were shown same exemplars or different exemplars of the study-objects in either the same viewpoint or at novel orientations. Biederman and Gerhardstein (1993) observed that recognition performance was affected by different exemplars but not by changes in viewpoint. In a further experiment, condition 3 (identical GSDs) was tested by rotating the objects around 45°. In one condition, this rotation changed the configuration of visible parts compared to the trained viewpoint. In the second condition, the same amount of rotation left the parts of an object visible that could be seen in the training viewpoint. As predicted, slower response times were observed in the parts-change condition, whereas no effect of rotation was found when there was no change of visible parts. The same result was obtained for unfamiliar geon-structured objects. In another experiment, Biederman and Gerhardstein (1993) substituted a different geon for each centre segment of a set of 10 line drawings of bent paper clips, a stimulus set that usually produces large view-dependent effects. The addition of the distinctive geon dramatically reduced rotation costs.

Biederman and Gerhardstein (1993) concluded that whether a change of orientation of an object yields slower recognition performance depends on whether the rotation produces a change in the geon structural description of that object. The often observed monotonic effect of orientation disparity in reaction times may therefore not necessarily be attributable to viewpoint-specific templates (e.g., Edelman & Bulthoff, 1992; Tarr & Pinker, 1989). Rather, violations of one of the criteria of object-invariance are likely to elicit "unsystematic selection of stimuli" (Biederman & Gerhardstein, 1993, p. 1180). In particular, when parts of objects are occluded or novel parts are revealed because of rotation, response times are likely to increase.

Apart from the evidence for geons in object recognition, there is further support for critical assumptions of RBC from the coding of categorical relations, the dynamic binding of attributes in object recognition, and the role of attention in selecting parts of objects. Concerning categorical spatial relations, there is recent evidence that the orientation between an object's parts is coded categorically as parallel, perpendicular or oblique similar to JIM. Rosielle and Cooper (2001) showed observers line drawings of novel objects in which the relative orientation of object parts varied by steps of 30°. Participants found it easier to judge that two objects were physically different when their parts had a different categorical orientation relationship (e.g., parallel 0° vs. oblique 30° or oblique 60° vs. perpendicular 90°) than when the parts of the two objects had the same relative orientation relationship (30° vs. 60°). In an object recognition task participants found it easier to classify objects that belong to the same category (i.e. decide if two objects shared the same set of parts) when the relative orientation of the parts did not cross a categorical boundary (objects with 30° vs. 60° relative orientations between their parts were more readily classified as belonging to the same class). Rosielle and Cooper (2001) conclude that relative orientation is coded categorically for both object recognition and physical discrimination.

A further crucial element of the model of Hummel and Biederman (1992) is that its structural description relies on binding mechanisms based on temporal synchrony. Conjoining features via temporal synchrony is time consuming and capacity limited and requires visual attention (Logan, 1994; Luck & Vogel, 1997; Treisman & Gelade, 1980). For example, Luck and Vogel (1997) demonstrated that observers typically retain information about only four features (e.g., colour or orientation) in visual working memory at one time. However, it is possible to retain both the colour and the orientation of four objects, suggesting that visual working memory stores integrated objects rather than a collection of isolated features. Furthermore, objects defined by a conjunction of 4 features were found to be retained in working memory just as well as objects defined by single features. Thus, 16 individual features could be retained when conjoined to four objects. Visual working memory seems to have a limited capacity for integrated objects that consist of dynamically conjoint features.

Concerning attentional processes, there is recent evidence for the role of part selection in visual attention (Vecera, Behrmann, & Fliapek, 2001; Vecera, Behrmann, & McGoldrick, 2000). In experiments by Vecera et al. (2001), participants were more accurate in reporting the attributes on the same part than attributes on different parts of a single object. This part-based effect was not influenced by the spatial distance between the parts. Their results suggest that visual attention may selectively process the parts of an object.

There is also neurophysiological evidence for part-based representations. Macaque inferior temporal (IT) neurons are highly shape selective and different neurons show different shape preferences. Tanaka (1993) demonstrated that these preferences can be elicited quite strongly to features of 'moderate complexity', similar to geons (Biederman, 2000). Vogels and colleagues (Vogels, Biederman, Bar, & Leuven, 1999; Vogels, Biederman, Bar, & Lorincz, 2001) tested macaque IT (area TE) to determine if cell activity was modulated differently by a geon change rather than a metric change. Although metric changes produced slightly greater image changes, a greater modulation in IT cell activity was found after geon (NAP) changes.

1.4.4 Evaluation of View-Based and Part-Based Models

Not surprisingly, the highly specified model of geon-based object recognition has some limitations. These mainly concern the extraction of contours from images, the extraction of geons from contours of a 2D image, the recognition of natural objects, effects of depth-rotations, and the need for attention and speed of processing. Although only the problems associated with the GSD approach will be considered in this section, many points implicitly address Marr's model as well.

One main assumption of RBC and JIM is that the visual system extracts edges from an image in such a way that it corresponds to a line-drawing of an object (see 1.4.3.3). A single view of that line drawing and the use of volumetric primitives allows view-invariant recognition. However, there are difficulties in extracting edges from real-world images in uncontrolled imaging situations (Dickinson et al., 1997; Edelman, 1997). Line drawings are idealised versions of the original edge information, and irrelevant edges (e.g., resulting from shading) are often omitted (Sanocki, Bowyer, Heath, & Sarkar, 1998). So far, no reliable algorithm has been found in computer vision that would allow the extraction of edges that only correspond to contours relevant for the representations of geons. Furthermore, Sanocki et al. (1998) found that object recognition was worse with line drawings than with colour photographs, and performance was even worse when other objects were present. A similar difficulty concerning the extraction of geon arises for computing the axes of symmetry. In JIM, this problem is circumvented by providing the model with a representation of the axes as part of the input (Hummel and Biederman, 1992). Finally, some geon attribute contrasts included in Biederman's (1987) RBC theory are not discriminated by JIM (e.g., whether a geon with nonparallel sides contracts to a point or is truncated).

A further problem is that the recovery of geon-based descriptions of natural objects is very difficult with the RBC approach. Many natural objects (such as trees, clouds) are not readily decomposable into constituent parts because they are highly complex or irregular (Edelman, 1997). Nevertheless, Biederman (Dickinson et al., 1997) pointed out that it is not necessary to extract parts from the image of a natural image such as an oak tree, because observers do not readily distinguish it from other oak trees. Thus, observers may not require the exact shape to make classification judgements for a natural object, and they do not necessarily generalise over variations in viewpoint for these types of objects. The shape variations that are not captured by geon theory may therefore correspond to distinctions that are hard to make for observers. Moreover, JIM may be able to represent an object such as a bush in terms of a texture description, based on low-level (layer 2) blob and contour units.

A more serious limitation of Biederman's structural description approach concerns the solution to the binding problem. The binding mechanism of JIM — correlated firing of units associated with geon properties, geons, and relations — may be too slow to produce the 100 ms real-time object and scene recognition that is so evident in human performance (Intraub, 1981). The model of Hummel and colleagues (Hummel, 2001; Hummel & Stankiewicz, 1996) described in a later section addresses this problem. Moreover, FELs in JIM operate with infinite speed, allowing two cells to synchronise instantly. In reality, there will be some time cost due to the propagation of an enabling signal across a FEL.

Apart from the admitted limitations of the RBC approach, there have been some experimental observations that challenge its predictions. RBC predicts view-invariant recognition of geons and geon-structured objects, as long as relations and parts are not changed and the conditions for view-point invariance are met (Biederman & Gerhardstein, 1993). However, some studies report recognition costs for objects rotated in depth and even for single geons (Hayward & Tarr, 1997; Tarr, Williams, Hayward, & Gauthier, 1998). There is an ongoing debate on the significance and the predictions of view-dependency across depth-rotations (Biederman & Bar, 2000; Biederman & Gerhardstein, 1993, 1995; Hayward & Tarr, 1997; Tarr, 1995; Tarr & Bulthoff, 1995).

As an argument against studies that report viewpoint-dependent performance, Biederman and Gerhardstein (1993) claimed that the activation of a representation of an object may be invariant even when performance sometimes does not exhibit this invariance. Biederman and colleagues (Biederman & Cooper, 1992; Biederman & Gerhardstein, 1993, 1995) argued that object recognition tasks such as matching may tap episodic recognition memory that involves a separate object processing system in addition to the system mediating object identification. For example, Biederman and Cooper (Biederman & Cooper, 1992) showed that object recognition was size invariant with a naming task, but not with an old-new recognition task, presumably because the episodic event of seeing an object includes its particular size, position, and viewpoint. According to this argument, object naming is mediated solely by an object identification system that is associated with the occipito-temporal pathways in the brain, whereas object matching (or old-new recognition tasks) may require access to exact spatial co-ordinates represented in occipito-parietal (dorsal) pathways (Ungerleider & Mishkin, 1982). Thus, reaction times may differ across view changes of the same object because latencies not only reflect the activation of object representations in the ventral pathway but are "contaminated" by episodic non-recognition processes. Evidence for such effects comes from priming studies using explicit and implicit tasks (Cooper, Schacter, Ballesteros, & Moore, 1992). One explanation favoured by Biederman and Gerhardstein was that the latter tasks depends on "feelings of familiarity that were influenced by both dorsal and ventral systems" (Biederman & Gerhardstein, 1993, p. 1163), whereas naming only includes ventral system representations. Similarly, mental rotation tasks in which handedness of rotated objects has to be determined (Shepard & Cooper, 1982) are deemed to be controlled by dorsal system processing. Thus, Biederman and Gerhardstein concluded it is necessary to assess the nature of object identification tasks and use experimental paradigms that do not rely on feelings of familiarity.

As another explanation for cases in which viewpoint-dependent effects are found, Biederman and Gerhardstein (1995) claimed that individuals may use viewpoint-dependent processes, such as searching for a distinctive GSD at a small scale. For example, one may look for a logo or brand name to distinguish different makes of similar watches or cars. Although the process of finding a GSD in such cases would be viewpoint-dependent and time-consuming, the shape representation could still be viewpoint-invariant. And finally, Biederman and Gerhardstein argued that effects of rotation or viewpoint may occur if objects are shown in accidental views (that yield ambiguous identification) or if objects are partially foreshortened or occluded (see Biederman and Gerhardstein, 1995).

Not surprisingly, there are also problems for view-based models but for different reasons to the part-based models. These severe limitations common to the view-based approaches include the role of contours and parts, translation and scale invariance, and generalisation across shapes.

In template matching, successful recognition depends on the number of points that match or mismatch between an image and a stored view, but there is no prediction concerning the locus of mismatch (an exception can be found in Ullman, 1989). However, Biederman (1987) showed that human recognition performance varied greatly depending where in the image contours have been deleted (vertices or line segments). Biederman and Cooper (1991b) demonstrated that an image in which half the vertices and edges from each of an object's parts were removed visually primed its complement (the image formed from the deleted contours) as much as it primed itself. In contrast, an image constructed by deleting half of the parts did not visually prime its complement. These results imply that there is a qualitative difference between contours - visual priming was predicted by the contours critical for the recognition of an object's parts, not just by the amount of contour.

As outlined earlier, human object recognition seems fairly invariant with retinal changes such as translation and size (Biederman & Cooper, 1991a, 1992). Most view-based models predict progressively poorer generalization — in the form of either weaker neural responses or diminished recognition performance — with increasing distance between a test view and any known view of an object. In addition, to computationally establish this kind of object invariance is no trivial problem for view-based models of object recognition (Riesenhuber & Poggio, 2000; Tarr, 1999). In contrast, concerning Ullman's model of view alignment, it is not clear whether it can account for the pattern of view-dependent recognition performance, because the alignment technique seems too powerful. As soon as a rotation angle has been computed, alignment can be performed in a single step (Edelman & Weinshall, 1998).

A further problem for object recognition theories based on whole views is that of image transformations concerning only a part of the object, as when a part is missing or an irrelevant part is added. These global transformations should in many cases lead to a total failure in recognition or classification, but this has not been observed in tests of recognition performance with such images (Biederman, 1987). Similarly, humans recognise a running dog as well as a sleeping dog, or a phone with the handle on or off the hook. Although the shapes projected in these different situations are radically different, they are mapped onto a common representation in memory - dog or phone. By assuming holistic template representations view-based models cannot easily explain a mechanism to compensate for these transformations in shape.

One common criticism of models that rely on view-based representations (Bulthoff & Edelman, 1992; Bulthoff, Edelman, & Tarr, 1995; Poggio & Edelman, 1990; Tarr, 1995) is that although image-based approaches allow specific recognition discriminations on a subordinate level (e.g., between a collie and a terrier), visual recognition predominantly taps the basic level (Biederman, Subramaniam, Bar et al., 1999). However, unlike RBC theory and its variants, most view-based models of human visual recognition cannot provide well-specified mechanisms for class-level recognition or further categorisation (Edelman, 1997; Hummel, 2001). Finally, a fundamental difficulty with view-based models is that they require stored representations before they can match two views of an object. View-based theories have, therefore, difficulties to account for viewpoint invariance with unfamiliar (novel) objects (Hummel & Biederman, 1992). Although the view-combination model requires only a small number of stored views to generalise over a large range of novel views, it still requires a sufficient number of representative views for each individual object. In contrast, the human visual system is able to obtain substantial generalisation on the basis of a single view of a novel object (Tarr & Gauthier, 1998). To address this problem, Ullman (1998) describes how such generalisations can be classed-based, that is, how generalisation to a new viewing direction can be obtained from a single image of a novel object on the basis of three other objects in the same class (e.g., faces). Still, it is unclear how the visual system would use the same process (linear combination of views) to both distinguish and generalise objects from the same class for example, a VW beetle from a (similar in outline shape) Saab.

The common problem for view-based approaches is that they represent object structure only implicitly (Edelman, 1997; Hummel, 2001). They have therefore difficulties to generalise to new views of novel objects, new views of familiar objects outside their training space (Hummel, 2000), and categorise objects in terms of their shape similarities.

Recently, view-based theorists have extended their models to incorporate properties reminiscent of feature matching (Edelman, 1998; Riesenhuber & Poggio, 1999). Feature matching models extract diagnostic features from an image for recognition (Selfridge & Neisser, 1960). These models use higher order image attributes such as parts, surfaces, contours, or vertices. The identity of an object is established in terms of multiple independent attributes and objects are seen as similar and subsumed under the same category if they share enough common visual features.

Feature matching has some theoretical advantages over template matching. In contrast to template models, feature models can compensate for transformations of object shape because they may code parts of an image. Furthermore, they can achieve view invariance by comparing features in different views of an object. However, the features used in such models are not encoded relative to one another in terms of their positions in space — they can be combined, but without reference to their location in the image; this property has attracted criticism (e.g., Hummel & Biederman, 1992). The problem is that location independence makes the representation insensitive to the spatial configuration of the features; thus, any configuration of the appropriate features could produce recognition of the object. As a consequence, even a scrambled version of an intact image that retains all features depicted in the original should produce equivalent recognition, which it does not. This point will be revisited in the General Discussion.

In summary, both view-based and part-based accounts have limitations. But it is also clear that both have important contributions towards our understanding of object recognition. Template matching models recently enjoyed considerable popularity in computer vision (e.g., Edelman & Poggio, 1990; Lowe, 1987; Ullman, 1989) as well as in behavioural studies on vision (Tarr, 1995). Nevertheless, many researchers in the computer vision community (Bergevin & Levine, 1993; Dickinson, Pentland, & Rosenfeld, 1992) and in psychology (Humphreys, Riddoch, & Quinlan, 1988; Sartori, Miozzo, & Job, 1993) assume a structural description stage in models of object knowledge. Indeed, RBC (JIM) remains one of the most detailed versions of this class of object recognition models (Dickinson et al., 1997; Tarr, 1995). Naturally, there are attempts to combine part-based and view-based models into hybrid accounts which will be described in the next section.

2. Chapter 2: Hybrid Accounts of Object Recognition

2.1 Introduction

As the above review on the two main classes of object recognition theories shows, both approaches offer theoretical advantages as well as disadvantages. Similarly, the data so far do not appear to clearly favour one family of object recognition theories. A number of researchers therefore attempted to formulate theories that incorporate properties of the view-based as well as part-based approaches to object recognition.

The following sections will give a short overview of hybrid approaches before turning to the hybrid model of object recognition by Hummel (Hummel, 2001; Hummel & Stankiewicz, 1996a) which is the basis for the present investigation. These approaches are broken down into those that stress the role of process or representation, although in many cases this distinction is not absolute.

2.2 Accounts Stressing the Role of Process

2.2.1 Mental Rotation

2.2.1.1 Dual Route Account (Jolicoeur, 1990)

Jolicoeur (1990) proposed that two functionally separate systems are working in parallel to identify disoriented patterns. These two systems are a mental rotation and a feature-based system. The former uses mental rotation to transform disoriented patterns by aligning them to the retinal upright and then matches the pattern with stored orientation-specific representations. The shortest path for aligning the input image is determined by a scheme similar to that proposed by Huttenlocher and Ullman (1987). According to Jolicoeur, the representations activated by the mental-rotation system are functionally spatially isomorphic to the stimuli. The feature-based system extracts shape attributes and surface attributes of objects which can be either orientation invariant or orientation sensitive. In both cases, Jolicoeur argues that effects of view changes on performance should be much smaller than those elicited by the mental-rotation system. An example of an orientation sensitive attribute would be symmetry, which is detected faster when the axis of symmetry is vertical than in other orientations (Jolicoeur, 1992). Thus, even if mental rotation is not used, recognition performance still may show some effects of orientation. Very small or no orientation sensitivity is found when the feature-based system is able to identify an object by extracting isolated features. Such orientation invariant attributes include texture, colour, size, particular types of line intersections, and global properties such as complexity, smoothness, etc. Since the mental-rotation system and the feature-based system work in parallel, both can yield a match to stored representations. Jolicoeur argues that his dual-systems model is able to account for studies that found mental rotation effects as well as for experiments in which no or rather small orientation sensitivity was observed (Lawson & Jolicoeur, 1998). An obvious weakness of this account is its low predictive power because the nature of the proposed representations appears highly underspecified. Furthermore, as discussed in a previous section (1.4.2.2), there is evidence that identification of disoriented images does not involve mental rotation (Jolicoeur et al., 1998; Lawson & Jolicoeur, 1998).

2.2.1.2 Double Checking (Corballis, 1988)

Corballis and his colleagues (Corballis, Zbrodoff, Shetzer, & Butler, 1978) found orientation dependency in a naming task with alphanumeric characters only in the condition where mirror-reversed stimuli were presented. Hence, Corballis (1988) concluded that mental rotation is usually not required to recognise misoriented objects, and that observers recognise familiar objects by extracting a description of a shape that is independent of any co-ordinate system. However, such a description would not allow the discrimination of mirror images, in which case observers have to mentally rotate a stimulus. In addition, such a description could not explain the view-dependent recognition performance for plane-rotated objects (Jolicoeur, 1985). Corballis (1988) proposed a "post access" process of mental rotation. According to this line of thought, initial access to stored representations is orientation-independent, since "mental rotation" of a misoriented object along the shortest path requires previous recognition. Once an object is identified, its internal axes are retrieved from long term memory and its orientation can be determined. After the initial access a normalisation process aligns the viewed object with the representation to verify the initial recognition or to distinguish it from its mirror image. According to Corballis, shapes may be encoded in a frame-free description consisting mainly of orientation invariant shape attributes. Information about the orientation of the pattern would be established after its identification and might be used occasionally merely to verify the identity of an already identified pattern. However, there is recent evidence that object identification does not employ mental rotation (Jolicoeur et al., 1998; Lawson, 1999). A more direct test of Corballis' "double checking" account involved unspeeded verification of briefly presented pictures (Lawson & Jolicoeur, 1998). Effects for plane rotations were found even in unspeeded conditions and did not attenuate with practice as often observed with effects on naming latencies. According to Lawson and Jolicoeur, these stable effects of rotation should not occur if objects are recognised irrespective of orientation. If mental rotation effects in latencies are due to "double checking" verification accuracy should not be affected in an unspeeded task. Therefore, a complete view-dependent recognition process in combination with orientation effects due to double-checking seems neither parsimonious nor supported by the data (Lawson & Jolicoeur, 1998, 1999).

2.2.2 Holistic and Analytic Processing

Farah (1990; 1991) has developed an account of object recognition that makes a distinction between two types of processing: holistic vs. analytic. In fact, her account is not exclusively one derived from process differences as these have a direct relationship to the representations used in object recognition. Moreover, Farah proposes a modular organisation of mental representations with separate systems for faces, objects and words.

Farah (1990) reviewed case studies of neuropsychological patients that showed specific impairment for word (alexia), object (visual agnosia), or face recognition (prosopagnosia). She found that there was strong evidence for a double dissociation between prosopagnosia and alexia and proposed that two major recognition capabilities are susceptible to neuronal damage. The first one is based on holistic analysis of stimuli that is mainly responsible for processing the overall configural structure of an object. The second system analyses visual stimuli in a part-based fashion. According to Farah (1990), patients with deficits in word recognition suffer from an impairment to the part-based or analytic system because letters and some objects need to be decomposed into parts to recognise them. In contrast, face recognition relies mainly on holistic processing. Therefore, damage to the holistic processing component should impair performance for processing faces. The recognition of objects would entail both holistic as well as analytic processes. Farah et. al. (Farah, Wilson, Drain, &

Tanaka, 1998) found psychophysical evidence for holistic processing of faces. They showed that a face matching task was more strongly impaired by an intervening mask consisting of a whole face relative to a mask consisting of scrambled parts of a face. A similar manipulation in an experiment where houses and words had to be matched produced a lesser or no disruption indicating that holistic processing seems particularly important for faces (for a critical review, see Humphreys & Rumiati, 1998). Although Farah proposed different types of representations, the notion of analytic versus holistic processes is more critical for the present investigation and will be addressed in subsequent sections.

2.2.3 Process Determined by Task-Demands

Tarr and Bulthoff (1995) proposed combining view-based and part-based approaches to object recognition in an approach that stressed the role of task demands. They suggest that human object recognition can be thought of as a continuum between pure exemplar-specific discriminations and categorical discriminations. According to this line of thinking, extreme cases of within-class discriminations allow for recognition exclusively achieved by viewpoint-dependent mechanisms. When objects are to be distinguished in broad categorical classes, recognition of objects may be exclusively achieved by viewpoint-invariant mechanisms. Hence the continuum reflects a "trade-off between efficiency of a representation and efficiency of recognition" (Tarr & Bulthoff, 1995, p. 1503). Shape discriminations usually fall within the extremes of the continuum and recognition is mediated by the viewpoint-dependent and the viewpoint-independent mechanisms according to the nature of the task, the similarity and familiarity of the stimuli, and other context conditions.

Although this account is plausible, its predictions are rather general and the experimental evidence is somewhat unclear. Supporting the task-demands argument, Hamm and McMullen (1998) found that picture-name matching of plane-rotated objects is view-dependent only on a subordinate rather than basic level. However, in a similar experiment with more distinctive entry-level categories than in the Hamm and McMullen study, Murray (1998) found that entry-level name-picture matching was highly view-dependent (sub-ordinate level naming was not tested). Dickerson and Humphreys (1999) also found view-dependent effects for naming plane-rotated objects on a basic level, which were accentuated when participants had to use subordinate-level names. Also, there were no recognition costs in superordinate level naming tasks, thus the pattern of performance was in agreement with Tarr and Bulthoff's (1995) prediction. However, in another test of the hypothesis that view dependency depends on task demands, Hayward and Williams (2000) showed that view-dependence and object discriminability of depth-rotated novel objects did not interact. The authors concluded that the degree of viewpoint-dependence is not a function of the ease of object discrimination.

Tarr and Bulthoff's (1995) proposal that subordinate-level recognition tasks are due to different processing demands compared to basic-level recognition tasks also does not necessarily contradict part-based accounts. Biederman et al. (1999) pointed out that exemplars of an object class may vary due to large geon or relation differences (e.g., a round table vs. a table with a square top), GSD differences on a small scale (e.g., distinguishing between different cars using their logos), and metric differences (e.g., discriminating between different drill bits). According to this logic, geon theory could also account for the findings of accentuated effects of plane-rotation changes for subordinate level tasks. Hummel and Biederman's (1992) GSD model JIM predicts a deterioration in recognition performance for plane-rotated objects. It is conceivable that looking for small geon differences affords a more detailed GSD processing. Analogue to Marr's (1982) idea of hierarchical part structures it is possible that structurally similar objects are parsed on a finer scale. This process may take longer for rotated objects as more geons and relations are perturbed compared to a geon description on a coarser scale. Therefore, accentuated view-point dependent performance as a function of task demands would be also predicted by a structural description system (Biederman et al, 1999).

2.3 Accounts Stressing Multiple Representations

There are a number of accounts that stress the importance of both global (presumably view-dependent) and more abstract (presumably view-invariant) representations in object recognition. Two of the earliest accounts (see section 1.3.1) by Posner (Posner, 1969; Posner & Keele, 1967) and Bartram (Bartram, 1976) concluded that matching of identical or rotated images involves different types of representations in object recognition: a view-specific representation that allows fast matching of identical views and more abstract representations that are employed when matching different views or different shape exemplars of an object. Ellis and his colleagues (Ellis & Allport, 1986; Ellis et al., 1989) extended Bartram's previous results by observing that the advantage for identical view matches over different view matches was only present at short ISIs (100 ms and 500 ms) but not at a long ISI (2,000 ms) or with an intervening mask. At the same time, the advantage for different view matches over different exemplar matches was maintained at long ISIs and in conditions with an intervening visual mask. These results were in support of Bartram's (1976) distinction between three types of representation in object recognition processes. First, a view-specific code mediates direct and fast identical view matches but dissipates quickly and is disrupted by masking. The second type is more abstract and mediates different view matches. This representation is derived more slowly but is less affected by visual masking. Finally, a semantic or name representation is involved when matching different exemplars of an objects.

There is also neuropsychological evidence in support of at least two different representations for objects: one that is view-specific and another one that is more abstract. Warrington and her associates (Warrington & James, 1988; Warrington & Taylor, 1973, 1978) tested neuropsychological patients at matching objects shown in a canonical view to a view in which important features were hard to extract or in which the object appeared foreshortened. Observers with damage to the right posterior areas of the brain were particularly poor at this test; therefore, Warrington and Taylor (1973; 1978) proposed that visual object recognition works in two main stages. In a first stage, perceptual object constancy is achieved, which relies heavily on right hemisphere processing. The second stage involves semantic categorisation, which taps primarily left hemisphere processing. Damage to the right hemisphere would therefore impair object constancy, so that only objects in highly familiar (canonical) views are recognisable which could then still get access to semantic categorisation. Indeed, patients with lesions to the right hemisphere are often reported having difficulty in recognising stimuli in unusual views, whereas objects in canonical views may still be recognised (Davidoff & Warrington, 1999; Warrington & James, 1988).

Somewhat different representations working in two parallel pathways were proposed by Humphreys and Riddoch (1984). They report five brain-damaged patients with deficits in achieving object constancy. The patients were asked to discriminate two different views of a target object from a photograph of a visually similar distracter object. Four of their patients with right-hemisphere damage were only impaired in this task when the principal axis of the target object was foreshortened in one of the photographs. These patients were unimpaired in matching views with a non-foreshortened principal axis. In contrast, the fifth patient (with damage to the left hemisphere) showed impaired matching only when the saliency of the target object's main distinctive feature was reduced, but foreshortening of the principal axis did not affect his performance. According to Riddoch and Humphreys (1984), this double dissociation indicates that two functionally independent routes are responsible for achieving object constancy, each of which may be damaged selectively. One route processes an object's local distinctive features whereas the second route encodes the object's structure relative to the frame of its principal axis.

Another more recent account based on hemispheric differences comes from Marsolek (1999). He proposed that the human visual system draws on two different subsystems in order to resolve contradictory demands of subordinate and superordinate classification of objects. An abstract-category recognition system, which is assumed to be dominant in the left brain hemisphere, serves the ability to map different input shapes to the same output representation for recognition in order to generalise across different exemplars of an object category. A second subsystem, the specific-exemplar subsystem, is thought to be working more effectively in the right hemisphere, and it is very sensitive to object shape and maps even slightly different shape exemplars to different output representations. Similar hemispheric subsystems specialised for metric and co-ordinate relations have been proposed by Kosslyn and his colleagues (Kosslyn, Chabris, Marsolek, & Koenig, 1992; Kosslyn et al., 1989). In Marsolek's theory, the abstract-category subsystem in the left hemisphere is assumed to represent features of objects independently, such as non-accidental properties (Lowe, 1985). This type of processing allows the visual system to generalise across different members of a category as they usually share a common subset of features. In contrast, the specific-exemplar subsystem in the right hemisphere processes object shape as a whole, that is features are not represented independently of each other. Evidence for the existence of different subsystems comes from studies that demonstrate abstract and form-specific recognition of word forms (Marsolek, Kosslyn, & Squire, 1992; Marsolek, Schacter, & Nicholas, 1996) and letterlike forms (Marsolek, 1995).

In further studies, Marsolek (1999) used shapes of common objects in repetition priming experiments. After encoding images of familiar objects that were shown centrally in a study phase the same images or different objects of the same name were presented briefly to one hemifield. Repetition priming was exemplar-abstract when presented directly to the LH but exemplar-specific when presented directly to the right hemisphere, confirming the notion about possible different subsystems underlying object recognition. An additional result was that priming for different exemplars was specifically visual in that different exemplars still elicited more priming than the written name of the object. The findings also suggest that both systems are working in parallel rather than in serial stages. A subsequent study (Burgund & Marsolek, 2000) also found that presenting the same object to the RH affected priming more than direct presentation to the LH, which is in line with Marsolek's argument that a more abstract visual subsystems is predominant in the LH that processes features of objects rather than whole shapes. In addition, Burgund and Marsolek found that priming was view-invariant in the left hemisphere, but view-dependent in the right hemisphere, indicating that two qualitatively different representations (or "subsystems") were activated.

One particular shortcoming of Marsolek's and other accounts discussed above is their lack of specification. In particular, it is not clear under what conditions the different representations are tapped separately or in combination. Also, the important role of attention for object recognition is mostly neglected. Therefore, the next section describes a hybrid account of object recognition that specifies the processes and representations that depend on visual attention.

2.4 The Hybrid Model of Object Recognition (Hummel, 2001)

2.4.1 Rationale for the Model

The review above showed that human object recognition reveals both view-based and part-based properties. According to Hummel (2001), this combination of properties together with the need of establishing object constancy over variations in shape is problematic for theories of object recognition that rely exclusively on the geometric properties of object shape — for example, by matching two-dimensional (2D) images to 3D models in memory (Lowe, 1987; Ullman, 1989) or by using mathematical interpolation to determine whether a given image is a "legal" projection of a familiar shape (Poggio & Edelman, 1990). A visual system that relied exclusively on the laws of projective geometry would be equally able to accommodate all variations in viewpoint (which the human visual system does not) but would not tolerate variations in object shape (which the human visual system does).

According to Hummel (2001), structural descriptions theories can better account for the properties of object recognition because in these models objects are visually represented by specifying their component parts in terms of their categorical relations (Biederman, 1987; Hummel & Biederman, 1992; Marr, 1982; Palmer, 1978). Using the previous example, a coffee mug might be represented as a curved cylinder (the handle) "side-attached" to a straight vertical cylinder (the body; see Biederman, 1987). Like human shape perception, this description is unaffected by translation across the visual field, changes in size, left-right reflection and some rotations in depth. However, it is sensitive to rotations about the line of sight (e.g., a 90° rotation changes the "side-attached" relation between the cup handle and body to an "on-top" of relation). The description also applies to many different mugs, permitting generalisation over metric variations in the shapes of different mugs.

One of the most important properties of a structural description is that it is an analytic representation, meaning that it specifies the relations among an object's parts both explicitly and independently. By contrast, representations based on metrically precise 2D or 3D models are holistic, in that they do not specify object features or parts independently of their location in the object as a whole (Hummel, 2000, 2001; Tanaka & Farah, 1993). Rather, a holistic representation of an object is "view-like" - the object structure is not represented explicitly but is based on a 2D coordinate system in which features are assigned to fixed locations (or coordinates) in a spatial reference frame. Therefore, elements in holistic view-like representations are not coded independently of their locations; they cannot uniquely specify object identity nor can they indicate similarity with another holistic representation containing the same element. The basic tenet of holistic representations in object recognition is that whole images are matched directly to image-like views stored in memory.

Consistent with the structural description account of shape perception, there is evidence that the visual system represents the relations among an object's parts both explicitly (Hummel & Stankiewicz, 1996b; Palmer, 1978; Tversky & Hemenway, 1984) and independently of the parts themselves (Saiki & Hummel, 1998). An important problem for structural descriptions concerns the way parts and their relations are bound into meaningful sets for example, straight cylinder with curved-cylinder for the description of a cup. One solution is dynamic binding, in which a single representational unit is used in different combinations. For example, one unit represents cylinders and another might represent the "side-attached" relation; a cylinder "side-attached" to another part would be represented by explicitly tagging these units as bound together (see section 1.4.3.3). As these tags are assigned in a dynamic fashion, the same units can be used in different conjunctions at other times. The second solution to the binding problem, however, called static binding, proposes that separate units are pre-dedicated for each conjunction. Thus, a unit might respond to cylinders "side-attached" to other parts, another might respond to cylinders "on-top" of other parts, and so forth.

To solve the dynamic binding problem some authors (Hummel & Biederman, 1992) propose synchrony of firing, that is to establish that two parts belong to one object (and are not combined with another part to form a different object) from the simultaneous firing of units that relate to parts which are grouped together. However, according to Hummel (2001), a theory of object recognition that relies solely on analytic descriptions based on dynamic binding faces a serious problem: dynamic binding is probably time-consuming and not free from errors, as units need time to synchronise (or desynchronise) their outputs. Also, units may fire in synchrony "accidentally" or bind parts together erroneously because of time constraints. Furthermore, some properties of shape perception are inconsistent with the properties of analytic representations. Binding independent visual dimensions (such as parts and relations) into a coherent analytic representation requires visual attention (Logan, 1994; Luck & Vogel, 1997; Treisman & Gelade, 1980). However, evidence of both negative priming of ignored images for overlapping stimuli (Murray, 1995b; Tipper, 1985; Tipper & Driver, 1988) and positive priming for spatially separated stimuli (Stankiewicz & Hummel, 2002; Stankiewicz et al., 1998) demonstrates that object recognition does not necessarily require visual attention. In addition, generating structural descriptions imposes a bottleneck on processing object information (Hummel, 2001). For example, in the Hummel and Biederman (1992) model geon attributes, geons and spatial relations are bound together by synchrony of firing. But both behavioural evidence (Intraub, 1981) and evidence from single-unit recording (Oram & Perrett, 1992) suggest that object recognition is too fast to depend on time-consuming analytic representations alone. Thus, it seems that dynamic binding is not sufficient to explain the observed efficiency and speed of human object recognition (Hummel, 2001; Hummel & Stankiewicz, 1996b). Object representations relying solely on structural descriptions that are based on dynamic binding seem unable to account for all the properties of human shape recognition performance. Thus, although representations in form of structural descriptions are needed, they must be complemented by representations that work fast and without attention.

2.4.2 The Hybrid Model

Hummel proposed a neural network model of object recognition: JIM.2 (Hummel & Stankiewicz, 1996a) and its successor JIM.3 (Hummel, 2001), both of which are based on the computational structural description model JIM by Hummel and Biederman (1992). In JIM.2 and JIM.3 object recognition is based on a hybrid analytic/holistic representation of object shape, with the analytic representation in form of a structural description, and the holistic representation resembling a more viewer-centred representation. In JIM.3 dynamic binding generates structural descriptions of object shape when an object is attended, but uses static binding to maintain the separation of an object's parts when an object is ignored. The model predicts that when the visual system segments an object successfully into its parts, shape perception exhibits the properties of a structural description. Recognition will be largely invariant with changes in viewpoint (Biederman, 1987), and part attributes will be represented independently of one another, and independently of the parts' interrelations. However, in case the visual system fails to segment an image into its parts (e.g., because of lack of attention or insufficient processing time), shape perception will exhibit the characteristics of a representation that uses static binding. Recognition will be more sensitive to variations in viewpoint, and part attributes will not be represented independently of their spatial relations, but rather in a crude viewer-centred reference frame.

The model will be discussed in some detail because it is the basis of the studies described in the following sections. JIM.3 consists of an eight-layer artificial neural network that can be trained to recognise line drawings of objects. Given the contours of an object's image JIM. 3 activates a representation of the object's identity as output (Figure 7). In the first three layers, units represent local image features such as contours (layer 1), vertices and axes of symmetry (layer 2), and the shape properties of surfaces (e.g., layer 3). The properties of each surface shape are categorically represented in terms of five categorical properties: (1) elliptical or not (i.e. bounded by a single smooth contour or bounded by more than one contour that converge at vertices); (2) possessing parallel, expanding, convex, or concave axes of symmetry; (3) possessing a curved or straight major axis; (4) truncated or pointed; (5) planar or curved in 3D. These categorical properties are derived from the vertices and axes of symmetry of a surface. In subsequent layers of the model these surface properties are used to extract the shape attributes of the corresponding geons.

Similar to JIM (Hummel & Biederman, 1992), the units in layers 1–3 fire in synchrony when they are activated by features of the same geon, and fire out of synchrony when the features correspond to different geons. Units in layer 4 gate the output of layer 3 to the two components in layer 5: the independent geon shape units (layer 5i) and the holistic surface map (layer 5s). In the model, the units of layer 4 are distributed spatially to cover the visual field. When synchrony can be established, the interactions between layers 3, 4, and 5 permit the model to dynamically bind features into parts-based sets. However, establishing synchrony (and asynchrony) of firing between units representing image features takes time. In the first moments after presentation (i.e., tens of ms for observers or during the first several iterations for the model) all the units activated by an image may fire at once even if they do not belong to the same geon. The model also assumes that inhibitory lateral interactions responsible for the asynchronous firing of features of separate geons require visual attention. Therefore, features of unattended objects will never group themselves into parts-based sets (Hummel & Stankiewicz, 1996a, 1998) and cannot be represented analytically. When attended, the synchrony relations established in layers 1–3 are preserved in layers 4– 6, where they are used to bind together the various shape attributes to form a geon and also to bind geons with their spatial relations.

The output of layer 4 is fed to layer 5 which is divided into two components (see Figure 7). The first component is termed the independent geon array (IGA) and is dedicated to represent geon attributes. The second component - the holistic surface map (HSM) - represents the shape attributes of an object's surfaces (corresponding to the shape attributes coded in layer 3). The IGA consists of a collection of units that represent the shape of a geon in terms of categorical attributes (Biederman, 1987): shape of its cross-section, shape of its major axis, whether its sides are parallel, expanding, convex or concave, and whether the geon is pointed or truncated (layer 5i). These shape attributes are coded independently of one another and independently of the geons' interrelations. Together, these units allow the coding of 31 different kinds of geons. For example, a cylinder has a round cross-section, a straight major axis, parallel sides, and is truncated. There are additional units that code a geon's aspect ratio (e.g., flat, intermediate, or elongated). Importantly, spatial relations units code whether a given geon is above, below, beside, larger-than, and/or smaller-than other geons in an object. Thus, the spatial arrangement of parts can be explicitly represented in the IGA.

The second component of layer 5 represents the shape attributes of an object's surfaces. The units of the HSM are arranged in a circular reference frame over 17 locations (layer 5s in the right-hand side of Figure 7). These units respond to the outputs of layer 3 (surface shape properties) routed via layer 4. The map is called holistic because each unit codes surface properties to certain locations in the map using static binding (see Hummel, 2000; Hummel & Stankiewicz, 1996a). Unlike in the IGA, the surface properties preserve their topological relations in the HSM. This means that surfaces that are in adjacent locations in layer 3 are mapped to corresponding adjacent locations in the HSM.

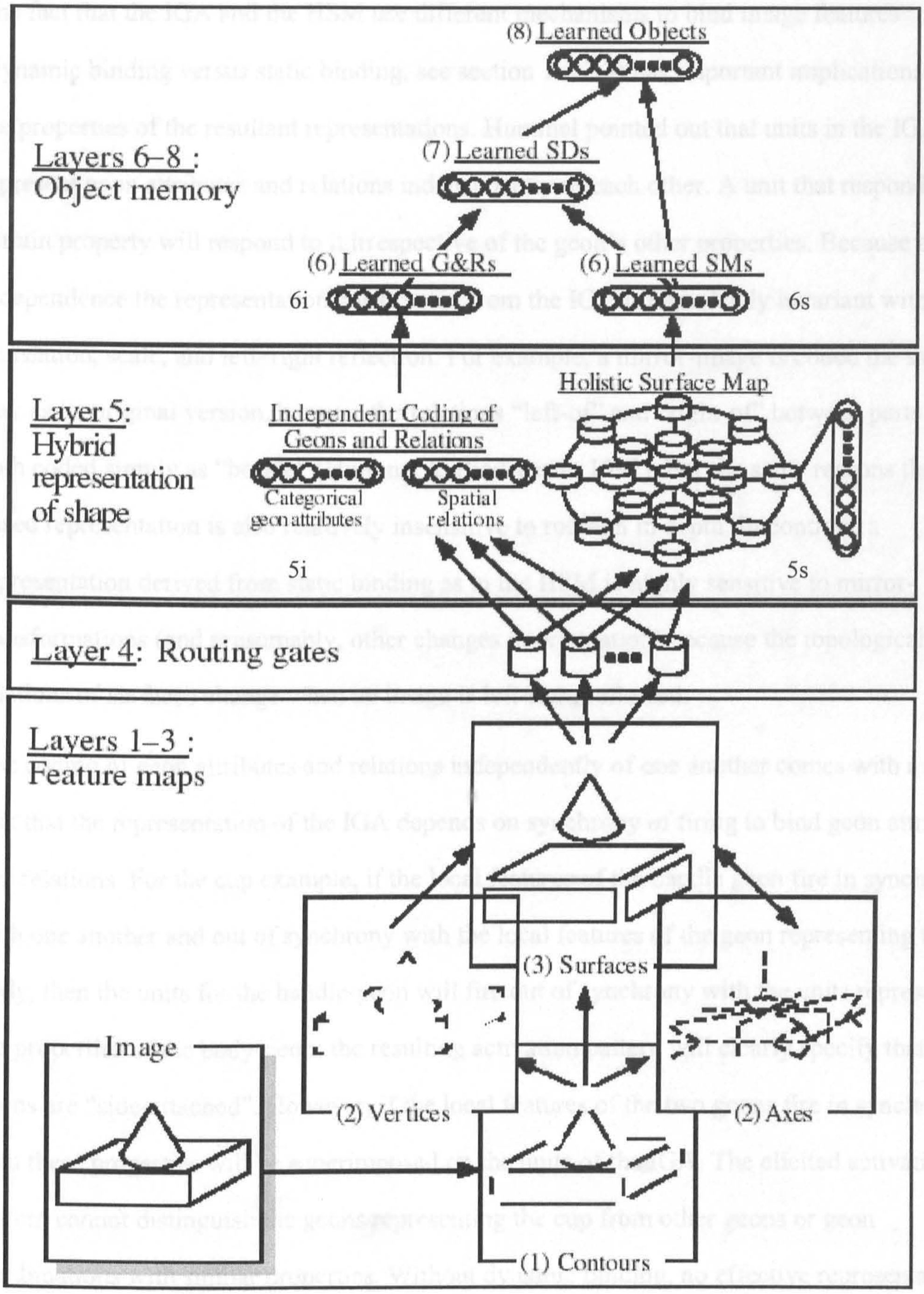


Figure 7: The architecture of JIM.3 (adapted from Hummel, 2001). Units in layer 1 are activated by the contours from an object's line drawing. Units in layer 2 represent vertices for coterminating contours and axes of symmetry between contours of the same surface. Units in layer 3 represent categorical shape properties of the object's surfaces. Routing gates in layer 4 propagate the output of layer 3 to units in layer 5 which has two components: The independent units represent the shape attributes of an object's geons, and the units in the surface map represent shape attributes of surfaces at each of 17 location in a circular reference frame. In layer 6, the activation patterns of both components are learned by individual units, which are then summed up in Layer 7 over time. Units in layer 8 code object identity.

The fact that the IGA and the HSM use different mechanisms to bind image features (dynamic binding versus static binding, see section 1.4.3.3) has important implications for the properties of the resultant representations. Hummel pointed out that units in the IGA represent geon attributes and relations independently of each other. A unit that responds to a certain property will respond to it irrespective of the geon's other properties. Because of this independence the representation constructed from the IGA is completely invariant with translation, scale, and left–right reflection. For example, a mirror-image is coded the same way as its original version, because the relations "left-of" and "right-of" between parts are both coded simply as "beside" (Hummel & Biederman, 1992). For the same reasons the geon based representation is also relatively insensitive to rotation in depth. In contrast, a representation derived from static binding as in the HSM is highly sensitive to mirror-transformations (and presumably, other changes in orientation) because the topological relations of surfaces change when an image is left-right reflected.

The coding of geon attributes and relations independently of one another comes with a cost. It is that the representation of the IGA depends on synchrony of firing to bind geon attributes and relations. For the cup example, if the local features of the handle geon fire in synchrony with one another and out of synchrony with the local features of the geon representing the body, then the units for the handle-geon will fire out of synchrony with the units representing the properties of the body-geon; the resulting activation pattern will clearly specify that the geons are "side-attached". However, if the local features of the two geons fire in synchrony, then these properties will be superimposed on the units of the IGA. The elicited activation pattern cannot distinguish the geons representing the cup from other geons or geon combinations with similar properties. Without dynamic binding, no effective representation will be established (Hummel, 2001; Hummel & Biederman, 1992)

In contrast to the independent geon attribute representation, the holistic surface map (HSM) is much less sensitive to errors resulting from dynamic binding. The units representing surface attributes in layer 5s are spatially separated. Therefore, when multiple geons fire simultaneously, the separation of the geon attributes is preserved because, like surfaces, different geons will be mapped to different locations in the HSM if they occupy different locations in the image. As a consequence, the representation formed on the surface map is sensitive to left–right reflections as well as to rotations in the picture plane and in depth. However, the HSM representation is invariant with translation and scale. Although the surfaces' topological relations are maintained in the mapping from layer 3 to the HSM, their absolute locations in the visual field and their size in the image are not. The units of the HSM are not confined to a particular location which allows them to "shrink-wrap" on a given object, no matter where it is in the visual field.

The properties of the surface map allow the model to recognise objects in familiar views even without dynamic binding. This is the case in the early moments after initial presentation of an input image. At first, the local features in the image (contours, vertices, axes, and surfaces) will tend to fire simultaneously because it takes time to establish synchrony of firing via lateral excitation and inhibition. The pattern of activation in the IGA will not be able to distinguish which properties belong to which geon in the image. However, the HSM keeps the geons spatially separate. Thus, even when the jumbled representation of the IGA does not allow recognition, the holistic representation generates an activation pattern that can be matched with a familiar (previously encountered and stored) view.

With time, the initial activation in layers 1–3 is followed by inhibitory signals when the image is attended. The excitatory signals force features belonging to the same geon to continue to fire in synchrony with one another, whereas the inhibitory signals force the features of different geons to fire out of synchrony with one another (see Hummel, 2001; Hummel & Stankiewicz, 1996a). When an object's geons come to fire out of synchrony the resulting series of activation patterns in layer 5 form a structural description. This representation specifies the object's geons, their spatial relations, and the topological relations of their constituent surfaces. This structural description allows the model to recognise an object's identity even if it is shown in a novel view, or even if the object is a novel member of a familiar object category (i.e., a never seen before type of cup).

In layers 6 to 8 the representation generated by the activation patterns of the units in layer 5 are encoded into the long-term memory of JIM.3. Both the patterns of activation generated in the IGA as well as the HSM are propagated to corresponding collections of units in layer 6. In the set of units that receives input from the IGA, each unit learns to respond to the shape attributes of one geon (or collection of geons) and its relations to other geons in the object (layer 6i). A collection of geons will only be activated when multiple geons fire in synchrony with one another - i.e., when there is a binding error. However, when the model is working correctly, only one geon and relation vector will be passed into Layer 6i. Likewise, the collection of units receiving input from the HSM dedicates one unit each to the arrangement of surfaces in a geon, that is only one geon's surface features to Layer 6s at a time. That is, the units in Layer 6 respond to the "instantaneous" outputs of Layer 5.

In layer 7 the input from both sets of units (6i and 6s) are summed up over time such that they respond to multiple, mutually desynchronised geons. Thus, the resultant pattern of activation represents separate geons as a single object. A representation of an object is formed as a whole in terms of a structural description, that is, its constituent geons and relations. This activation pattern in layer 7 is fed to units in layer 8 that code object identity (and other semantic aspects). In test simulations (Hummel, 2001), the activation output in the object identity layer is the measure for successful object recognition. An important detail of the model is that there is a direct connection from layer 6s (representing learned holistic surface maps) to layer 8 (object identity) which allows the model to identify an object (in a familiar view) via a fast holistic route that by-passes any analytic representation.

Both the analytic and holistic components of the model support the recognition of attended images, but only the holistic representation is involved in the recognition of ignored images. The analytic representation is a structural description and therefore has orientation-independent properties as described by Hummel and Biederman (1992). The holistic component is orientation-sensitive because it represents the shape attributes of an object's surfaces in relation to a view-centred reference frame. The model predicts that attended images should prime themselves, translated, scaled and left-right reflected versions of themselves; ignored images should prime themselves, translated and scaled versions of themselves, but not their left-right reflections because the holistic representation is orientation-sensitive. This view-sensitivity is only short-lived, however, and object recognition will become more robust to orientation changes as long-term priming is only mediated by the structural description component. These predictions have been corroborated with test simulations on the model (Hummel, 2001).

There is empirical support for the predicted relationship between attention and patterns of visual priming across variations in viewpoint. Stankiewicz, Hummel, and Cooper (1998) observed that attended objects were visually primed in both the same view and in the left-right reflected view; ignored objects were primed only in the same view. The effects of attention (attended vs. ignored) and view (same vs. reflected) were strictly additive, that is the priming component from the analytic representation was independent of the priming component from the holistic component. Priming became view insensitive after several minutes. Further support for the model comes from Stankiewicz and Hummel (2002) who found that ignored images primed not only themselves, but also their scaled and translated versions in a short-term priming paradigm. Finally, the notion that recognition will be more view-dependent early in processing (due to a fast holistic route) rather than in later stages (after several hundred milliseconds in the analytic route) fits with earlier observations. Performance in picture matching tasks is usually view-sensitive at short interstimulus intervals (ISI), but becomes view-invariant after longer ISIs (Ellis & Allport, 1986).

In summary, JIM.3 is a model of the visual system in which two solutions to the binding problem for the representation of object shape are employed. These solutions have complementary properties. First, dynamic binding is used to generate analytic representations, which is a process that requires both visual attention and time to establish. The advantage of the resulting structural descriptions is that they specify object properties independently, which makes them highly flexible representations (e.g., that are robust to variations in viewpoint and the metric properties of an object's shape). Second, static binding does not require attention nor time to establish. However, static binding results in a representation that lacks the flexibility of a structural description, and is therefore susceptible to some (mirror-reflections, plane-rotations) but not all (translation, scale) changes.

JIM.3 is a promising model to investigate properties of object recognition. It is highly specified and allows relative clear predictions in particular about the effect of attention on object recognition and the form of representation (analytic vs. holistic) that results from attending or ignoring an image. In this thesis, only the part of the model will be considered that is concerned with the distinction between holistic vs. analytic processing, that is layers 5 and 6 in particular (as they consist of two qualitatively different components). These components of the model allow behavioural predictions concerning view and shape invariance in dependence of attention, which is the focus of this and earlier research (Stankiewicz et al. 1998; Stankiewicz & Hummel, 2002). The findings concerning the analytic and holistic representation obtained so far are summarised in Table 2. The next sections will give supplementary evidence for the hybrid representations and the role of attention in visual object recognition.

Analytic Representation	Holistic Representation
Requires Attention	Activated without attention
Invariant with left-right reflection (Biederman & Cooper, 1991a; Stankiewicz, et al., 1998)	Sensitive to left-right reflection (Stankiewicz, et al., 1998)
Long-lived priming (> ~5 minutes Stankiewicz et al. 1998)	Short-lived priming (> 3 seconds <~5 minutes; Stankiewicz et al. 1998)
Invariant with location (Biederman & Cooper, 1991a)	Invariant with translation (Stankiewicz & Hummel, 2002)
Invariant with scale (Biederman & Cooper, 1992)	Invariant with scale (Stankiewicz & Hummel, 2002)
Invariant with some rotations in depth (Biederman & Gerhardstein, 1993; but see Tarr, 1995)	Unknown
Part-based (Biederman & Cooper, 1991b)	Unknown

Table 2. Properties of the analytic and holistic representations of object shape (adapted from Stankiewicz & Hummel, 2002).

2.4.3 Evidence for Hybrid Representations in the Brain

As discussed in an earlier section (2.3), evidence from neuropsychological studies led to theories that proposed multiple representations of object shape (Humphreys & Riddoch, 1984; Warrington & James, 1988; Warrington & Taylor, 1978). Recent neuropsychological evidence (Davidoff & Warrington, 1999, 2001; Warrington & Davidoff, 2000) gives further support for the notion of a hybrid representation of shape in human object recognition. Davidoff and Warrington (1999, 2001) tested patients with severe difficulties in recognising objects. They were extremely impaired at recognising object parts or "exploded" objects (when parts were disconnected) but they could nevertheless name intact objects. Object recognition was also limited to familiar views. In terms of the hybrid model, their data could be interpreted to mean that the holistic route of these patient is intact, allowing them to recognise objects presented in intact familiar views, whereas their analytic route seemed to be impaired, preventing recognition of object parts or from unfamiliar views.

Other neuropsychological studies also indicate that object information may be stored in a view-dependent as well as a view-invariant manner. Turnbull (1997) described two patients that showed a clear double dissociation with respect to object recognition and object orientation performances. One patient with a visuo-spatial disorder was unable to identify the upright canonical orientation of objects shown as line drawings, although she could correctly name the objects. A second patient with visual object agnosia was unable to name many objects but could establish the upright orientation even for objects that he could not name. Together with other reports of orientation agnosia (Turnbull, Beschin, & Della Sala, 1997; Turnbull, Laws, & McCarthy, 1995) these case studies seem to support the claim for the independence of determining object orientation and object identity. Turnbull (1997) concludes that object information may be processed in two routes: A view-invariant route permits recognition but not necessarily determining the orientation of an object, and a view-dependent route that can by-pass the former description.

Neurophysiological evidence also suggests that two types of representation are involved in object recognition. Recently, more and more studies report evidence for both view-dependent and view-independent neurons. Booth and Rolls (Booth & Rolls, 1998) measured neuronal responses of IT cells in monkey brains to different views of familiar object. The majority of the visual neurons recorded were responsive to some views of some objects, as was observed in previous studies. A small subset of these neurons, however, were responsive to all views of particular objects. Booth and Rolls conclude that this finding provides evidence that these neurons were coding for objects rather than simply for individual views or visual features within the image.

A number of studies seem to indicate that anterior areas along the ventral pathway are less sensitive to image transformations such as translation in the visual field and scaling as compared to changes in orientation or configuration (Grill-Spector et al., 1999; Logothetis et al., 1995; Lueschow et al., 1994). This finding indirectly supports the hybrid model's prediction that scaling and translation are compensated by both the time-consuming analytic and the fast holistic component. More direct support for a neuronal basis of a view-dependent but size-independent representation comes from a recent fMRI study. Vuilleumier, Henson, Driver, and Dolan (Vuilleumier, Henson, Driver, & Dolan, 2002) used a repetition priming method presenting visual stimuli either with the same appearance or with changes in size, viewpoint or exemplar. Repetition of different exemplars with the same name affected only the left inferior frontal cortex whereas priming-induced decreases in activity of the right fusiform cortex depended on whether the objects were repeated with the same viewpoint, regardless of retinal image size. The decreases of activation found in the left fusiform were independent of both viewpoint and size. These findings strongly suggest that dissociable subsystems in ventral visual cortex maintain distinct view-dependent and view-invariant object representations.

Thus, in summary, although the hybrid model was driven by behavioural and computational motivations (Hummel, 1997) the neuropsychological and neurophysiological evidence seems to support the idea that representations mediating shape recognition in the brain rely on different processing systems.

2.4.4 Visual Attention

In this thesis, the formats used in object recognition and their dependence on attention will be studied following the hybrid model of shape recognition (Hummel, 2001; Hummel & Stankiewicz, 1996a). A brief review of the attention literature is therefore given here as far as it concerns the attentional paradigm and theoretical aspects relevant for the model.

Attention has become a major area of research in cognitive psychology. Associated with attention are mainly the ideas of selection and capacity limits (Pashler, 1995). The former refers to selecting a particular stimulus from a variety of other stimuli (as a familiar voice at a party) for further processing, since our sensory systems are always confronted with more stimuli than is consciously noticed. The latter refers to the constraints of human information processing systems (such as sensory or motor systems) that allow only a limited amount of information or tasks to be processed simultaneously. In the capacity approach (Kahneman, 1973), attention is described as a limited mental resource that is essential for information processing and can be allocated flexibly to various sources of information. As such, this approach was primarily addressed in divided attention studies. In contrast, selection has been discussed typically within a theoretical framework about the selective aspects of attention. A general agreement among researchers appears to be that irrelevant (unattended) information is processed differently from relevant information.

How are visual stimuli selected? According to location-based (sometimes termed space-based) accounts, objects are selected for processing by directing attention towards locations in the visual field (Cave & Bichot, 1999). In contrast, object-based accounts proposed that attention may select from mental representations of objects rather than from their location in the environment (Baylis & Driver, 1992; Driver & Baylis, 1989; Duncan, 1984; Lavie & Driver, 1996; Vecera et al., 2000). There is an ongoing debate about the role of space-based vs. object-based visual attention, which will not be repeated here (see Driver & Baylis, 1998). Important for the experimental paradigm reported in this thesis is the former group because of the use of a spatial cue for attentional selection.

It has been proposed that focused visual attention resembles a spotlight (Eriksen & Eriksen, 1974). This metaphor implies that attention may work like a spotlight of a torch in the dark across a region in the visual field. Attention can be moved within the visual field like a spotlight, and perceptual information is only processed if it falls within the "beam" of that spotlight. There have been variants of this idea, notably the zoom-lens model (Eriksen & St. James, 1986). This metaphor relates to the idea that the area (diameter) of focal attention can be increased or decreased dynamically.

The important role of location in selective visual processing has been demonstrated by numerous studies with a variety of experimental paradigms (Eriksen & Hoffman, 1974; Posner, Snyder, & Davidson, 1980). These studies have shown that advance knowledge of stimulus location facilitates stimulus processing even in the absence of nontarget items. Nissen (1985) has found that when participants were required to report the location and shape of items defined by colour, correct shape responses depended on correct localisations. Nissen concluded that the localisation of items is required for the correct integration of their individual features (e. g., colour and shape). The participants in the study by Tsal and Lavie

(1988) were presented with circular arrays of letters from which they had to report first a target specified by colour or by shape and then any other letters they could identify. The reported letters tended to be those adjacent to the previously seen target rather than those similar to the target in the cued property (e. g., other red letters). These and similar (Tsal & Lavie, 1993) findings suggest that the selective processing of targets is accomplished by attending to their locations.

In general, current thinking about attention in vision presumes that in attention tasks some information is excluded on the basis of location (Pashler, 1995). The locus of this filtering is probably "early", prior to object recognition, but probably after some initial feature analysis. Treisman and Gelade (1980) addressed this issue by using a visual search task in which observers had to look for a target among distracters, which varied in the number of feature dimensions (e.g., only colour or orientation). Also, the absolute number of distracters was varied. Treisman and Gelade reported that targets defined by a unique salient colour (e.g., red among green) or orientation (e.g., vertical among horizontal) were apparently processed in parallel, as the set size of distracters did not affect performance. By contrast, search for specific conjunctions of the same orientations and colours (e.g., red vertical among green vertical and red horizontal, where the target is unique only in its combination of these features) depended on the set size. Thus, processing of stimuli defined by multiple feature dimensions appeared serial. According to Treisman and Gelade (1980), individual features (colour, orientation) can be extracted 'preattentively' and in parallel, whereas feature integration requires observers to serially attend to the location of each item in turn. Treisman and colleagues (Treisman, 1986; Treisman & Schmidt, 1982) subsequently produced further evidence for the feature integration theory. According to this theory, features can only be integrated by use of focused visual attention mediated by the parietal lobes (Treisman, 1998).

There is a growing body of evidence in support of Treisman's theory that a primary function of visual attention is to permit the dynamic binding of independent object attributes (Logan, 1994; Luck & Beach, 1998; Luck & Vogel, 1997). In addition to behavioral evidence, there are supporting data from several studies concerned with the activation of parietal lobes which are known to play a role in attention (Ashbridge, Walsh, & Cowey, 1997; Corbetta, Shulman, Miezin, & Petersen, 1995; Driver & Mattingley, 1998). The Corbetta et al. (1995)

study showed distinctive activation of the parietal lobes during conjunction search in multiple element displays. A further striking demonstration of the role of attention for binding features used transcranial magnetic stimulation (Ashbridge et al., 1997). Transcranial magnetic stimulation (TMS) is a non-invasive temporary 'lesion' technique, which is associated with detrimentally affecting attention when applied over the parietal visual cortex. Participants were performing single feature ("popout") or conjunction visual search tasks. Although magnetic stimulation had no negative effect on the performance of pop-out search it did significantly increase reaction times during conjunction search. This finding indicates that attention is important for binding of visual attributes.

The studies on the role of attention for binding naturally concern the analytic representation of the hybrid model. If attention is necessary to bind features into objects, the question arises whether ignored objects can be recognised? A mechanism would be needed to group the elements of an image into objects that can be recognised without attention as proposed in the holistic route of the hybrid model. One possibility is that visual information is parsed according to the Gestalt principles of organization before attention is allocated within a scene. For example, Moore and Egeth (1997) found that Gestalt grouping does occur without attention but that these grouped patterns may not be encoded in memory without attention. However, there is a debate about the nature of the stimuli processed without attention and how much visual information can be processed at different stages.

There is little agreement on whether withholding attention results in the exclusion of irrelevant information from perception ("early selection" approach, see Treisman, 1969) or whether selective attention can affect only later processes, such as memory or responses ("late selection" view, e.g., Duncan, 1980; see Lavie & Tsal, 1994, for a review on the late and early selection debate). Concerning the latter, an active inhibition view of selective attention has been proposed (e.g., Tipper, 1985; for a review, see Fox, 1995). According to this approach, perceptual processing is not limited to attended information. Both relevant and irrelevant information are processed but attention can suppress or inhibit responses to ignored stimuli. Evidence for this view comes from studies showing negative priming, which is the slowing down of responses to overlapping line drawings of objects that were previously ignored (e.g., Tipper, 1985). Further evidence for processing of ignored images was demonstrated by Stankiewicz and colleagues (Stankiewicz & Hummel, 2002; Stankiewicz et al., 1998) who showed positive priming for ignored objects that were spatially separated from the attended item in the prime display.

The issue of the role of early vs. late selection will be revisited in the General Discussion. This review must be limited to those aspects of visual attention relevant for this investigation which are the formats of object recognition with and without attention, but not attentional mechanisms per se. In summary, important for the present study are the findings that location is a major determinant for attentional selection, that attention is crucial for binding features into objects, and that ignored objects can be processed to the level of identification in the visual system.

These considerations show that object recognition studies need to investigate what types of representations are involved in object recognition in dependence of visual attention. Stankiewicz and his colleagues have used an attentional priming paradigm to test object invariance for mirror-reflected, scaled and translated images. However, as can be seen from Table 2, the model's predictions concerning the holistic representations for plane-rotated and depth-rotated images have not been tested. Moreover, it is not known whether the holistic representation is affected by changes in the global structure rather than parts. These topics are the motivation and main focus of this thesis and will be investigated with a short-term priming paradigm. Naming responses to previously attended and ignored objects will indicate whether the predictions hold for changes in viewpoint and overall configuration.

3. Chapter 3: Experiments on Priming for Attended and Ignored Images

3.1 Rationale for the Use of the Priming Paradigm

Common objects are usually identified at their basic level (see section 1.3.2). In the context of experimental studies the term recognition may not always capture the basic level category. Observers sometimes have to make old-new recognition judgements, same-different or match-to-sample judgements. These explicit recognition tasks require the participant to refer back to specific, previously studied objects. In contrast, implicit recognition tasks such as naming and object-possibility judgements do not require any such reference to a previously studied object (at least not to a specific exemplar of that class). In general naming tasks are considered as an adequate measure for basic-level recognition (Biederman & Gerhardstein, 1993). To investigate the format of representations involved in basic-level object recognition Biederman and Gerhardstein (1993; 1995) used repetition priming for object naming. Repetition priming of recognition is the facilitation of recognising an item due to earlier exposure.

Repetition priming has two main advantages. First, repetition priming is one of a number of examples of "implicit" memory, which may be robust in circumstances (e.g., short presentation times) in which explicit memory is absent or poor. Second, priming may transfer between different item formats or domains and can therefore be utilised to investigate the perceptual processes of object, face, and word recognition. Inferences about the nature of the representations involved can be made on the assumption that a visual stimulus changes the state of the recognition system. Subsequent presentation of the same or similar item benefits recognition performance to the extent that the same representational units or pathways become activated. If priming is reduced by a change in item format it may indicate that the changed feature is a component of the representational description (e.g., see Biederman & Cooper, 1992). Naming has been favoured in priming tasks over other tasks such as matching by some researchers because the demands on naming seem additive with the effects of perceptual change (Biederman & Cooper, 1992). Recently, Bruce and her colleagues (Bruce, Carson, Burton, & Ellis, 2000) have found that repetition priming with naming tasks seems more sensitive to changes in perceptual properties of an item from study to test than other (semantic) tasks (such as deciding whether an object usually is found within of outside a house).

Researchers including Biederman (e.g., Biederman & Gerhardstein, 1993) and Stankiewicz et al. (1998) have used the repetition priming paradigm with a naming task to tap the representations that exclusively mediate object identification. All but one of the experiments described below will use the short-term priming paradigm of Stankiewicz and his colleagues.

3.2 Experiments 1-3: Priming of Intact and Split Objects

3.2.1 Introduction

In order to give a coherent account of analytic and holistic properties of object recognition, Hummel & Stankiewicz (1996a; Hummel, 2001) proposed a hybrid model of shape processing. The model generates both a holistic and an analytic representation of an attended image but only a holistic representation of an ignored image (see section 2.4.2). These two types of representation make independent contributions towards object recognition by providing alternative access to stored representations in long-term memory (LTM). To test their model, Stankiewicz et al. (1998) chose to investigate priming for mirror image reflections. The model predicts visual priming from the analytic representation, meaning that attended images will visually prime both themselves and their left-right reflections. The model also predicts a priming component from the holistic representation, which means that ignored images will prime themselves but not their reflections. The automatic holistic representation, acting independently of attention, allows the prediction of equal differences in priming across view changes for both attended and ignored conditions. Consequently, priming resulting from the analytic and holistic representations is additive. The predictions were tested and confirmed by Stankiewicz et al. (1998). However, these predictions are indirect in the sense that they are predictions about the effect of attention on priming across variations in viewpoint. Although view-based models of human object recognition may not have predicted the observed priming pattern, it is still possible that view-matching processes are involved. For example, the observed priming advantage for objects in the same view may be due to the fact that a different view has to be rotated to (or interpolated with) the previously encountered view, and that attention is simply more effective for this process. The hybrid model's prediction is that the visual system will generate a holistic representation of shape in response to an ignored image, and an analytic representation in response to an attended image. Thus, instead of indirect predictions involving view changes, a more suitable investigation would involve the manipulation of the holistic properties of an image.

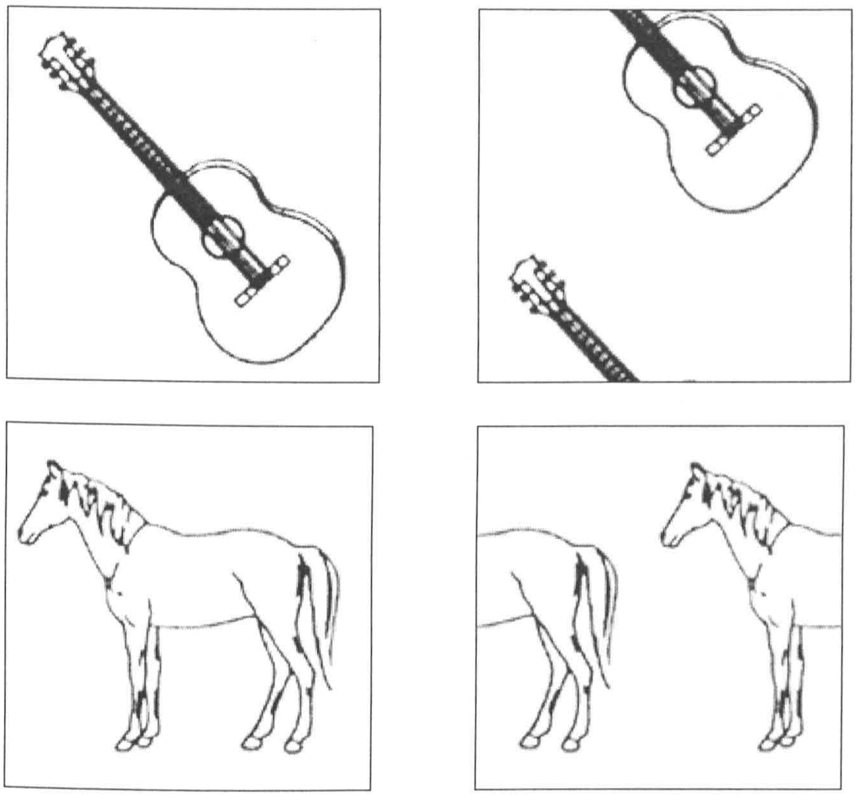


Figure 8: Examples of intact and split images used in Experiments 1 - 3 (images were shown without frames).

The first three experiments directly tested the model's predictions concerning analytic and holistic representations using the priming paradigm of Stankiewicz et al. (1998). The critical experimental manipulation was based on the logic that analytic representations should be more robust to configural distortions (as when parts are misplaced relative to one another) than are holistic representations. For example, the split and intact images of a guitar and a horse in Figure 8 depict all of the same parts and many of the same spatial relations. As a result, an analytic representation of the split image will specify many of the same parts and relations as an analytic representation of the intact image. By contrast, the locations of all the specific image features (e.g., lines, vertices, etc.) differ between the intact and split versions of the image. Consequently, a holistic representation of the split image is expected to have little or nothing in common with a holistic representation of its intact counterpart. These considerations suggest that a split image should visually prime its intact counterpart in an analytic (attended) representation of shape, but not in a holistic (ignored) representation.

Experiment 1 investigated the role of attention in priming for split and intact object images. Participants named objects in pairs of prime-probe trials. The first trial in a pair served as a prime, and presented two object images of which only one was spatially pre-cued. The probe object was always presented as an intact image, and an intact or split version served either as the attended (cued) prime image, as the ignored prime, or was preceded by an image of a completely different object which the subject had not previously seen in the experiment (which served as a baseline). The hybrid holistic/analytic model predicts that intact images will prime themselves whether they are attended or not although priming should be substantially greater for attended images than ignored images (Stankiewicz et al., 1998). Split images will prime their intact counterparts only when they are attended. Geon theory would predict equivalent priming for attended-split and for attended-intact images, but no priming for ignored images because binding processes necessary for structural descriptions are thought to depend on attention (Hummel & Biederman, 1992). View-based theories predict no or considerably reduced priming for attended-split images because the representational template has to be matched with a considerably changed image. View-based representations would also not necessarily predict priming in the ignored route (Olshausen, Anderson, & Van Essen, 1993), except with the assumption of a low-level matching process for identical images.

Experiment 2 served to estimate what fraction of the priming observed in Experiment 1 was specifically visual (as opposed to name or concept priming). Experiment 3 tested whether priming for ignored images in Experiment 1 could be attributed to low-level visual representations (e.g., of the local features in an object's image) rather than holistic memory representations. On the former account, ignored split images should prime themselves as much as ignored intact images primed themselves. On the latter (hybrid model) account, only ignored intact images should prime themselves.

3.2.2 Experiment 1: Priming for Split and Intact Images

3.2.2.1 Method

Participants

Forty-two native English speakers with normal or corrected-to-normal vision participated for credit in introductory psychology courses at the University of California, Los Angeles.

Materials

The experimental program was generated in E-Prime 1.0 (PSN). Black-and-white line drawings of 84 asymmetrical objects from the Snodgrass and Vanderwart (1980) set were displayed on a PC monitor. Response times were collected with a dynamic trigger microphone attached to an Interface Box. Participants sat approximately 90 cm from the display. The images were standardised in size to subtend 4.0° of visual angle. For each object a "split" version was created by using a 50% "offset" filter in Adobe Photoshop 5.5, resulting in images that appeared to be cut in two halves that were relocated to the opposite side of the original area (vertically or horizontally, depending on the main axis of the object). The manipulation did not alter the number or local configuration of any image features, except that some lines were necessarily broken at the location of the cut nor did it alter the total number of pixels in an image (see Figure 8).

Procedure

The experimental conditions in which objects appeared were counterbalanced across participants by placing each image into one of 14 clusters, each containing six images. Each cluster (and thus each image) was placed into one of seven conditions (attended-intact, attended-split, attended-not probed, ignored-intact, ignored-split, ignored-not probed, and unprimed) for any given participant, and all images appeared in all seven conditions equally often across participants. An image appeared in only one trial pair for any given participant. The ordering of the trials and the pairing of attended and ignored objects on prime trials were randomised for each participant. The participants read instructions, which they then paraphrased back to the experimenter. The experimental session began with 18 practice trials using a set of images different from the experimental set. After the practice trials, the participants were asked whether they had any questions.

The sequence of events in a trial is depicted in Figure 9. An unfilled circle remained in the centre of the screen until the participant pressed the space bar. The circle was then replaced with a fixation cross, which remained on the screen for 495 ms, followed by a blank white screen (for 30 ms). An attentional cueing square (4.5° of visual angle on a side) was then presented either to the left or right of the fixation cross, centred 4.0° from fixation. After 75 ms, two object images were displayed simultaneously for 120 ms, with the attended image inside the square, and the unattended image centred 4.0° from fixation on the other side of the screen. Each prime image was either intact or split. After the images disappeared, a blank screen was shown for 30 ms, followed by a random-line pattern mask that covered the entire screen (15.6° of visual angle) for 495 ms. The entire prime display lasted less than 200 ms; a duration that is too short to permit a saccade to the cueing square or either object. The participant's task was to say the name of the cued (attended) object as quickly and as accurately as possible. Response times were recorded by the computer through a voice key attached to a microphone on the table.

After the prime display, a blank screen was displayed for 1,995 ms, followed by a fixation cross (495 ms). Following a 30 ms blank screen, the probe image was displayed in the centre of the screen for 150 ms. The probe depicted either the attended object (attended conditions) or the ignored object (ignored conditions) or an object the participant had not seen previously in the experiment (unprimed baseline condition). The probe image was always the intact version. In total, 3,015 ms elapsed between the end of the prime display and the beginning of the probe display. The probe display was followed by a single pattern mask (4.6°) shown for 495 ms. The participant's task was to name the probe object as quickly and as accurately as possible. The computer then displayed the names of the attended prime and the probe object, as well as the probe response time. At the end of each trial-pair, the experimenter used the keyboard to record the participant's accuracy as well as voice key errors (i.e., when the voice key triggered erroneously). The participant could then initiate the next trial with a key press.

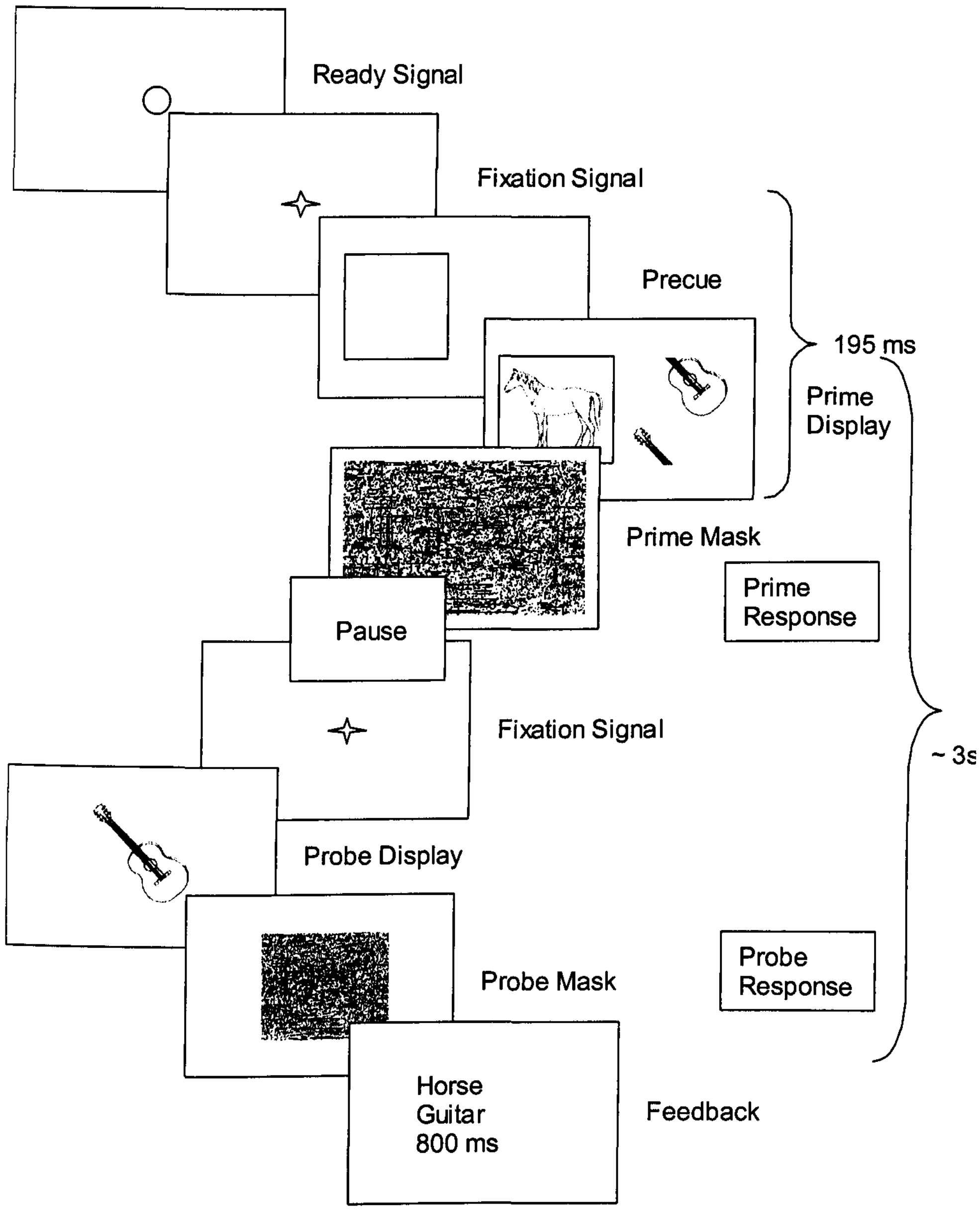


Figure 9: Sequence of displays in Experiment 1.

3.2.2.2 Results

In all conditions, priming was calculated as the participant's mean response time (RT) in the unprimed (baseline) condition minus their mean RT in the corresponding experimental condition. Trials on which either the prime or probe responses were incorrect (12.63 %) were excluded from the statistical analysis, as were voice key errors (4.03 %). Voice key errors were defined as recorded response times under 300 ms or recordings in which a noise different from the participants voice triggered the microphone. A 2 (Attention: attended vs. ignored) x 2 (Configuration: intact vs. split) within-subjects analysis of variance (ANOVA) revealed a reliable main effect of Attention, F (1, 41) = 167.35, $p < .001$, and Configuration, F (1, 41) = 18.61, $p < .001$, but the interaction between Attention and Configuration did not approach reliability, F (1, 41) < 1 (see Figure 10). No separate item analysis was necessary as objects were counterbalanced over subjects (for a discussion, see Raajimakers, Schrijnemakers, & Gremmen, 1999). Matched pairs t tests were conducted on each priming condition to determine which type of prime display caused savings in response time for the probe display (i.e., faster naming responses relative to unprimed probes). Priming was reliably greater than zero in the attended-intact, t (41) = 13.85, $p < .001$; attended-split, t (41) = 9.44, $p < .001$; and ignored-intact conditions, t (41) = 3.36, $p < .01$, but not in the ignored-split condition, t (41) < 1. A Friedman ANOVA over errors for probe trials showed no significant difference in the four priming conditions, Chi Sqr. (3) = 1.27, $p > .73$. There were no indications of a speed-accuracy trade-off in any condition.

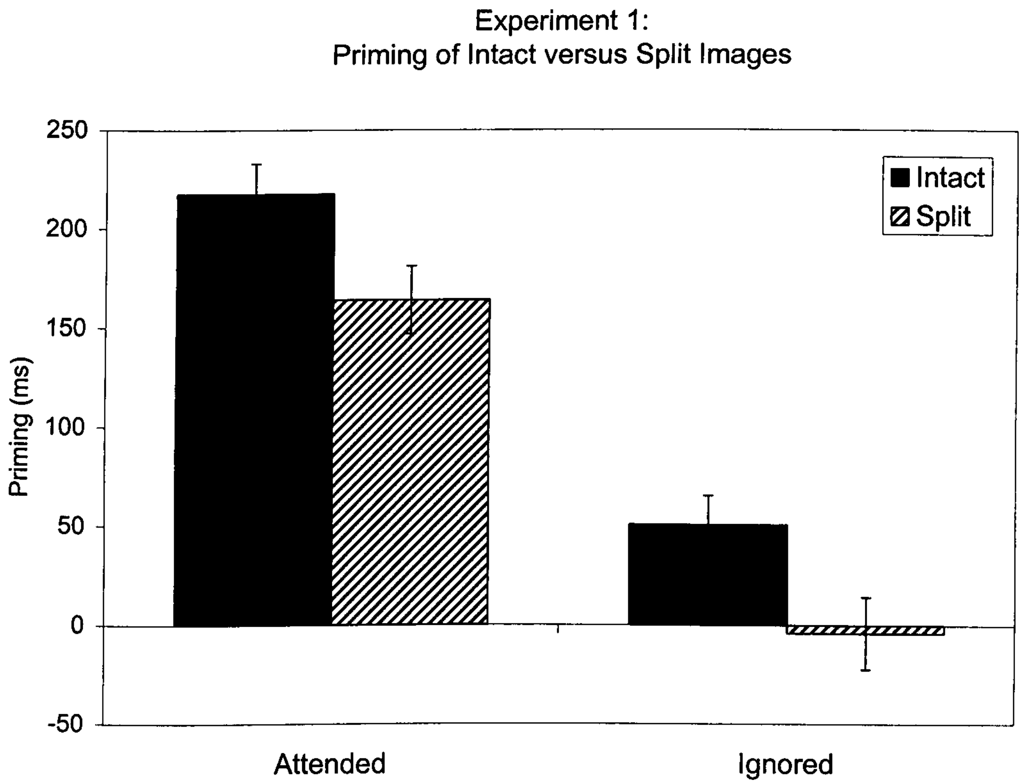


Figure 10: Priming (baseline RT minus RT in each experimental condition) means and standard errors for Experiment 1 as a function of whether the prime image was (a) attended or ignored and (b) intact or split (n = 42).

	Attended Intact	Attended Split	Ignored Intact	Ignored Split	Unprimed
% Errors (SE)	7.14 (1.80)	15.48 (2.69)	11.11 (2.10)	13.49 (2.62)	14.29 (1.97)

Table 3: Mean error rates for Experiment 1 (standard error in parantheses).

3.2.2.3 Discussion

The results of Experiment 1 are in line with previous studies that object recognition can be observed in the absence of attention (Tipper & Driver, 1988; Treisman & DeSchepper, 1996; Stankiewicz et al., 1998). In particular, as predicted by the hybrid model, both attended and ignored intact images primed their intact counterparts, but split images primed their intact counterparts only when they (the split images) were attended. Ignored objects were primed only in the identical view, suggesting that the representation mediating recognition without attention is sensitive to manipulations that affect the holistic shape of the object. There was a reliable priming advantage for intact primes over split primes which was almost identical in both attended and ignored conditions (~50 ms). Thus, the effects of attention (attended vs. ignored) and configuration (intact vs. split) were strictly additive. These results strongly support the hypothesis that two qualitatively different representations of shape mediate priming in object recognition: an analytic representation that is relatively robust to configural distortions but requires attention, and a holistic representation that is sensitive to configural distortion but does not require attention.

As Stankiewicz et al. (1998) pointed out, ignored intact images primed themselves although the prime and probe images appeared in different locations of the visual field (left or right of fixation in the prime display and at fixation in the probe displays). Even for ignored images, priming was obtained despite translation across the visual field. Stankiewicz et al. concluded that the intact-identical priming for ignored images does not simply reflect priming in an early representation of local (i.e., retinotopic) image features. That priming occurs despite location shifts from prime to probe display also suggests that the visual system may achieve invariance with translation even without attention. Stankiewicz and Hummel (2002) confirmed this invariance for ignored images with more systematic and larger translations across the visual field.

An important limitation of Experiment 1 is that attention was confounded with naming because participants named the cued images and did not name the uncued images. It is possible, therefore, that all the observed priming for split images in the attended condition was simply name or concept priming. Thus, it could be that visual representations that mediate object recognition are completely image-specific, and only the name and concept are invariant with configural changes. Experiment 2 was designed to test this possibility by investigating how much of the priming for attended split objects is name or concept priming, and how much is specifically visual priming.

3.2.3 Experiment 2: Visual Priming for Split Images

3.2.3.1 Introduction

Experiment 1 necessarily confounded attention with naming; if and only if the participants attended to an image did they name it. Therefore, it could be argued, that all the priming in the attended conditions derived from associative or name priming. Previous research (Stankiewicz et al., 1998) has shown that the priming from attended mirror images contained a large visual component but this is not necessarily due to the activation of an analytic representation. It is possible that the visual priming across view changes in their experiment was achieved by the mental rotation of a holistic representation (see Tarr & Pinker, 1990). However, the distorted configuration of split images should not provide sufficient visual similarity to promote priming from any holistic processing. Thus, demonstrating visual priming for split images would provide more direct evidence for a strictly analytic visual component in the hybrid model.

Experiment 2 was designed to estimate what fraction of the priming observed in Experiment 1 was due to visual as opposed to associative and/or name processing in the attended condition. To this end, the identical-image condition of Experiment 1 was replaced with objects having the same basic-level name as the corresponding probe object, but with a different shape (same-name-different-exemplar - SNDE). For example, if a jumbo jet (basic level name "aeroplane") served as a probe object, then Experiment 2 presented an intact image of a small private plane ("aeroplane") in the SNDE prime condition (instead of an identical intact jumbo jet), and a split version of the image of a jumbo jet in the split conditions. If any of the priming observed for attended objects in the split condition in Experiment 2 is specifically visual, then more priming with split image (jumbo jet) primes than with SNDE (private plane) primes is expected. By contrast, if all the priming observed with split images in Experiment 1 was simply name or concept priming, then the SNDE images should prime as much as (or more than) the split images (Biederman & Cooper, 1991a)

3.2.3.2 Method

Participants

Forty-two native English speakers with normal or corrected-to-normal vision participated for credit in introductory psychology courses at the University of California, Los Angeles, and at Goldsmiths' College University of London.

Materials

The Experiment used a set of 84 objects in 42 SNDE pairs. Half were taken from the Snodgrass and Vanderwart (1980) set used in Experiment 1, and the corresponding SNDE exemplars were line drawings of similar style.

Procedure

The conditions in Experiment 2 were identical to those of Experiment 1 except that the identical image condition was replaced by the SNDE condition. In the SNDE condition, each prime image was paired with an intact image of a different object with the same name (rather than being paired with itself, as in Experiment 1). Each SNDE object appeared equally often in each of the seven probe conditions.

3.2.3.3 Results

Trials in which either the prime or probe were named incorrectly were excluded (19.9 %), as were voice key errors (3.9 %). Figure 11 shows the priming results in each condition. A 2 (Attention: attended vs. ignored) x 2 (Prime Type: SNDE vs. split image) within-subjects analysis of variance (ANOVA) revealed a reliable main effect of Attention, $F(1, 41) = 13.63$, $p < .001$, but only a marginally significant effect of Prime Type, $F(1, 41) = 3.45$, $p = .071$. The interaction between Attention and Prime Type was reliable, $F(1, 41) = 4.73$, $p < .05$. Matched pairs t tests showed that the difference between the attended-SNDE and attended-split conditions was statistically reliable, $t(41) = 3.13$, $p < .01$, but the difference between the ignored-SNDE and ignored-split conditions was not, $t(41) < 1$. A Friedman ANOVA on errors for probe trials showed no significant difference in the four priming conditions, Chi Sqr. $(3) = 6.11$, $p > .10$.

Matched pairs t tests were conducted on each priming condition to determine which type of prime display caused savings in RT for the probe display. Priming was reliably greater than zero in the attended-SNDE, t (41) = 2.31, p < .05; attended-split, t (41) = 5.53, p < .01, but neither in the ignored-SNDE condition, t (41) < 1, nor in the ignored-split condition, t (41) < 1. Thus, SNDE and split images primed the corresponding probe image when attended, but not when ignored.

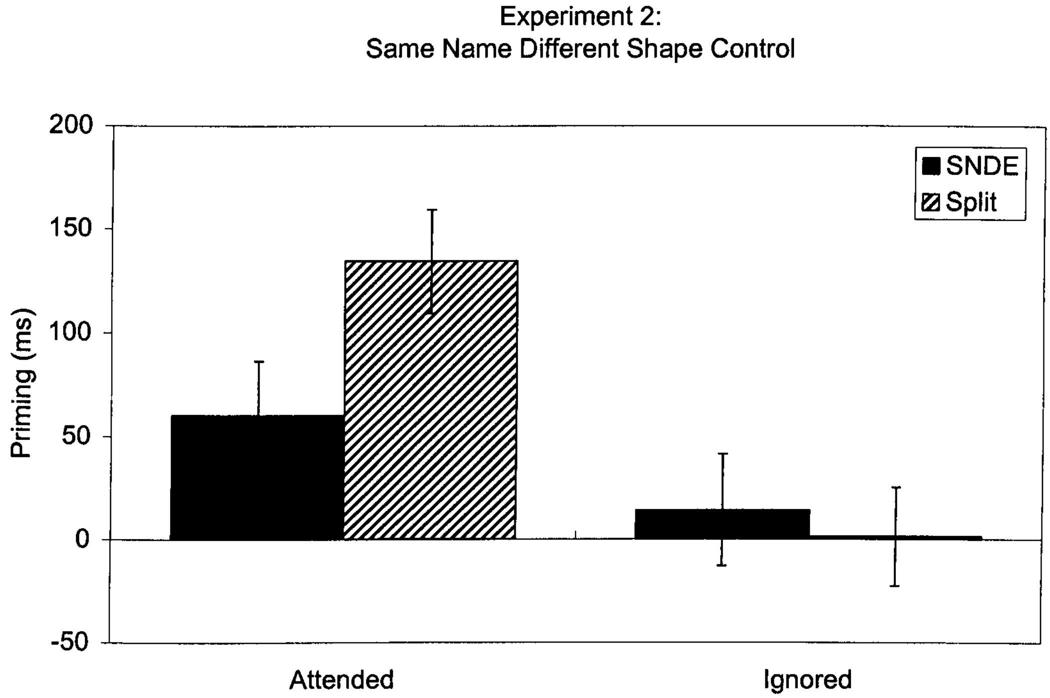


Figure 11: Priming means in ms and standard errors for Experiment 2 as a function of whether the prime object was attended or ignored and whether it was a same-name-different exemplar (SNDE) image or as a split image (n = 42).

	Attended SNDE	Attend. Split	Ignored SNDE	Ignored Split	Unprimed
% Errors (SE)	18.25 (3.45)	23.02 (3.85)	20.63 (3.93)	14.29 (3.62)	23.41 (2.96)

Table 4: Mean error rates for Experiment 2 (standard error in parantheses).

3.2.3.4 Discussion

An intact probe was primed more by an attended split image than by an attended intact different exemplar with the same name. Therefore, the results of Experiment 2 demonstrated that a substantial fraction of the priming for attended split images was specifically visual and not attributable solely to the same name or concept. Moreover, the visual priming obtained for split images constitutes substantial evidence that the recognition of attended images is at least partly mediated by an analytic representation because a strictly view based (holistic) representation should only be activated by an intact view (Hummel, 2000).

Experiment 2 found no priming in either of the ignored conditions. It therefore not only replicated the lack of priming for ignored split images in Experiment 1 but also produced no priming that was reliably greater than zero for SNDE prime images (replicating Stankiewicz et al., 1998; Experiment 2). These last results are in contrast to the findings of Tipper (1985) from a negative priming paradigm in that they provide no evidence for processing of unattended images to a semantic level of representation. This point will be discussed in a later section (4.4.1).

3.2.4 Experiment 3: Priming for Identical Split and Intact Images

3.2.4.1 Introduction

Experiments 1 and 2 together demonstrated visual priming for intact and split images in the attended conditions, but only intact images were primed in the ignored conditions. This pattern of priming is expected on the account that the visual system generates holistic representations of ignored images and analytic representations of attended images (Hummel, 2001; Hummel & Stankiewicz, 1996a). However, an alternative interpretation of the results of Experiments 1 and 2 is possible. It could simply be that non-identical images prime less than identical images and that ignored images prime less than attended images. Together, these effects could conspire to yield an absence of priming from ignored split images to their intact counterparts. Such an alternative interpretation would imply that the priming advantage for identical images compared to split images was due to simple low-level priming rather than the result of a shape representation matches (see Bartram, 1976; Ellis & Allport, 1986; Ellis et al., 1989). Although such a low-level priming account is challenged by the results of Stankiewicz and Hummel (2002), who showed that priming for ignored images is invariant with translation and scale (i.e., ignored images can prime non-identical versions of themselves), it cannot be ruled out based only on the results of Experiments 1 and 2.

Experiment 3 was designed to establish whether the results of Experiments 1 and 2 simply reflect a general decrease of priming for ignored non-identical images, or whether they reflect a reliance on holistic processing for ignored images as predicted by the hybrid model (Hummel, 2001). The logic of Experiment 3 is based on an assumption about the locus of the visual priming observed in these and other experiments. Priming is a form of learning, so one likely locus of visual priming is the point where visual representations of object shape are matched to representations stored in LTM (Biederman & Cooper, 1991a; Cooper, Biederman & Hummel, 1992; Hummel, 2001). In the hybrid model, priming is predicted between layers 5 and 6 (Hummel, 2001). Consequently, all priming, including that for an ignored object image, must reflect activation of pre-existing (stored) representations in LTM. If an image does not have a matching representation in LTM, then ignoring that image on one occasion should not even prime recognition of the very same image on a subsequent occasion. Consistent with this reasoning, Stankiewicz (1997) showed that ignoring an upside-down image on one trial does not prime recognition of the very same (i.e., still upside-down) image on the next trial.

Applied to the current paradigm, the logic is as follows: If the patterns of results of Experiments 1 and 2 for ignored images reflect the role of holistic processing, and if objects are encoded in LTM in an intact (rather than split) format, then ignoring a split image on one occasion should not prime even the very same image on a subsequent occasion. At the same time, as predicted by the hybrid model, attending to a split image on one trial should permit the encoding—and therefore priming—of that image. However, if the results of Experiments 1 and 2 are due to the alternative "identical primes more and attended primes more" account, then ignoring a split image trial **should** prime recognition of that image on a subsequent trial.

The design of Experiment 3 was identical to that of Experiment 1 with the exception that the probe image following an intact or split image was correspondingly either intact or split as well. Priming for both intact and split images was measured with regard to their respective baselines. The alternative hypothesis based on low-level visual similarity predicts visual priming across configurations; thus, ignored split images should prime themselves as much as ignored intact images prime themselves. By contrast, the hybrid model predicts that intact ignored images should prime themselves whereas ignored split images should not.

3.2.4.2 Method

Participants

Thirty-six English speakers with normal or corrected-to-normal vision participated for money or for credit in introductory psychology courses at Goldsmiths' College University of London.

Materials

The experiment used a set of 84 objects; 36 were used in prime-probe target pairs, the rest were fillers for attended and ignored primes. The items were taken from the Snodgrass and Vanderwart (1980) set and were similar to those used in Experiment 1.

Procedure

The procedure was equivalent to that of Experiment 1 except that in half the conditions the probe was a split image. The configuration (split vs. intact) of the probe image was always the same as the configuration of the prime image of interest (attended vs. ignored).

3.2.4.3 Results

Trials in which either the prime or probe were named incorrectly were excluded (15.7%), together with voice key errors (4.9 %). Priming was calculated in the same way as in Experiment 1 except that two different baseline conditions (unprimed-split and unprimed-intact) were used to calculate the corresponding priming for split and for intact conditions. The mean response time for the baseline intact configuration was 830 ms (SE 25) and 887 ms (SE 24) for the split configuration, a significant difference $t(35) = 2.28$, $p < .05$. The mean error rate for the unprimed intact images was 12.04 % (SE 2.36) and 19.91 % (SE 2.56) for the baseline split images, a significant difference, $t(35) = 2.22$, $p < .05$.

Figure 12 depicts the amount of priming observed in each condition. A 2 (Attention: attended vs. ignored) x 2 (Configuration: intact vs. split) within-subjects analysis of variance (ANOVA) revealed a reliable main effect of Attention, $F(1, 35) = 119.42$, $p < .001$, but not Configuration, $F(1, 35) < 1$. The interaction between Attention and Configuration was reliable, $F(1, 35) = 12.54$, $p < 0.01$. The difference between the ignored-intact and ignored-split conditions was statistically reliable, $t(35) = 2.10$, $p < .05$, but not the difference between attended-intact and attended-split and conditions, $t(35) < 1$.

Further matched pairs t tests were also conducted on each priming condition to determine which type of prime display caused savings in response time for the probe display (i.e., faster naming responses relative to unprimed probes). Priming was reliably greater than zero in the attended-intact, $t(35) = 7.67$, $p < .001$; attended-split, $t(35) = 11.93$, $p < .001$; and ignored-intact conditions, $t(35) = 2.81$, $p < .01$, but was not reliably greater than zero in the ignored-split condition, $t(35) < 1$.

A Friedman ANOVA for errors in probe trials was performed. There was a significant difference between the four priming conditions, Chi Sqr. $(3) = 11.41$, $p < .01$. A Wilcoxon Rank test revealed that the attended-intact conditions elicited significantly fewer errors than the attended split-condition, $z = 2.48$, $p < .05$, the ignored-intact condition, $z = 3.01$, $p < .01$, and the ignored-split condition, $z = 2.32$, $p < .05$.

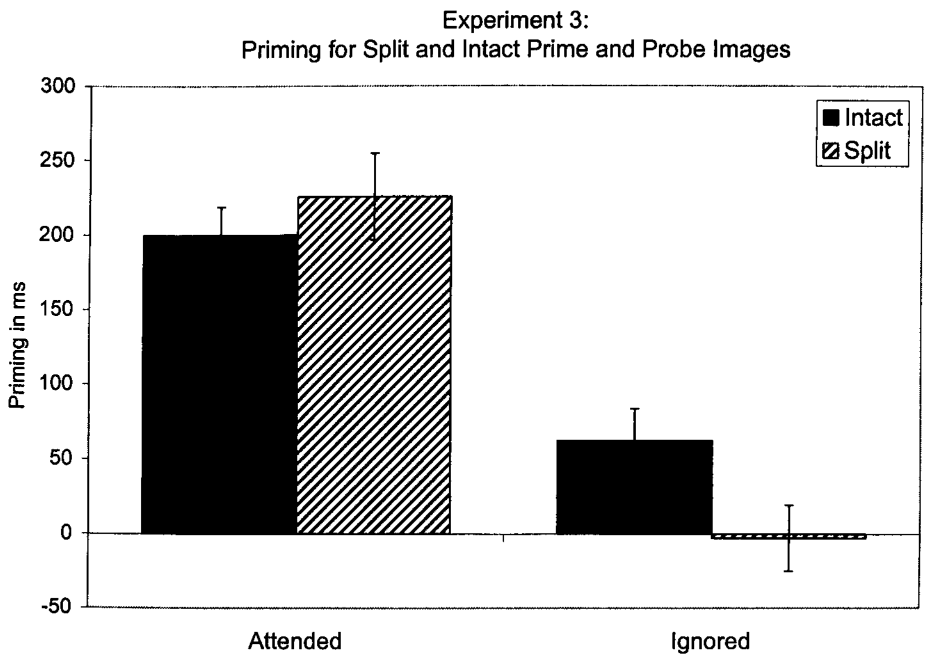


Figure 12: Priming means in ms and standard errors for Experiment 3 as a function of whether the prime object was attended or ignored and whether both prime and probe were split or intact (n = 36).

	Attended Intact	Attend. Split	Ignored Intact	Ignored Split	Unprimed Intact	Unprimed Split
% Errors (SE)	7.87 (1.81)	19.90 (3.17)	17.13 (2.70)	17.13 (3.01)	12.04 (2.36)	19.91 (2.56)

Table 5: Mean error rates for Experiment 3 (standard error in parantheses).

Additional ANOVAs on errors and latencies including unprimed conditions were performed to check for any speed accuracy trade-off. A 3 (Primetype: attended, ignored, unprimed) by 2 (Configuration: intact vs. split) ANOVA on error rates (excluding voice key errors) revealed only a reliable main effect of Configuration, $F(1, 35) = 12.87$, $p < .01$ with error rates for split images higher than for intact images. There was a trend towards an interaction between Attention and Configuration with a greater difference in error rates for the attended conditions $F(2, 70) = 3.09$, $p = .052$. A 3 (Primetype: attended, ignored, unprimed) by 2 (Configuration: intact vs. split) ANOVA on probe latencies revealed a main effect of Primetype, $F(2, 70) = 94.71$, $p < .001$, a main effect of Configuration, $F(1, 35) = 29.83$, $p < .001$, and a significant interaction, $F(2, 70) = 5.54$, $p < .01$. Latencies for split images compared to intact images were longer for both attended, $t(35) = 2.33$, $p < .05$, and unprimed conditions, $t(35) = 2.28$, $p < .05$, and somewhat longer still in the ignored conditions, $t(35) = 5.63$, $p < .001$. There was no evidence for a speed accuracy trade-off in the latency data indicating that observers responded to split images faster while making more errors. In general, higher error rates were accompanied by higher response times.

3.2.4.4 Discussion

Experiment 3 showed that a split image primed itself when attended but not when ignored, whereas an intact image primed itself in both conditions. Critically in the ignored conditions priming was found only for intact images but not for split images. Experiment 3 thus demonstrated that the lack of priming for ignored split images in Experiment 1 cannot be attributed to a general decrease of priming in response to split images. The priming pattern is predictable from the hybrid model and in contrast to the alternative hypothesis that would have predicted equal priming in both ignored conditions.

The pattern of priming for attended conditions differs somewhat from that of Experiment 1. When attended, a split image primed itself just as much as an intact image primed itself. This result could seem surprising, especially given the failure of ignored split images to prime themselves at all. A possible interpretation could have been that the equivalent priming for attended split and intact conditions is due to a speed-accuracy trade-off because participants risk more errors while responding as fast to split images as to intact images. However, the analysis of error rates and latencies suggests that this is not the case. Alternatively, it is also possible that because split images are harder to recognise, they profit more from priming than do intact images. This interpretation is supported by the analyses of error rates and probe latencies; overall, there were both more errors and longer latencies for split images. Moreover, Srinivas (1993) found similar effects with more priming for unusual (foreshortened) than usual (canonical) views.

Whatever the reason for equivalent priming in the attended conditions, the main result of this experiment comes from the difference in priming for the ignored conditions. Ignored split images do not prime themselves while ignored intact images do prime their identical counterpart. Thus, it is not simply the case that identical ignored images necessarily enjoy greater priming than non-identical ones, and attended images enjoy more priming than ignored ones. Rather, the pattern of priming characterising ignored images differs qualitatively from that characterising attended images. Namely, attended images prime themselves, translated, scaled, reflected and split versions of themselves, whereas ignored images prime themselves, translated and scaled versions of themselves, but not reflections or split versions of themselves. Experiment 3 clearly suggests that the priming pattern observed in Experiment 1 cannot be explained by a simple "more priming for identical images" (low-level processing) account.

3.2.5 General Discussion of Experiments 1-3

Experiment 1 showed that attending to either an intact or split image primed its intact version. However, whereas ignoring a split image did not prime its intact version, an ignored intact image did prime itself. Experiment 2 demonstrated that a substantial fraction (at least 80 ms) of the observed priming for attended objects was specifically visual (rather than simply name or concept priming). Experiment 3 showed that ignored split images did not prime themselves while ignored intact images did. These findings strongly suggest that, as predicted by the hybrid analytic/holistic model of Hummel and Stankiewicz (1996a;

Hummel, 2001), the visual system represents attended images both analytically and holistically, and represents ignored images only holistically.

The interpretation of Experiments 1 to 3 rests on the assumption that splitting an image will seriously disrupt holistic processing, but not substantially affect the structural description representation. Cave and Kosslyn (1993) tested observers on their performance with intact versus non-intact images from the Snodgrass and Vanderwart set. Non-intact images were created by either removing "natural" parts as assessed prior to the studies (e.g., the handle of a desktop phone, or the arm of a pair of glasses), by disconnecting the image along natural part boundaries, or disconnecting the image "unnaturally" (i.e. not parsing at part boundaries). Furthermore, both fragment types of disconnected images could be scrambled randomly on the page so that their spatial relations were perturbed relative to the original image and their merely disconnected versions. Cave and Kosslyn found that non-intact images were generally named more slowly than intact images. The performance for non-intact images did not depend on how the images were split when participants had 1s to recognise an object. However, with presentation times of only 200 ms, images with scrambled relations of parts were harder to identify than images that were merely disconnected or those that lacked a certain part. Also, unnaturally parsed images with scrambled relations were harder to identify than naturally parsed and scrambled images in the 200 ms condition. Thus, the pattern of these results confirms the assumption that it does not matter so much how the holistic properties of an image are eliminated (part-removal or disconnection, and presumably, splitting) as long as the spatial relations between the parts are still intact. Furthermore, spatial relations among parts do matter when processing capacity is limited (shorter presentation time), as does the nature of parsing (natural vs. unnatural). Cave and Kosslyn interpreted their results as evidence that the participants only processed parts and spatial relations when processing resources are limited. Because performance was equally bad for all non-intact images in the 1s condition, the authors conjectured that their participants used parts-based processing as a "fallback strategy". In terms of the hybrid model, the conclusion would be that parts and relations are always processed but that the disruption of parts-based processing is only evident when capacity is limited (i.e. only 200 ms for recognition) and attentional binding fails. When observers had

1 s to identify an object, there was enough time to establish temporal binding even among parts of disconnected and scrambled images. The holistic representation, however, was always disrupted when parts of an object were disconnected, which caused longer recognition times. In fact, Cave and Kosslyn's data seem to fit perfectly well with the predictions of the hybrid model.

The experiments presented here have important implications for object recognition theories in general and for the hybrid model in particular. The findings strongly suggest that both analytic and holistic representations work in parallel rather than in a serial manner. It is not simply a matter of "early" priming for ignored images and "early and late" priming for attended images. Such a serial model can in principle account for the short-term priming effects observed by Stankiewicz et. al. (1998), where the finding of less priming after a change in viewpoint (left-right reflection) was associated with presenting non-identical rather than identical images in prime and probe trials. However, such a serial account can not explain the priming effects observed in Experiment 3. Identical images do not necessarily prime themselves when ignored; they need to be holistic representations. At the same time, if they are attended, split images prime themselves and their intact versions. Thus, both holistic and analytic representations can make contact with object memory independently of each other.

Split images allowed a more direct test of whether processing involves analytic representations because in contrast to left-right reflections they do not permit the holistic processing of an image. Nevertheless, one of the most striking aspects of the findings presented here is that the results of Experiment 1 (Figure 10) are nearly an exact numerical replication of the findings of Stankiewicz et al. (1998, Experiment 1), who used left-right reflections rather than split images (see Figure 13). In both experiments, identical images enjoyed a priming advantage of approximately 50 ms over non-identical (i.e., split or reflected) images in both the attended and ignored conditions. The fact that configural distortions (splitting an image) and left-right reflection have such similar effects on attended and unattended images suggests that similar mechanisms are at work in both cases. Namely, attention permits the visual system to generate an analytic representation robust to both configural distortions and left-right reflection, whereas in the absence of attention, object recognition must rely on holistic representations that are robust to neither of these manipulations.

The use of split images had a further advantage in allowing us to establish the locus at which priming of ignored images manifests itself. A split image, when it appears on a prime trial, necessarily depicts all the same features as on the probe trial (albeit at different locations in the visual field), and represents the same object shape. But in spite of this equivalence, split images did not prime themselves when they were ignored in Experiment 3. The data presented here suggest that the locus for the observed priming of ignored intact images does not reside in the representation of image features, or even in the representation of object shape. Split images do not have pre-existing holistic representations in LTM - a reasonable assumption given that participants have probably not spent time viewing split images of objects. Therefore, when a split image is ignored, the resulting holistic representation of its shape simply has nothing to prime in LTM (see also Stankiewicz, 1997). By contrast, an intact image presumably does have a pre-existing holistic representation in LTM provided it depicts the object in a familiar view (Hummel & Stankiewicz, 1996a). As a result, the holistic representation generated by an ignored intact image did have something to prime, hence the approximately 50 ms of priming for intact ignored images. In addition, the results of Experiment 2, which showed that an ignored image did not prime a different member of the same basic-level category, suggest that priming for ignored images does not reside at the level of concepts or object names. Together, these results suggest that the priming resides at the interface between shape perception and object memory as proposed by Cooper, Biederman and Hummel (1992).

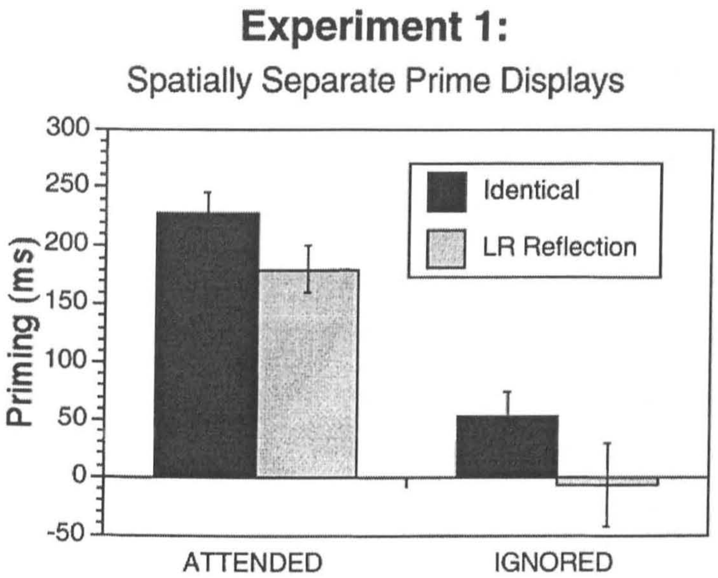


Figure 13: Priming means in ms and standard errors for Stankiewicz et al.'s (1998) Experiment 1 as a function of whether the object was attended or ignored and whether the object was identical to the prime image or a left—right (LR) reflection of the prime (adapted from Stankiewicz et al. 1998).

Experiments 1 to 3 showed that images with a configural distortion still yield priming effects when attended, but not when ignored, indicating that two qualitatively different representations are involved in object recognition. However, much of the debate in recent years has centred around the question of how the visual system achieves object constancy across changes in viewpoint (Biederman & Gerhardstein, 1993; Tarr & Bulthoff, 1995). As outlined in the introduction, competing object recognition theories proposed different mechanisms to achieve object constancy across view changes. Therefore, the following experiments will test whether the hybrid model can account for the properties of human recognition concerning orientation changes. Experiment 4 to 6 will be concerned with plane rotations, and Experiments 7 to 9 will investigate priming after depth-rotations.

3.3 Experiments 4-6: Priming of Upright and Rotated Objects

3.3.1 Introduction

Experiments 1 to 3 tested the predictions of the hybrid model with intact and split images to distinguish between holistic versus analytic processing. The results showed that images with a configural distortion still visually primed an intact image when attended, but not when ignored, indicating that two qualitatively different representations are involved in object recognition. The question now arises whether the effects of priming predicted by the hybrid model can account for view changes in the picture plane. Viewpoint-dependent theories of object recognition hold that representations of familiar objects are stored in memory in a single canonical view (Palmer, Rosch, & Chase, 1981; Ullman, 1989) or in multiple views (Bulthoff & Edelman, 1992; Tarr, 1995; Tarr & Pinker, 1989). Objects in unfamiliar or novel views are recognised by normalising them to an upright or familiar view (Tarr & Pinker, 1989) or by using mathematical interpolation to determine whether a given image is a "legal" projection of a familiar shape (e.g., Poggio & Edelman, 1990).

Both view-based and structural description accounts of object recognition predict considerable performance costs for recognition of plane-rotated images (Hummel & Biederman, 1992). For example, Tarr and Pinker (1989) assume an analogue mental rotation process to normalise a misoriented object to its upright position. The time to normalise an object is assumed to be a function of the degree of its misorientation in the picture plane which explains the linear increases in response times. Hummel and Biederman (1992), however, claim that this pattern of view-dependent performance could be a result of a mismatch between perceived and stored part structure. For example, when a coffee mug is rotated 90° in the picture plane, the handle appears below-of instead of "side-attached" to the body. The visual system still recognises the object because the units coding the critical parts of the mug description are activated, but recognition is more time consuming because of the compensation necessary for the non-matching of units coding the relative spatial arrangements. Thus, the mismatch between perceived and stored spatial relations affects identification because the visual system needs time to correct this structural perturbation. According to Hummel and Biederman (1992), the greater the angular disorientation in the picture plane the more part relations are disturbed; this causes the linear increases in response performance. We will return to this point in a later section.

In their critical study with primed mirror images, Stankiewicz et al. (1998) showed viewpoint constancy in the attended component versus viewpoint sensitivity in the ignored component. However, the particular viewpoint manipulation of left-right reflection is special (Murray, 1997; Rollenhagen & Olson, 2000; Warrington & Davidoff, 2000) and maybe interpreted as 2D (flip) or 3D (rotation) manipulations of its original. Mirror-image equivalence may not be the same as viewpoint invariance because for example, neurons responding best to views that are mirror images of objects may respond differently to other views (Logothetis et al., 1995). The principal aim of the Experiments 4 to 6 is to generalise the model's predictions to objects that are rotated in the picture plane. An immediate concern is that, unlike for mirror images, rotated versions of the objects normally employed in object recognition suffer substantial recognition costs when rotated. Differences in baseline responding could promote difficulties in assessing priming (see Experiment 3). Fortunately, recognition costs may not be associated with all types of objects (Vannucci & Viggiano, 2000).

When observers rated the goodness of views for common objects, one cluster – termed "poly-oriented" objects – did not yield a preferred "upright" orientation (Verfaillie & Boutsen, 1995). The consequent distinction between base (objects with a preferred upright) and no-base objects has been found to have importance for both behavioural (Vanucci & Viggiano, 2000) and neuropsychological (Davidoff & Warrington, 1999) investigations of object orientation. In particular, no-base objects can be completely insensitive to orientation and no baseline naming differences would be predicted. Experiment 4 is therefore conducted to test whether images of no-base objects elicit equivalent naming latencies in different orientations in the picture plane, whereas objects with a definite base should incur recognition costs after these manipulations. In Experiments 5 and 6, the predictions of the hybrid model for plane-rotated objects are tested by using the short-term prime probe paradigm of Stankiewicz et al. (1998).

3.3.2 Experiment 4: Naming of Objects Rotated in the Picture Plane

3.3.2.1 Introduction

Experiment 4 aims to extend the findings of Vanucci and Viggiano (2000) that object decision depended on the degree of plane rotation but only for animals and inanimate objects with a definite base (e.g., a house). Objects without a definite base (e.g., hammer) were decided upon equally well at all orientations. The aim of Experiment 4 is to verify that the same pattern of results will hold for the speeded naming tasks to be used in the next studies. Moreover, in Vannucci and Viggiano (2000), insufficient account was given to the matching of base and no-base object sets on familiarity and visual complexity; both might be argued to have affected the pattern of performance. Nevertheless, even after more comprehensive matching, it is predicted that recognition latencies for animals and objects with a definite base will increase with rotation away from the canonical (upright) view and no-base objects will be identified equally well from all orientations. This pattern is predicted by most theories of object recognition that rely on stored views but also by structural description theories (see Hummel & Biederman, 1992). Given this prediction turns out to be correct, the distinction of no-base and base objects will be employed in the short-term priming paradigm.

3.3.2.2 Method

Participants

Twenty-nine native English speakers with normal or corrected-to-normal vision participated for credit in introductory psychology courses at Goldsmiths' College University of London.

Materials

Three subsets of 24 objects each (animals, base objects, and no-base objects) were taken from the Snodgrass and Vanderwart (1980) set. The base and no-base object subsets were matched for familiarity and visual complexity (means: 3.75 and 2.64 vs. 3.69 and 2.77). The means for familiarity and visual complexity of the animal subset were 2.69 and 3.70, highly significantly different from the former two subsets ($p < .001$). For each object, the standard view (as obtained from the original set) was assigned as the 0° view. Counter-clockwise rotations in the picture plane resulted in 60° and 120° orientations for each object (see Figure 14).

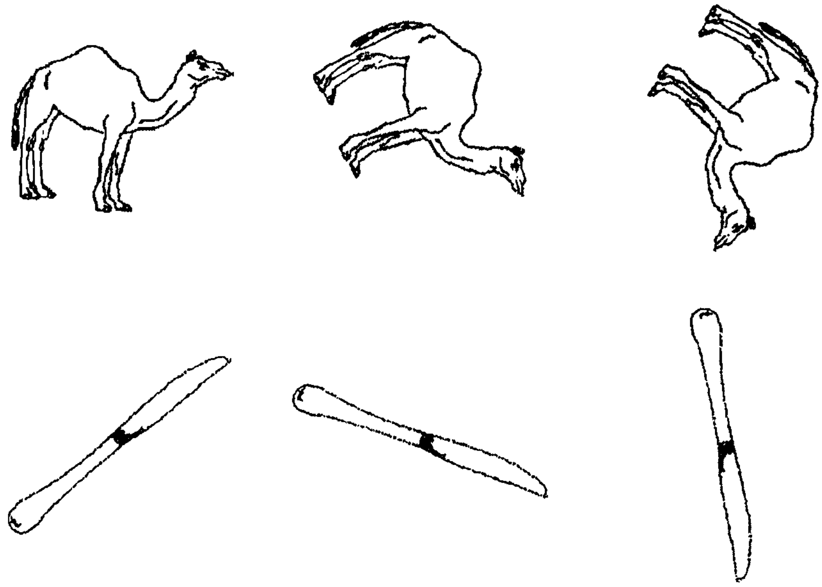


Figure 14: Examples of object images used in Experiment 4.

Procedure

The allocation of objects to the experimental conditions was randomised for each participant. Thus, there were 8 different objects in each of the three object sets for each of the three orientation conditions (0°, 60°, and 120°) constituting a total of 72 trials per participant.

The participants read instructions after which they paraphrased them back to the experimenter. After four practice trials with objects different from the experimental set participants were asked if they had any questions. Each subsequent test trial was initiated by the participant. A trial began with an unfilled circle (subtending 0.032° of visual angle) in the centre of the screen that was replaced by the participant's key-press with a fixation cross for 495 ms. An object (subtending 4.57° of visual angle) was then shown in the centre of the screen for 195 ms followed by a single pattern mask for 495 ms. The participant's task was to name the probe as quickly and as accurately as possible. After the response, a feedback display with the name of the object and the response time was shown. At the end of each trial, the experimenter used the keyboard to record the participant's accuracy and any voice key errors.

3.3.2.3 Results

The overall error rate was 6.32 % (including voice key errors). Response times for correct trials and error rates were submitted to a 3 (Object type: animals vs. base vs. no-base) by 3 (Rotation: 0° vs. 60° vs. 120°) ANOVA. For latencies, there were significant effects of Object type, $F(2, 56) = 33.41$, $p < 0.001$, Rotation, $F(2, 56) = 21.84$, $p < 0.001$, as well as for the interaction, $F(4, 112) = 4.15$, $p < .01$ (see Figure 15). A similar pattern was found for an ANOVA over errors. Significant effects were again found for object type, $F(2, 56) = 15.75$, $p < 0.001$; rotation, $F(2, 56) = 12.06$, $p < 0.001$, and the interaction, $F(4, 112) = 9.07$, $p < .001$.

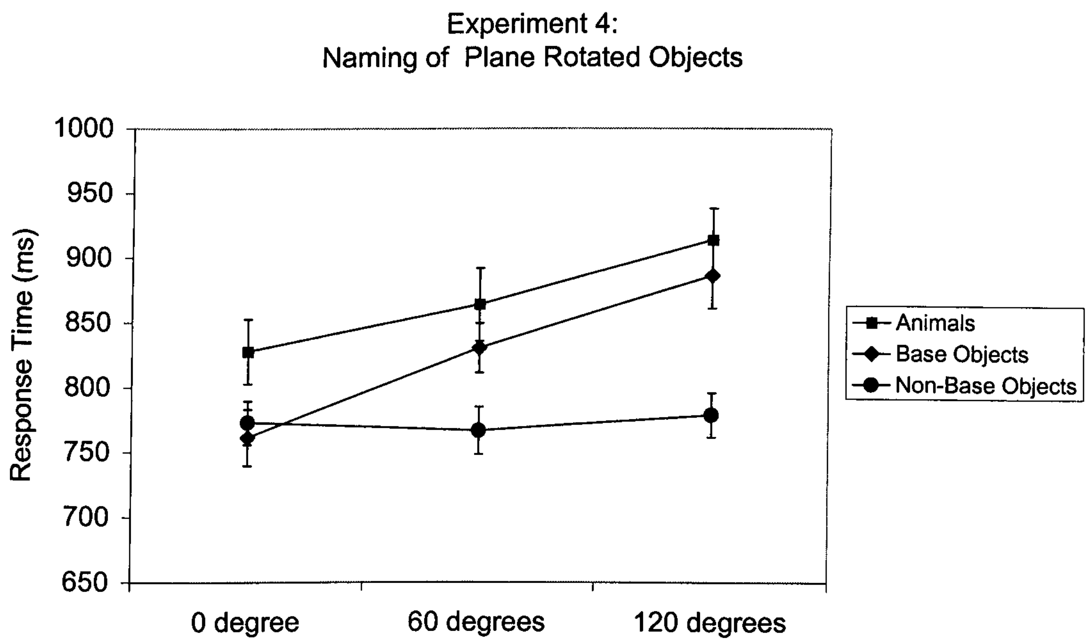


Figure 15: Response time means and standard errors for Experiment 4 as a function of the degree of rotation in the picture plane and the type of objects (n=29).

% Errors (SE)	Animals	Base Objects	No Base Objects
0°	3.88 (1.26)	2.59 (1.14)	2.59 (1.14)
60°	6.03 (1.71)	4.74 (1.44)	2.59 (0.96)
120°	17.24 (2.51)	8.19 (1.99)	0.43 (0.04)

Table 6: Mean error rates for Experiment 4 (standard error in parantheses).

Concerning latencies, post-hoc comparisons using Tukey's HSD test revealed that, for animals, the increase in response times from 0° compared to 120° rotation was significant ($p < .01$). Similar analyses revealed that, for base objects, the differences between 0° and 60°

($p <. 05$) and between 0° and 120° ($p < .001$) were significant. For no-base objects, however, there were no conditions in which the differences even approached significance ($p > .05$). Error rates were small and post-hoc comparisons revealed significant effects only for animals. There were significant differences between 60° and 120° ($p < .001$) and between 0° and 120° conditions ($p < .001$). For the other two object types, no significant differences were found for any other comparison ($p > .05$).

3.3.2.4 Discussion

Experiment 4 clearly demonstrated that orientation change in the picture plane had differential effects on base and no-base objects. It was only objects with a definite base that incurred increasing recognition performance costs when rotated. Even though animals could not be matched for familiarity and visual complexity and were, in general, harder to identify, they showed the same pattern of increasing response times and error rates as the other base objects. However, no-base objects were equally recognisable in all picture plane orientations. Thus, the present results extend the findings of Vannucci and Viggiano (2000) with object decision to the current object naming task. The results for base objects are also very similar to those for naming artefacts and natural objects with a distinctive top and bottom in Experiment 1 of Dickerson and Humphreys (1999).

The results of Experiment 4 do not distinguish between structural description and multiple views accounts of object recognition. They would be predicted from the multiple views account because objects that are seen often from different viewpoints would be stored in a like manner (Tarr & Pinker, 1989). Hence, recognition latencies would be equivalent over the range of stored views for no-base objects, because they are arguably encountered in a variety of orientations. In contrast, structural description accounts would explain the results in terms of different spatial relations used for base and no-base objects (Hummel & Biederman, 1992). In contrast to base objects or animals the parts of objects without a definite base are not coded in terms of spatial relations such as "on-top-of" or "below-of". Rather, parts of objects like a hammer or scissors may be described as being "side-attached", which is less disturbed when an object is rotated in the picture plane. Both multiple views and structural description accounts predict response time costs for rotated base objects and animals.

The goal of Experiment 4 was not to distinguish between view-based accounts and structural description theories for object constancy over plane rotations. Rather, the aim was to verify that no-base objects would provide images to test the predictions from the hybrid theory of Stankiewicz et al. (1998). These priming studies require measurements against a baseline. No-base objects will provide a means of accurately matching baseline latencies for different rotations of an object because it has now been shown that these are identical for this class of object.

3.3.3 Experiment 5: Effects of Viewpoint and Attention on Priming

3.3.3.1 Introduction

The aim of Experiment 5 is to extend the priming results of Stankiewicz et al. (1998) to the more common picture plane rotations compared to the relatively special case of mirror image reflections. The hybrid model would be more compelling if it were shown to have properties that would predict the effects of more common rotational transformations. While picture plane rotations are common experimental manipulations, unfortunately for priming studies, such transformations can produce large variations in naming latencies (Lawson, 1999). The data from Experiment 4 allow us to use no-base objects for which naming latencies do not vary with rotation. As in Stankiewicz et al. (1998), we expect visual priming for attended objects in both the same and a rotated view but priming for ignored objects should only be observed for objects shown in the same view. There should be no priming from ignored objects in a different view. In addition, in the attended condition, the prediction would be that same views should prime more than rotated views. On the Stankiewicz et al. model, the reason for the additional priming is that same views would activate both the analytic and holistic components whereas a rotated view only activates the analytic component.

Concerning attended images, the structural description account of Biederman (1987) would predict equal priming for identical and rotated views of no-base objects because access would be to the same geon structure. In contrast, the multiple views account would favour the same predictions as the Stankiewicz et al. (1998) account that priming should decrease for non-identical (rotated) views but for a different reason. On the multiple views account, more priming would occur if the identical view-based representation is activated by prime and probe displays. In the ignored condition, structural description accounts would predict no priming for ignored objects whatever their orientation because attentional processes are necessary to code parts and their relations into structural descriptions (Hummel, 2001; Hummel & Biederman, 1992). Multiple views accounts do not allow strong predictions for the ignored conditions as they do not explicitly incorporate attention. However, multiple view accounts are unlikely to predict access to stored object representations for ignored stimuli. For example, Olshausen and his colleagues proposed a model of object recognition that predicts that attention is necessary to map retinotopically organised visual information from V1 to a scale-and translation-invariant representation in IT (Olshausen et al., 1993).

3.3.3.2 Method

Participants

Thirty native English speakers with normal or corrected-to-normal vision participated for credit in introductory psychology courses at Goldsmiths' College University of London.

Materials

A set of 56 no-base objects similar to the ones in Experiment 1 was used in Experiment 2. Of these, 24 were target items and 32 filler items that were never used as probes.

Procedure

The basic procedure was as in Stankiewicz et al. (1998). The target items were counterbalanced across participants by placing each object in one of six clusters, with each cluster containing four objects. The clusters (and thus, each object) were placed into one of six conditions (attended-same, attended-rotated, ignored-same, ignored-rotated, unprimed-same-view and unprimed-rotated-view). Thus, an object appeared in only one trial during a session. The target objects appeared in all six conditions equally often across participants. The filler objects appeared in the unprobed conditions. Prime and probe objects were shown in the standard view (as in the original Snodgrass and Vanderwart set) or rotated 90° in the picture plane. The two views appeared in all conditions equally often.

The procedure was similar to Experiment 1 except that participants read the names of the objects before starting the practice trials, and the number of practice trials was reduced to 10. Another change in design relative to Experiment 1 was that each view of the no-base objects (standard and rotated) was used in the standard condition half the time and in the rotated condition in the other half of the trials to counterbalance possible idiosyncratic effects.

3.3.3.3 Results

Trials on which either the prime or probe responses were incorrect were excluded from the analysis (13.1 %) as were voice key errors (5.0 %). The mean response time for the (unprimed) unrotated condition was 814 ms (SE 36.7) and 824 ms (SE 34.2) for the (unprimed) rotated condition, a non-significant difference, $t(29) < 1$. The mean error rates for the unrotated condition was 15.0 % (SE 3.5) and 14.2 % (SE 3.1) for the rotated condition.

For all conditions, priming was calculated as the difference between each participant's mean latency in the unprimed (baseline) condition and the participant's mean latency in each of the other probe conditions. A 2 (Attention: attended vs. ignored) × 2 (Rotation: same view vs. rotated view) within-subjects ANOVA was performed on priming latencies. The analysis revealed a reliable main effect of attention, $F(1, 29) = 10.87$, $p < .01$ and a main effect of rotation, $F(1, 29) = 14.79$, $p < .001$. The interaction between attention and rotation was not reliable, $F(1, 29) < 1$ (see Figure 16). A Friedman ANOVA on probe errors revealed no significant differences in the four conditions, Chi Sqr. $(3) = .29$, $p > .96$.

Matched pairs t tests revealed priming reliably greater than zero in the attended-same, $t(29) = 4.27$, $p < .001$; attended-rotated condition, $t(29) = 2.85$, $p < .01$; and ignored-same conditions, $t(29) = 2.12$, $p < .05$, but not in the ignored-rotated condition, $t(29) < 1$, $p > .05$. Thus, attended images in the prime display primed the probe image in both the same and the rotated view but ignored images primed the probe object only when it was presented in the same view.

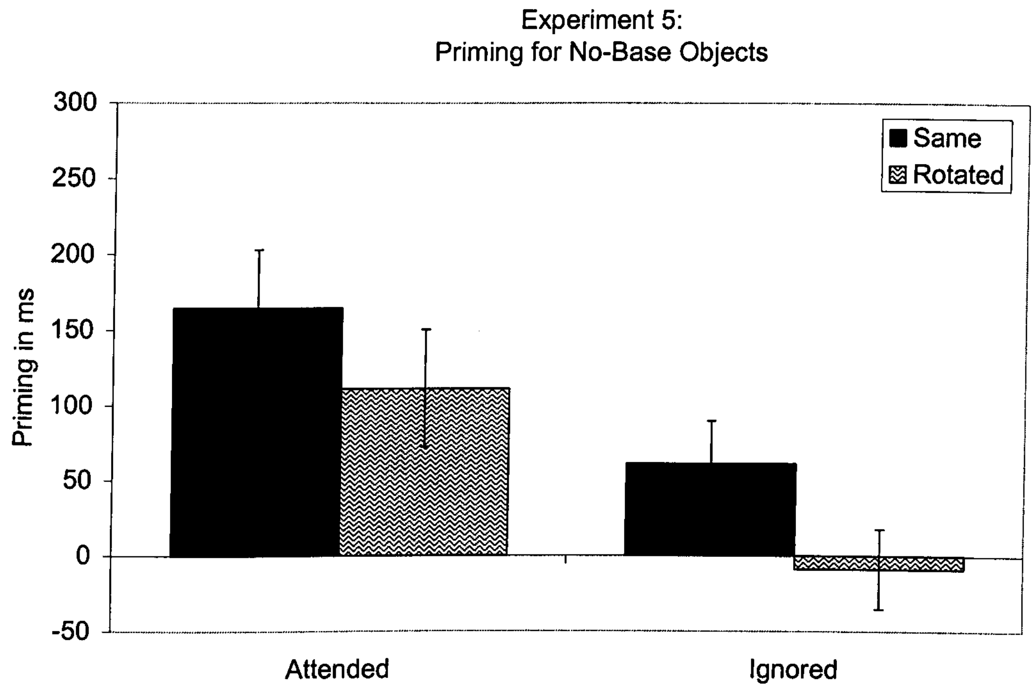


Figure 16: Priming means in ms and standard errors for no-base objects in Experiment 5 as a function of whether the object was attended or ignored in the prime display prior to the probe and whether the probe objects were presented in the same orientation as the prime image or rotated in the picture plane (n = 30).

	Attended Same	Attend. Rotated	Ignored Same	Ignored Rotated	Unprimed Same	Unprimed Rotated
% Errors (SE)	11.67 (2.87)	8.33 (2.77)	15.00 (4.42)	14.17 (3.32)	15.00 (3.52)	14.17 (3.10)

Table 7: Mean error rates for Experiment 5 (standard error in parantheses).

3.3.3.4 Discussion

The pattern of priming effects observed in Experiment 5 clearly replicated the findings of Stankiewicz et al. (1998). On the additive account of Stankiewicz et al. (1998), the pattern of performance in the same attended condition is due to the combined effects of analytical and holistic processing. Attended images primed both themselves and their plane-rotated versions whereas ignored images only primed themselves and are not primed by a plane-rotated view. Importantly for the hybrid model, there was an equivalent priming advantage for same views relative to rotated views in both attended and ignored conditions. The results are not due to recognition differences between unrotated and rotated views because Experiment 4 showed that no-base objects did not show an effect of viewpoint, confirming that they are equally identifiable across plane-rotations.

The priming effects obtained in Experiment 5 do not fit with aspects of geon theory (Biederman, 1987). On that account, no-base objects should be primed equally by themselves and by their plane-rotated versions. The visible part-structure is the same in all orientations and their structural description would not contain explicit "above" or "below" relations. In addition, geon theory would not predict view-dependent priming in the ignored conditions, as binding of parts should require attention (Hummel, 2001).

View-based models would have no problem in accounting for the data from the attended conditions. Objects may be stored in a variety of orientations in which they are encountered, and therefore there are no recognition costs for rotation in the baseline conditions. The priming advantage for identical relative to rotated conditions can be attributed to higher activation of the previously encountered view. The probe image would be matched more efficiently with the same view-specific representation than with any other stored view of that image. However, many multiple views accounts (e.g., Tarr & Pinker, 1989) do not explicitly mention the role of attention and would therefore not necessarily predict priming in the ignored conditions. One model related to the view-based approach that might explain the effects in the ignored conditions is the model of Olshausen, Anderson and Van Essen (1993) in which attention serves a gating function in early visual processing.

According to Olshausen et al. (1993), the outputs of retinotopic visual neurons (as found in V1 and V2) are mapped to neurons whose receptive fields are invariant with translation and scale (and possibly other variations in viewpoint) in higher visual areas such as infero-temporal cortex (IT). Ignored information is either not mapped from V1 to IT or, if it is mapped, then it is sensitive to metric variations such as translation, scaling and rotation (see Olshausen et. al, 1993). In the current paradigm the position of the prime (left or right of fixation) and probe (at fixation) image changes between prime and probe trial. The Olshausen et al. model would therefore not account for the obtained priming of ignored identical images described here because translation invariance did not depend on attention. Nevertheless, one might still argue that the hybrid account is not needed to explain the priming of ignored identical images. An identical image could prime itself when ignored not because it accesses a stored object shape description (presumably in IT) but simply because the same low-level features (edges and contours) are activated when responding to a previously seen image (see Experiment 3, section 3.2.4). According to this account, every image should prime its identical version from prime to probe trial, which could explain the priming in the ignored conditions in Experiment 5. In contrast, the hybrid model predicts that ignored images access stored representations. Experiment 6 seeks to rule out such low-level activation as the cause of view-dependent priming in the ignored conditions by testing whether images in unfamiliar (rotated) views prime themselves when ignored.

3.3.4 Experiment 6: Priming for Upright or Rotated Images of Base Objects

3.3.4.1 Introduction

The goal of this Experiment was to establish whether the pattern of priming observed in the identical ignored condition of Experiment 5 was due to activation of view-dependent object representations or to simple facilitation based on identical low-level features from study to test. In Experiment 5, either explanation could hold because rotated images were primed by a different image. However, on the low-level matching account, it should follow that priming would be solely dependent on view. A different prediction would follow from the hybrid model of object recognition. On that model, it is only identical familiar views of objects that cause priming in the ignored condition. Priming in the ignored conditions occurs from the activation of holistic representations but these exist only for familiar views. Therefore, identical unfamiliar views would not show priming from ignored images.

Experiment 6 considers only priming from objects with a definite base because they have only one familiar view - the upright orientation; no-base objects are equally well recognised after plane rotation (see Experiment 4). In the priming conditions of Experiment 6, both the prime and the probe objects are either shown in their identical familiar (upright) view or both are shown in their identical unfamiliar (90° plane-rotated view, as in Experiment 5). The particular interest is in the ignored trials. If ignored images make contact with stored representations in object memory, then objects that have a definite base and are seen almost exclusively in an upright position (e.g., a house or a car) should exhibit no priming when both the prime and the probe object are shown in an identical rotated (unfamiliar) view. At the same time, they should exhibit priming when the prime and probe objects are presented in the identical upright (familiar) view. If, however, the priming observed for ignored objects is due to simple low-level priming, then even unfamiliar (rotated) views of base objects should prime themselves. Thus, the critical difference of Experiment 6 compared to the previous Experiment 5 is that for attended and ignored conditions, both the prime and the probe appear always upright or both appear in a rotated view (similar to Experiment 3 in which both the prime and the probe appeared in an intact configuration or they both appeared in their split configuration).

3.3.4.2 Method

Participants

Thirty native English speakers with normal or corrected-to-normal vision participated for credits in introductory psychology courses at Goldsmiths' College University of London.

Materials

A set of 84 objects was taken from the Snodgrass and Vanderwart (1980) set. There were 36 target objects with a definite base and 48 filler objects that were never used as probes.

Procedure

The basic procedure was similar to that of Experiment 5 with the following two differences. First, only base objects were used as targets; second, prime and probe images were always shown in the same orientation in a single trial – either rotated (90°) or upright.

3.3.4.3 Results

Trials on which either the prime or probe responses were incorrect (8.7 %) were excluded from the analysis of latencies, as were voice key errors (3.3 %). The mean response time for the baseline standard view was 787.3 ms (SE 31.2) and 805.6 ms (SE 19.8) for the rotated view, a non-significant difference $t(29) < 1$. However, the mean error rate for the upright view was 4.4 % (SE 1.5) and 13.8 % (SE 2.5) for the rotated view, a significant difference, $t(29) = 3.32$, $p < .01$.

The analysis of the data was similar to Experiment 5, with the exception that different baselines (upright vs. rotated) were used to calculate priming for the corresponding conditions. A 2 (Attention: attended vs. ignored) × 2 (View: upright view vs. rotated view) within-subjects analysis of variance (ANOVA) for priming revealed a reliable main effect of Attention, $F(1, 29) = 123.40$, $p < .001$, but no reliable main effect of View, $F(1, 29) = 1.70$, p > .05. The interaction between Attention and View was reliable, F (1, 29) = 7.64, p < 0.01 (see Figure 17). The difference between the ignored-upright and ignored-rotated conditions was statistically reliable, t (29) = 2.27, p < .05, but not the difference between attended-upright and attended-rotated and conditions, t (29) < 1. A Friedman ANOVA on probe errors revealed no significant differences between the four conditions, Chi Sqr.(3) = 1.81, p > .61.

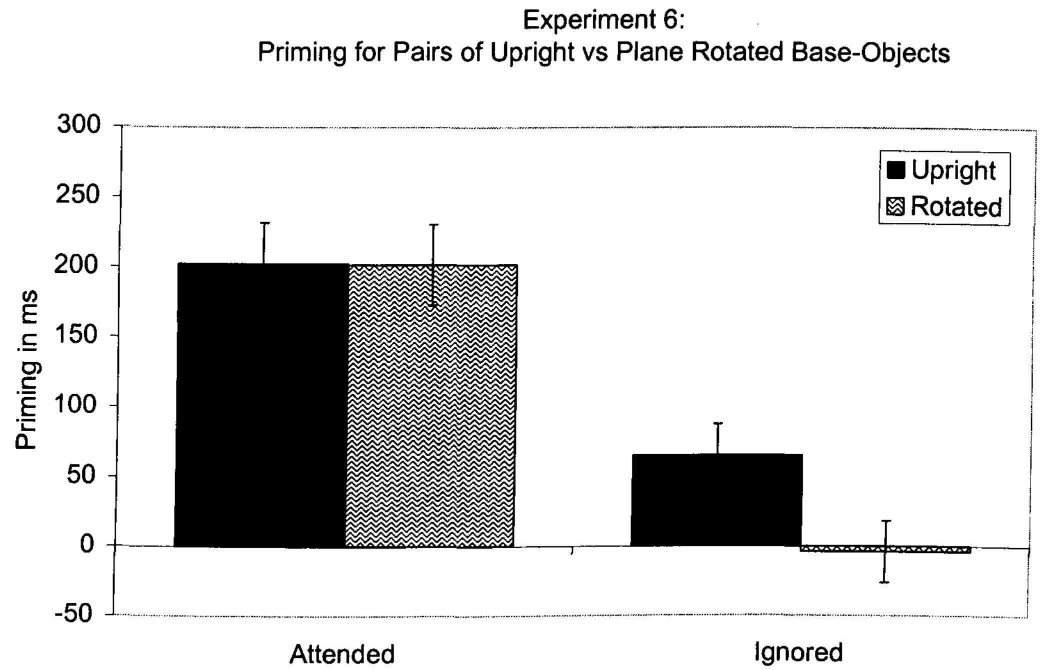


Figure 17: Priming means in ms and standard errors for base objects in Experiment 6 as a function of whether the object was attended or ignored in the prime display prior to the probe and whether the prime and the probe objects were presented in the upright orientation or rotated in the picture plane (n = 30).

	Attended Same	Attend. Rotated	Ignored Same	Ignored Rotated	Unprimed Same	Unprimed Rotated
% Errors (SE)	5.56 (1.66)	17.22 (2.93)	6.11 (2.33)	5.00 (1.41)	4.44 (1.58)	13.89 (2.53)

Table 8: Mean error rates for Experiment 6 (standard error in parantheses).

Matched pairs t tests were conducted on each priming condition to determine savings in response time when naming the probe object (i.e., faster naming responses relative to unprimed probes). Priming was reliably greater than zero in the attended-upright, t (29) = 6.92, p < .001; attended-rotated, t (29) = 9.04, p < .001; and ignored-upright conditions, t (29) = 2.26, p < .05, but not in the ignored-rotated condition, t (29) < 1. Thus, attended images in the prime display primed themselves in both the upright and the rotated view but the ignored images primed themselves only when presented as an upright view.

A deviation from the analysis in Experiment 5 was that different baselines were required (unprimed-upright vs. unprimed-rotated) to calculate priming because better performance was expected for upright images than for rotated objects. In the previous experiment, only a single baseline was required because of equal performance across views of no-base objects. In order to compare the priming pattern with Experiment 5 an additional ANOVA was run over priming data obtained by pooling baselines (i.e. priming was established from the mean of unprimed-upright and unprimed-rotated condition). There were reliable main effects of Attention, $F(1, 29) = 123.39$, $p < .001$, and View, $F(1, 29) = 28.45$, $p < .001$, and the interaction between Attention and View was reliable, $F(1, 29) = 7.64$, $p < 0.01$. Thus, the interaction effect found in this Experiment was not due to the way the baseline was established. However, pooling the baseline to calculate priming accentuated absolute priming differences between upright and rotated conditions, hence the now significant main effect of rotation. Nevertheless, the basic priming pattern remained the same for all conditions.

3.3.4.4 Discussion

The critical result of Experiment 6 is the replication of the pattern of performance found with ignored images in Experiment 5. Once more, we find a significant amount of priming in one condition and no priming in the other. The lack of priming in Experiment 5 was for ignored rotated images that were familiar. Importantly, the lack of priming here in Experiment 6 was for ignored *identical* views that were unfamiliar. Thus, the priming results in the ignored conditions of Experiment 5 cannot be attributed to simple low-level priming of images resulting from changes in early visual stimulation and cannot be trivially attributed to the amount of featural overlap between prime and target views. The priming pattern found here (and in Experiment 3) is perhaps the most direct evidence that images of ignored objects achieve priming from access to stored familiar views. Very similar results were obtained by Stankiewicz (1997). He showed in one experiment that attended upside-down images (i.e. flipped over the horizontal axis) primed themselves as much as upright images primed themselves, whereas ignored images primed themselves only in the upright condition. Unattended upside-down images of base objects did not prime themselves.

In general, the data of Experiment 6 are similar to the pattern of priming effects in Stankiewicz et al. (1998) for same view versus left-right reflection. However, they reported an additive effect of attention and viewpoint in the attended condition. Here, in Experiment 6, there was equivalent priming (no additivity) for attended objects i.e., there was a larger amount of priming for attended rotated (unfamiliar) objects than in Experiment 5. There are at least two alternative explanations for these data. First, it could simply be that a larger absolute value for the baseline would act to increase the magnitude of priming. The parts of the rotated prime would activate the same analytic components of the object representation as an upright base object. If the rotated images start out further from ceiling, this could allow extra analytic priming. Indeed, the analyses of latencies and errors show main effects of orientation for base objects (see Appendix), confirming the findings of Experiment 4. Second, an alternative explanation for the equivalent priming in the attended conditions could be that by attending to and recognising the rotated image, a holistic representation is thereby encoded. Hence, on subsequent presentation of the target, both upright and rotated images benefit equally from analytic and holistic representations. This interpretation fits with experimental data reported elsewhere showing that previously attended rotated objects are normalised more quickly during subsequent presentation, whereas formerly ignored objects do not show such an advantage of prior exposure (Murray, 1995c). It also fits with earlier studies that report reduced effects of orientation on performance after repeated presentation of stimuli (Jolicoeur, 1985; Jolicoeur & Milliken, 1989; Jolicoeur, Snow, & Murray, 1987). The present data cannot differentiate between these alternative explanations.

Whatever the best explanation of the equal priming in the attended conditions, it in no way detracts from the findings for the ignored conditions. Even if priming for rotated images had an advantage over upright images because of baseline differences we would expect such an advantage for ignored-rotated conditions. In the ignored conditions, we found no priming for unfamiliar views of objects. Any effects of increased baseline for rotated over upright images ought to increase the amount of priming, giving the ignored-rotated images an even greater chance of priming. This was not the case here. Ignored images prime themselves only when they can be matched with a familiar view in memory. This finding is indirectly supported by a recent event-related fMRI study. Henson, Shallice and Dolan (2000) found that repetition priming resulted in different patterns of attenuation in the right fusiform area depending on whether the repeated stimuli were familiar or unfamiliar. This effect persisted over multiple repetitions. Henson et al. concluded that priming-related responses depend on the presence or absence of pre-existing stimulus representations.

3.3.5 General Discussion of Experiments 4-6

The data from Experiments 5 and 6 support the role of attention in object constancy and are consistent with the hybrid model of Hummel and colleagues (Hummel, 2001; Stankiewicz et al., 1998). Objects primed themselves and their rotated versions when attended but they primed themselves only in the identical orientation when ignored. Critically, in Experiment 6, ignored objects in unfamiliar views did not prime themselves. This indicates that the priming observed for ignored images must be due to access to a stored (holistic) representation. Priming advantages for conditions of Experiment 5 appeared additive and thus lend support for analytic part-based representations.

One possible limitation of Experiments 5 and 6 is that recognition and naming were confounded in the "attended" conditions. The priming observed is therefore likely to contain a semantic or name prime component, as well as a component for visual priming. Priming measured with naming tasks has reliably been shown to be sensitive to image changes from study to test (Bruce et al., 2000). Furthermore, as described earlier, Stankiewicz et al. (1998; Experiment 2) estimated visual priming by substituting the image in the identical conditions with a different object that had the same basic-level name (e.g., a grand piano instead of an upright piano). These "same-name-different-exemplars" produced no priming in the ignored condition, and significantly less priming than reflected images in the attended conditions. Subtraction produced a conservative estimate of about 80 ms for purely visual priming that is not due to the holistic component. This is almost exactly the difference between attended-intact and attended-SNDE images obtained in Experiment 2 of the present thesis. There was significantly more facilitation for the split version of the prime image than for an intact SNDE version indicating that there was a large visual priming component for the former. Taken together, the evidence suggests that the priming found in the attended conditions contained a significant and large visual component.

Theoretical accounts that propose only a single format of representations for objects – either view-dependent or view-independent – are currently not powerful enough to explain these findings. Structural description accounts (Hummel & Biederman, 1992) would not predict priming (let alone view-dependent priming) in the ignored route, because attention is needed to bind parts and their relations together to form a representation that can be matched with a stored object model. Multiple views accounts would not have predicted the combination of priming patterns found in Experiments 5 and 6, as priming seems view-dependent not because of normalisation or low-level processes but because of a holistic representation that makes contact with object memory independently of attention. View-based theories that propose only a single representation of object shape would have to assume that priming is simply greater for identical images than for non-identical images, and greater for attended images than ignored images. Such a theory would predict uniform effects of attention on patterns of view-invariance in visual priming. For example, in the model of Olshausen et al. (1993) attention is used to map retinotopically organised visual information from V1 to a scale-and translation-invariant representation in IT. However, Experiment 6 (equivalent to Experiment 3) showed that identical images prime themselves only in familiar views when ignored. Thus, the present data indicate that attention is not always needed to map information from V1 to a shape representation in IT. These results together with the findings of Stankiewicz and Hummel (2002) are inconsistent with the hypothesis that the function of visual attention is to generalise over variations in the size or location of an object's image.

If non-identical images do not always incur reduction in priming relative to identical images, then the results of Experiment 5 are not easily explained by multiple views accounts. Recall that there were equivalent reductions in priming for both attended and unattended no-base images when they were rotated from prime to probe trial. Experiments 4 and 5 (equivalent performance across views in baseline conditions) have established that mental rotation or any other transformation account is not necessary to recognise these objects from rotated orientations. Therefore, multiple views theories have to make additional assumptions to account for the lack of priming in the ignored rotated conditons in Experiment 5.

The experiments so far tested the hybrid model by manipulating the holistic properties between the prime and probe image. In the next set of experiments depth-rotation is used to manipulate the analytic properties between prime and probe image.

3.4 Experiments 7-9: Priming for Depth-Rotated Objects

3.4.1 Introduction

The experiments so far have shown that two different representations may work in parallel in object recognition - one that is analytical (e.g., a structural description) and view-invariant, and one that is holistic (e.g., a view template) and highly sensitive to changes in viewpoint such as mirror-reflection (Stankiewicz et al., 1998), and plane-rotation (Experiments 5 and 6). As outlined in previous sections, a vigorous debate has surrounded the question of view-dependency in object recognition and its theoretical implications. Many researchers have shown that recognition performance reliably drops with rotations from a familiar or trained view-point. Since both view-based accounts as well as structural description accounts predict performance costs for plane-rotations (Hummel & Biederman, 1992) the focus of research on view-dependency has recently shifted to rotations in depth.

Bulthoff and Edelman (1992) demonstrated recognition costs for rotations of novel objects in depth. Similarly, Tarr (1995) showed observers objects consisting of connected cubes, which differed in their spatial arrangement. He, too, found view-dependent performance that was consistent with a view-interpolation account. Participants were trained on certain views, and recognition performance for objects in new views dropped increasingly with the degree of rotation. However, it is not certain that depth-rotations necessarily produce decrements in performance. Biederman and Gerhardstein (1993) criticised studies using novel objects and proposed three 'conditions for invariance' claimed to be 'typical' of human object recognition: (a) objects must be decomposable into parts; (b) each object in the recognition set must be composed of a distinct configuration of parts; (c) different viewpoints of the same object must show the same configuration of parts (see section 1.4.3.3). Researchers responded to Biederman and Gerhardstein's (1993) proposal and tested recognition performance using experimental designs that satisfied their conditions for view-invariance. The bulk of the results from these studies strongly indicate recognition costs for depth-rotated objects (Hayward, 1998; Hayward & Tarr, 1997; Tarr, 1995; Tarr, Bulthoff, Zabinski, & Blanz, 1997) even when these conditions are satisfied; but see also Biederman (2000) for a critical review of view-dependent effects on recognition of depth-rotated objects.

The debate between proponents of structural description accounts and view-based accounts shows that the underlying nature of the view-dependency found for depth-rotations is still unclear. However, the hybrid theory of object recognition may offer a comprehensive solution for the effects of depth-rotation without making any further assumptions. Most depth-rotations should substantially change activation in the holistic component and lead to temporary reduction in priming or recognition performance. Therefore, unlike geon theory, the hybrid model would predict that any substantial change in depth orientation from prime to probe trial in a short-term priming (or sequential matching) paradigm should result in recognition costs; this is because the holistic representation will always be affected regardless of whether or not part changes occur. Thus, the model can potentially account for the results of studies that found recognition costs for matching depth-rotated objects (Hayward & Tarr, 1997; Lawson & Humphreys, 1996; Tarr et al., 1998). It would potentially also be able to account for findings that with longer ISIs between sequential presentations in picture matching the effects of rotation are attenuated (Lawson & Humphreys, 1996), since the activation of the holistic representation is assumed to be rather short-lived (Stankiewicz et al., 1998).

Similar to Biederman and Gerhardstein (1993), the hybrid model would also predict additional reductions in priming for attended objects after some depth-rotations. If parts are revealed or occluded between depth-rotated views the activation in the analytic representation should also be affected. According to the model, units in Layer 6i (independent geon array) fire maximally on repeated presentation of an object when activated by the same visible parts but less so if parts are missing or a new parts are visible. In these cases, the hybrid model predicts an interaction between attention and view change. If the number or type of visible parts of an object is changed after depth-rotations, then we would expect a larger reduction in priming between the attended conditions (reduction in analytic and holistic activation after depth-rotation) relative to the ignored conditions (reduction only in holistic activation after depth-rotation). The next series of experiments aims to explore whether the predictions derived from the hybrid model can account for priming patterns of attended and ignored objects that are presented in the identical view or rotated in depth between prime and probe view.

3.4.2 Experiment 7: Priming for Mirror Images

3.4.2.1 Introduction

The primary goal of this Experiment was to replicate the findings of Stankiewicz et al. (1998) in which mirror images were found to prime their original version in the attended conditions, but not in ignored conditions. Mirror images can be considered as depth-rotations if the object has an axis of symmetry that can be aligned with the line of sight, which was the case for most of the objects in this experiment. A new set of objects was used consisting of grey-level photorealistic objects instead of black-and-white line drawings (see Figure 18). Experiments with the latter type of stimuli have the advantage of being able to control for low-level differences between objects and views such as shading and texture. However, careful rendering and counterbalancing of realistically depicted objects and their views should indicate if the findings obtained with line drawings generalise to ecologically more plausible images.

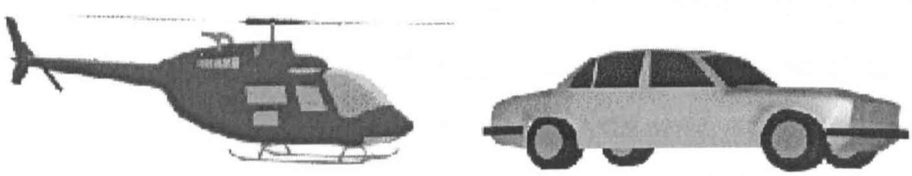


Figure 18: Examples of objects shown in Experiment 7.

There was an additional interest in this experiment. In former experiments, it was assumed that participants did not pay any attention to the ignored objects. In order to test this assumption, directly after the last trial participants were asked if they had been able to see the to-be-ignored object and if so whether they could identify it. If the manipulation was successful and participants only paid attention to the precued image, then ignored images should not be reportable.

The hybrid model predicts that attended images prime themselves and mirror-reflections, but ignored images only prime themselves, not their reflected versions. The effects of view and attention should be additive. The priming component resulting from the view change should be seen as an equivalent reduction in priming for both attended and ignored images. Again, theories of object recognition that rely on a single format of representation (Biederman, 1987; Tarr, 1995; Tarr & Pinker, 1989) would not predict priming in the ignored route.

3.4.2.2 Method

Participants

Twenty-eight native English speakers with normal or corrected-to-normal vision participated for pay or for credit in introductory psychology courses at Goldsmiths College University of London.

Material

Fifty-six common everyday objects were used. The objects were obtained from various open sources on the internet as 3D-meshes in the 3D Max (Autodesk) format. Each object was oriented in a standard 0° orientation, in which the main axis of elongation and/or the symmetry axis coincided with the line of sight. Each object was then rotated slightly between 5° and 10° in azimuth (y-axis) to give it a more canonical view (Blanz et al., 1999) as if the observer's vantage point was slightly elevated. Each object was then rotated in standard orientations in depth (z-axis), ranging from 30°, 60°, and 90° rotation of depth. All objects were rendered in 3D Max Studio (R3) using a 25 degree field of view, which gave a slight impression of perspective without drastically changing the perceived relative size of objects' parts. The objects were surface rendered with realistic overhead lighting but no cast shadows. The size of the images was then standardised. In this experiment, objects were only shown in the 60° viewpoint and their mirror reflection (see Figure 18).

Procedure

The basic procedure was identical to former experiments using the paradigm by Stankiewicz et al. (1998). Prime and probe objects were either shown in the identical view or a mirror-reflected view. Each object was placed into one of seven conditions (attended—identical, attended—reflected, attended—not probed, ignored—identical, ignored—reflected, ignored—not probed, and unprimed). All objects appeared in all seven conditions equally often. The ordering of the trials was randomised for each participant, as was the pairing of attended and ignored objects on prime trials.

After reading and paraphrasing the instructions, the participant read the names of the objects on the screen and then received 12 practice trials. A departure from former experiments was that directly after the last trial participants were asked if they had been able to see the to-be-ignored object and if so whether they could identify it to establish whether the participants were paying attention to the ignored objects.

3.4.2.3 Results

Trials on which either the prime or probe responses were incorrect were excluded from the analysis (18.6 %) as were voice key errors (6.1 %). For all conditions, priming was calculated as the difference between each participant's mean latency in the unprimed (baseline) condition and the participant's mean latency in each of the other probe conditions.

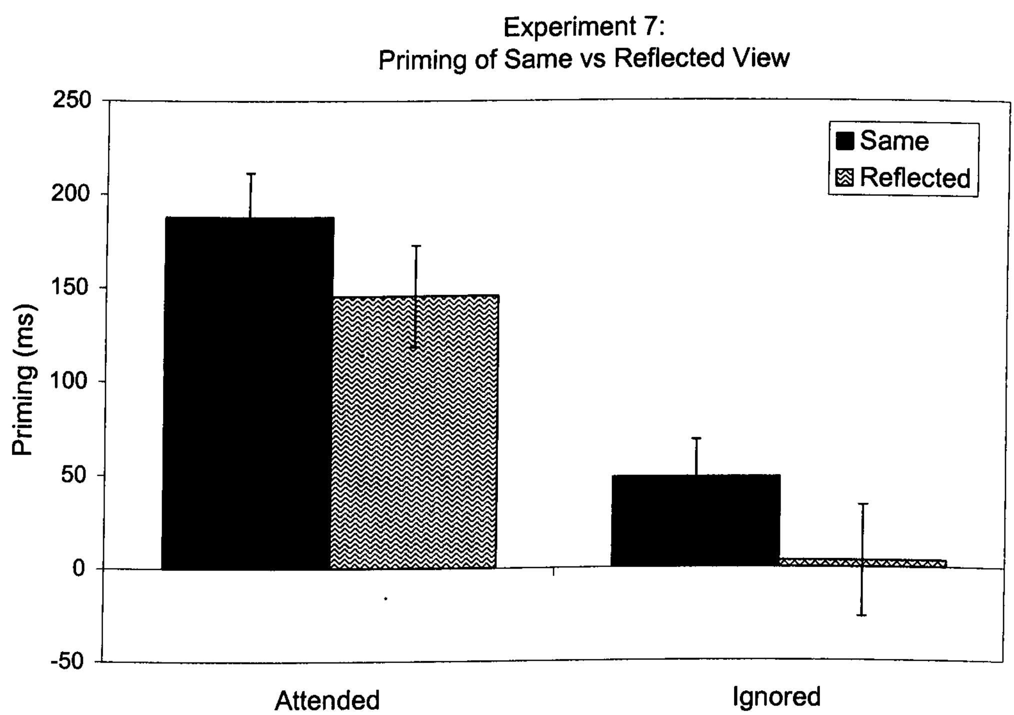


Figure 19: Priming means in ms and standard errors for Experiment 7 as a function of whether the object was attended or ignored in the prime display and whether the probe objects were presented in the same orientation or mirror-reflected (n = 28).

	Attended Same	Attended Reflected	Ignored Same	Ignored Reflected	Unprimed Same	Unprimed Reflected
% Errors (SE)	16.07 (2.94)	17.86 (3.37)	25.00 (3.86)	20.54 (3.42)	16.96 (3.87)	15.18 (3.48)

Table 9: Mean error rates for Experiment 7 (standard error in parantheses).

A 2 (Attention: attended vs. ignored) × 2 (View: same view vs. reflected view) within-subjects ANOVA was performed on priming latencies. The analysis revealed a reliable main effect of attention, $F(1, 27) = 42.92$, $p < .001$ and a main effect of rotation, $F(1, 27) = 4.77$, $p < .05$. The interaction between attention and rotation was not reliable, $F(1, 27) < 1$ (see Figure 19). A Friedman ANOVA on probe errors revealed no significant effects, Chi Sqr. $(3) = 2.56$, $p > .46$.

Matched pairs t tests showed priming reliably greater than zero in the attended-same, $t(27) = 7.93$, $p < .001$; attended-reflected condition, $t(27) = 5.37$, $p < .001$; and ignored-same condition, $t(27) = 2.41$, $p < .05$, but not in the ignored-reflected condition, $t(27) < 1$, $p > .05$. Thus, attended images in the prime display primed the probe image in both the same and the reflected view but ignored images primed the probe object only when it was presented in the same view.

The last trial was followed by the question of whether the participant had recognised the ignored object. Twenty-six observers responded with "No"; two responded with yes, but guessed the object's identity incorrectly.

3.4.2.4 Discussion

Experiment 7 replicates the results of Stankiewicz et al. (1998) that were obtained with line-drawings of objects: Attended objects prime both themselves and their reflected versions, whereas ignored objects only prime themselves but not their mirror-versions. The effects of attention and viewpoint were additive, meaning that the advantage for same versus reflected images was equivalent in both the attended and ignored conditions. The findings of an advantage for same views over mirror images replicate observations from matching tasks (Lawson & Humphreys, 1996). An additional result of Experiment 7 was the fact that observers seemed to comply with the instructions and did not pay attention to the non-cued (ignored) object: No participant could identify the ignored object in the last trial.

The results obtained with this stimulus set are important because they replicate the findings of Stankiewizc et al. (1998) with line drawings. Thus, grey-shaded images and line drawings of objects seem to produce the same priming effects, implying the possible involvement of edge-based representations (encoding parts derived from vertices and contours, e.g., geons) in object recognition. Edge-based approaches (Hummel & Biederman, 1992; Lowe, 1987), have been criticised (Sanocki et al., 1998) because in many studies or simulations their proponents mainly use line-drawings that contain no ambiguous contours or edges (e.g., resulting from shading and highlights) as compared to real objects (or photographs). The critical argument is that line drawings are created by artists who have used global interpretations that result from high-level vision in terms of meaningful visual structures (e.g., figure and ground, shadows, regions and volumes). These interpretations have probably eliminated many local ambiguities that are associated with edge extraction. Although Experiment 7 is not a direct test of whether realistically rendered grey-level images and line-drawings are treated equivalently by the visual system, the priming pattern previous results obtained with line drawings. Thus, the differences between priming conditions or lack thereof found in earlier studies with line-drawings using the attentional priming paradigm cannot be explained away by possible differences in extracting low-level features such as edges and contours.

The result of Experiment 7 encourages the use of grey-level images of photorealistic models of objects for studies in depth-rotation. Mirror-reflections may be considered as a special case of depth-rotation for objects that have an axis of symmetry which is perpendicular to the axis of rotation; this was the case for many stimuli in Experiment 7. In the following experiments, the effects of rotations in depth on priming of attended and ignored images will be examined more directly.

3.4.3 Experiment 8: Priming for Depth-Rotated Objects with Part Changes

3.4.3.1 Introduction

In previous experiments, the hybrid model was tested by manipulating only the holistic properties of an image. For example, changing the orientation (plane-rotation, reflection) or splitting of an image is predicted to affect mainly the holistic component, whereas the analytic component should not (or less so) be affected because the same parts were visible between prime and probe displays. Experiment 7 showed that the effects of viewpoint and attention are additive when the part-structure is kept constant between viewpoints by using mirror-reflections. The differences in priming between same and rotated conditions are due to the missing priming component from the holistic representation.

In contrast to mirror reflections, rotations in depth between study and test may affect the analytic representation because visible parts may be occluded or new parts may be revealed (Biederman & Gerhardstein, 1993). Depth-rotations should therefore provide an opportunity to further test the theory that two representations work in parallel because it affects both representational components (analytic and holistic) instead of just one (holistic). Note that according to the hybrid model manipulating the holistic properties should always affect priming for both attended and ignored conditions, whereas the manipulation of the structural properties should affect priming only in the attended conditions. For example, depth-rotations that produce changes in the structural description of the object (such when parts of an object are occluded between views) should affect the holistic representation as much as depth-rotations that do not change the structural representation. In contrast, part changes from study to test should only affect the representation that relies on dynamic binding, but not the holistic representation. In principle, this means the prediction of an interaction effect between attention and view for depth-rotations that cause a change in visible part structure.

The aim of this experiment is to test whether depth-rotation affects priming for attended objects (analytic route) more than for ignored objects (holistic route). The logic of the experiment is in three parts: First, according to the hybrid model, all viewpoint changes (except translation and scaling) should affect the holistic component. Second, because the holistic representation works with and without attention, changes in viewpoint by depth-rotations should equally decrease the amount of priming in both attended and ignored conditions compared to priming in the identical viewpoint. Third, depth-rotations that affect the perceived part structure of an object should additionally reduce the amount of priming for attended images, but not for ignored images. In summary, if a part-based representation is involved for attended images but not for ignored ones, then the manipulation of the part changes should affect priming for attended images (holistic and analytic change) more than for ignored images (holistic change only).

In Experiment 8 objects are rotated in depth to produce an altered part-structure between views. One problem is to assess the degree of part change for natural objects rotated in depth because it is not yet known at what scale part decomposition is achieved. For example, rotating a four-legged animal such as a horse may occlude or reveal a leg (large scale) or an ear (small scale). Moreover, the degree of part changes is not always systematically related to the degree of angular rotation for both natural (Lawson, 1999) and novel objects (Willems & Wagemans, 2001). For example, Figure 20 shows an object (a dog) in three different views rotated in depth. Although view b is rotated further away from view a than from view c, it shares more visible parts with view a, because two of the legs are hidden in view c. View c is called an accidental view (Biederman, 1987; Blanz et al., 1999) because the axis of symmetry and elongation are orthogonal to the line of sight. Such a complete side view of an object often occludes parts of an object or makes the extraction of parts (e.g., geons) more difficult. The criterion for an accidental view is that small changes in orientation produce considerable changes in the part structure (Biederman & Gerhardstein, 1993). This applies to complete side views because slight rotations would reveal new parts or new contours and surfaces of parts. Srinivas (1995) used a similar logic to create "part-occluded" objects.

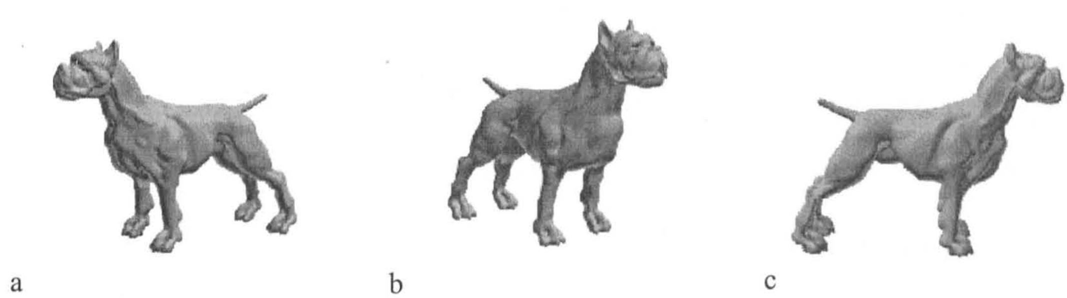


Figure 20: Three views of an example object as used in the pilot study for Experiment 8. View b is rotated further away (90°) from view a than from view c (60°), but the object shares more visible parts with view a, because two of the legs (and one ear) are hidden in view c.

To achieve a qualitative change in view orientation objects in Experiment 8 were depicted in two views. A complete side view (or "planar" view, Blanz et al., 1999) was chosen as a depth-rotated view that would be primed by a more conventional view or vice versa. Figure

21 depicts two objects and their corresponding views employed in Experiment 8. The effect of part-change was investigated in a pilot study.

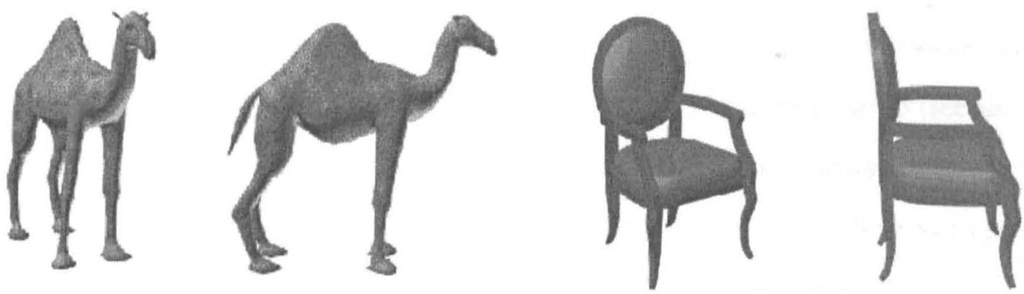


Figure 21: Examples of objects in the two views in which objects were shown in Experiment 8. Some parts (legs) are occluded with rotation.

The goal of Experiment 8 was to establish whether depth-rotations that cause changes in part structure affect the amount of priming more in the attended condition than in the ignored condition as predicted by the hybrid theory. In contrast, theories that rely on a single representation for object recognition based on views (e.g., Tarr & Pinker, 1989) or on structural descriptions (Biederman, 1987) would not predict that attended images will show a different decrease in priming than ignored images after depth rotations.

3.4.3.2 Method

Participants

Twenty-eight native English speakers with normal or corrected-to-normal vision participated for credit in introductory psychology courses at Goldsmiths' College University of London.

Materials

A pilot study was conducted to test whether the pairs of +30° and +90 views (Figure 21) produced perceived part changes. To assess if these two views of objects differ qualitatively from each other in their part structure compared to other pairs of depth-rotated views a rating study was conducted. Seven independent observers from the Goldsmiths' College student community were shown 87 objects in two pairs of views each on a computer screen. Three views were constructed by rotating an object -60°, +30° and +90° (views a, b, c in Figure 20) from a standard frontal view where the axis of elongation or the symmetry axis of the object coincides with the viewing direction of the observer (line of sight). The participants had to compare the views of each object. They saw an object in the -60° versus the +30° view in one trial as well as in the +30° versus the +90° view in another trial. The task of the observers was to indicate whether crucial parts of an object were visible only in one of the two views. To provide the participants with a scale of what is meant by a part they were introduced to the concept of geons and geon-built objects (Biederman, 1987) by using Figure 5 (see also Appendix 2). The order of trial (view-pair) presentation was completely randomised. Participants had as long as they wished to press the "P" key if they thought there was a part change or the "S" key if basically the same parts were visible in both views.

For the subset of the objects used in Experiment 7, two one-way ANOVAs on the factor View-pair were performed, first with objects and then participants as random factor. The factor levels were the two types of view-pairs separated by depth-rotation (-60° and +30° vs. +30° and +90°) with the number of "P" (i.e. part change) responses as dependent variable. An ANOVA with participants as random factor revealed a main effect of type of rotation, $F(1,6) = 26.35$, $p < .01$, as did the ANOVA over items, $F(1,86) = 49.07$, $p < .001$. Thus, objects shown in the +30° and +90° view pair were perveived to exhibit more part changes (mean 4.11, SE .29) across the two views than when shown in the -60° and +30° view pair (mean 1.91, SE .26).

In the priming experiment the same 56 objects as in Experiment 7 were used. Prime objects were depicted in two different views counterbalanced across two groups: objects were shown in an orientation rotated 90° off the line of sight in group 1, and rotated 30° off the line of sight in group 2 (Figure 21). The probes were displayed in either one of the two views depending on the experimental condition. The objects were counterbalanced across participants so that each object would serve in each condition equally often. The general set-up of the experiment was the same as in Experiment 7.

Procedure

The procedure was the same as in Experiment 7, except that participants were not asked whether they recognised the ignored image in the last trial. There were six priming conditions (attended-same, attended-rotated, ignored-same, ignored-rotated, unprimed-same-view and unprimed-rotated-view) in which each of the objects appeared equally often.

3.4.3.3 Results

In Figure 22 the priming results of Experiment 8 are given as savings in response times relative to the baseline (unprimed) condition. Trials on which either the prime or probe responses were incorrect were excluded from statistical analysis (20.2%), as were voice key errors (4.1 %). The baseline latencies for each of the two probe views was 866 ms (SE 38.5) for the 30° probe view and 857 ms (SE 27.7) for the 90° probe view (collapsed over groups).

For all conditions, priming was calculated as the difference between each participant's mean response time in the relevant baseline (unprimed) condition and the participant's mean response times in each of the corresponding priming conditions (Figure 22). A 2 (Group: prime view 30° vs. 90°) x 2 (Attention: attended vs. ignored) × 2 (View: same vs. rotated) mixed analysis of variance (ANOVA) revealed no reliable effect of group (i.e. the two orientations primed their corresponding probe equivalently), $F(1, 26) < 1$, a reliable main effect of attention, $F(1, 26) = 47.15$, $p < .001$, and View, $F(1, 26) = 8.37$, $p < .001$. The only significant interaction was between Attention and View, $F(1, 26) = 5.04$, $p < .05$. The difference between the attended-same and attended-rotated conditions (collapsed over groups) was statistically reliable, $t(27) = 3.73$, $p < .001$, but not the difference between the ignored-same and ignored-rotated conditions, $t(27) = 1.27$, $p > .05$. A Friedman ANOVA on probe errors for each priming condition revealed no significant effects, Chi Sqr. (3) = 1.48, $p > .68$.

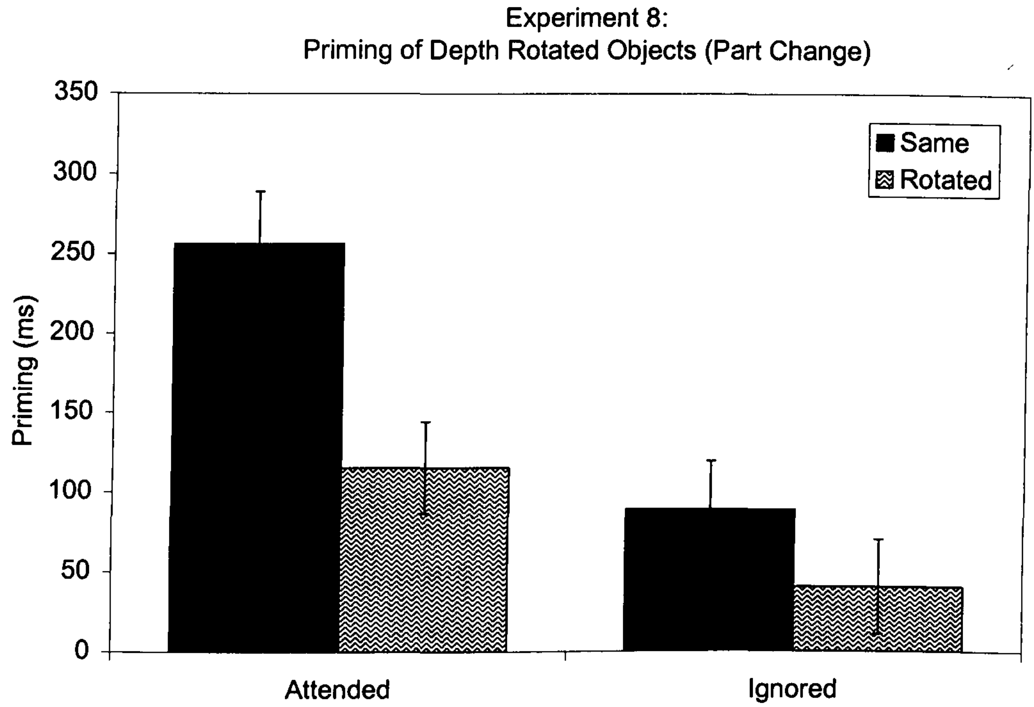


Figure 22: Priming means in ms and standard errors Experiment 8 as a function of whether the object was attended or ignored in the prime display and whether the probe objects were presented in the same orientation or rotated in depth (n = 28).

	Attended Same	Attend. Rotated	Ignored Same	Ignored Rotated	Unprimed Same	Unprimed Rotated
% Errors (SE)	16.07 (3.46)	20.54 (3.87)	21.42 (3.57)	25.89 (3.74)	16.07 (3.90)	21.43 (4.40)

Table 10: Mean error rates for Experiment 8 (standard error in parantheses).

An ANOVA over all errors revealed only a main effect of Group, $F(1, 26) = 6.90$, $p < .05$. None of the interactions approached significance, all $F < 1$. Independent t-tests for errors collapsed over priming conditions revealed a significant group difference for prime errors, $t(26) = 2.75$, $p < .05$, but not for probe errors, $t(26) = 1.31$, $p > .05$. Thus, the participants in the group with the prime views in the 30° orientation made more errors across conditions.

Matched pairs t tests were conducted on each priming condition to determine which type of prime display caused savings in response time for the probe display. Priming was reliably greater than zero in the attended-same condition, $t(27) = 7.78$, $p < .001$; attended-rotated condition, $t(27) = 4.00$, $p < .001$; and ignored-same condition, $t(27) = 2.95$, $p < .01$, but not in the ignored-rotated condition, $t(27) = 1.37$, $p > .05$. Attended images in the prime display primed the probe image in both the same and rotated view, but ignored images primed the probe object only when presented in the same view.

3.4.3.4 Discussion

The results of Experiment 8 replicate the previous findings of priming for attended images in the same view and in a changed (here: depth-rotated) orientation, and that ignored objects were only primed if they are depicted in the same view. However, the data show a difference in priming effects between ignored and attended images; the difference between identical and rotated views was significantly greater for attended than for ignored images. This priming pattern is in line with the prediction of the hybrid model that depth-rotations may cause qualitative changes in representation which are more detrimental for priming when attending to an image than when ignoring it.

The finding that some depth-rotations may reveal priming differences in attended compared to ignored conditions confirms the prediction of the hybrid model that two qualitatively different representations are employed. In contrast, if the visual system always relied on a single type of representations based on metric properties to align new views with stored views we would expect additive effects of attention and viewpoint. Most current view-based accounts do not specify the role of attention and therefore could not have predicted the results described here. However, if attention plays a role in matching representations based on metric properties, one would expect less effects on rotation in attended conditions relative to ignored conditions, because attention would serve to aid the matching process (Olshausen et al., 1993).

The results also do not fit entirely with geon theory. Part-based accounts would not have predicted priming in the ignored predictions because structural descriptions rely on attention to actively bind local features into parts and then parts and relations to objects. However, this account would have predicted the larger rotation costs obtained for rotations that include accidental views (Biederman, 2000) compared to mirror reflected views. Accidental views that cause part occlusions may change the activation pattern of a GSD representation between prime and probe view, causing a reduction in priming compared to exactly the same parts being visible in both events (e.g., in mirror images). Thus, the results for the attended conditions are in line with structural description accounts.

The results found here cannot be attributed to difficulties with the specific orientation in depth of objects because the priming pattern was the same independent of the prime and probe views employed. The present experiment as well as other studies (e.g., Hayward, 1998) have found that planar views (as the +90° view) do incur more recognition costs after a change in orientation in prime/probe paradigms than do other non-planar and non-foreshortened (canonical) views. Thus, it seems that planar views are not generally harder to recognise or less familiar than non-accidental views. Rather, the finding that the priming differences between attended rotated images are larger than between ignored rotated images suggest that qualitative changes affect the attended recognition route more than the ignored route. That part changes such as occlusion account for the striking priming differences in the attended conditions seems supported by the findings of Srinivas (1995; Experiment 2). She manipulated part changes over depth-rotations in a similar way as in the present experiment (selecting views in which parts of photographed objects were occluded by other parts). Srinivas' participants were shown photographs of object rotated in depth (67°, 130°, and a part-occlusion rotation) for 300 ms in the prime display. Matching object identity did not affect latencies in the view conditions with all parts visible, but increased response times in the part-occluded condition.

There was no reliable difference in priming between the ignored-rotated conditions in Experiment 8. Priming for the ignored condition in the hybrid model is due to the holistic surface map which responds to the visible surfaces in an image. Priming should therefore be maximal for identical images (although regardless of image size and position in the visual field), and decrease with changes in orientation because units in the holistic surface map respond less strongly when surfaces are changed or removed from their receptive fields. However, the activation pattern may not change completely between two views (from prime to probe) inasmuch there is an overlap of the same surfaces. This may explain the lack of statistical difference between the identical and rotated conditions. An estimate of metric similarity could be provided with the Lades et al. (1993) model. Indeed, the two viewpoints used in this experiment were rotated from each other within one side off the line of sight, which means that some parts and surfaces correspond to the same relative location on a hypothetical grid of units. For example, the trunk and head of an animal in both views would correspond to similar units when superimposed on a (flexible) holistic surface map (see

Figure 7). Thus, the hybrid model generally predicts that the larger the metric changes (given no additional analytic changes), the larger the differences in priming for ignored conditions.

Experiment 9 was designed to test the possibility that the lack of priming difference in the ignored conditions is due to an overlap of holistic properties. If the overlap in priming (i.e. non-significant difference between identical and rotated conditions) for the ignored trials in Experiment 8 was due to an overlap of surfaces or other metric features in a holistic representation, then a larger rotation between identical and rotated view should accentuate this difference in priming. Larger view-differences than those used in Experiment 8 with rotations across the line of sight should produce larger metric changes between views which in turn should accentuate priming differences between identical and rotated views. At the same time such a rotation should not affect the structural representation if the part structure does not change substantially across these views (Biederman & Gerhardstein, 1993). In this case, the hybrid model would predict additive effects of priming for attention and view as found in previous Experiments and in Stankiewicz et al. (1998). Only the holistic component - which works with and without attention - is affected by view changes. Therefore the short-term priming paradigm in Experiment 9 employed familiar views that differed considerably in depth orientation but not in part structure.

3.4.4 Experiment 9: Priming for Depth-Rotated Objects without Part Changes

3.4.4.1 Introduction

This experiment was designed to test whether the priming pattern observed for mirror images (Experiment 7; Stankiewicz et al., 1998) can be replicated with depth-rotated objects. The critical assumption is that depth-rotated objects - like mirror images - can be shown in orientations that reveal equivalent part structures but differ significantly in their metric similarity. To this end, pairs of depth-rotated views from the rating study were used that elicited fewer "part-change" responses than the pairs of orientations used in Experiment 8. In addition, there was a greater degree of angular separation (90°) between these views than in Experiment 8 (60°). The two views in Experiment 9 were "off-axis" views that showed the front and the side of an object, that is they are shown in orientations that do not fall in the line of sight or perpendicular to it. Objects in these orientations are usually easier to recognise when they are shown in "off-axis" (or "canonical") views. In these views a large number of surfaces are usually maximally visible and they are often rated as the most typical views in which objects appear (Blanz et al., 1999; Palmer et al., 1981; Verfaillie & Boutsen, 1995). These views have been found to be the easiest to recognise (Boutsen, Lamberts, & Verfaillie, 1998; Palmer et al., 1981). In addition, only those objects were used that were bilaterally symmetric such that a rotation across the line of sight produces a view in which roughly the same parts are visible (see Figure 23).

The hybrid model of Hummel (2001) would predict additive effects of viewpoint and attention for these views rotated 90° from each other. The analytic representation should not or only be slightly affected by these view changes, which means that its contribution towards overall priming remains equivalent between the identical and the depth-rotated view. In contrast, the holistic representation should change considerably due to the depth-rotation, because different surfaces of the same object map to the units in layer 5s. Its contribution to priming should go towards zero compared to the presentation of an identical view in the probe display.

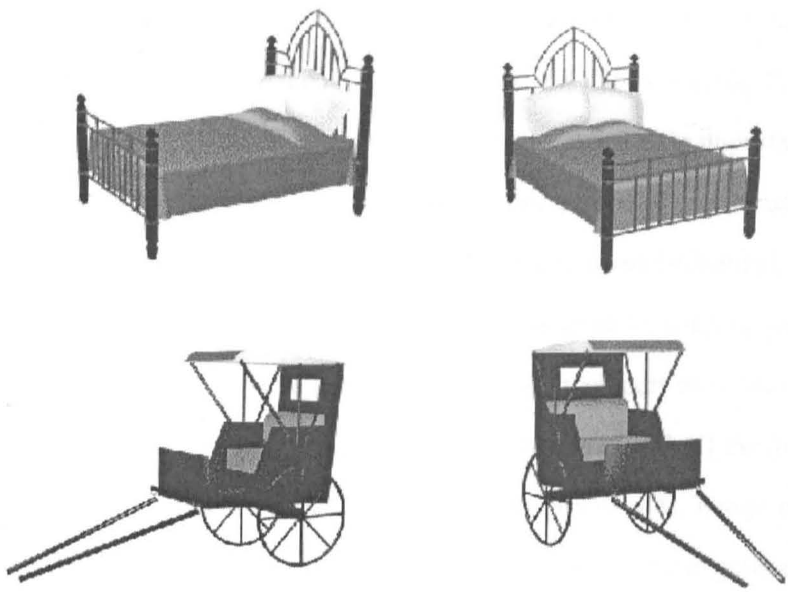


Figure 23: Examples of bilateral symmetric objects in the two orientations used in Experiment 9; objects are rotated from each other 90° (-60° and 30° off the line of sight).

View-based models and structural description models do not necessarily predict priming in the ignored conditions (see discussion of Experiments 3 and 6). Geon theory would predict no or only slight differences in priming between the same and rotated attended views. The predictions of multiple-views-plus-transformation accounts are not clear. The likely alternatives are a) no effect of rotation because both views of the objects are canonical and probably highly familiar (Blanz et al., 1999; Palmer et al., 1981) or b) a reduction of priming is predicted for the rotated attended condition, either because mental transformation is required to match prime and probe view or because of an identity advantage of the same versus the rotated stored view, that is the non-identical views are less activated.

3.4.4.2 Method

Participants

Forty native English speakers with normal or corrected-to-normal vision participated for credit in introductory psychology courses at Goldsmiths College University of London.

Materials

A set of 84 objects was used containing most of the objects as in Experiment 7 and 8. All objects were shown in two standard views (see Figure 23). These were separated by 90° from each other, with an orientation which was created by rotating +30° (one of the views in

Experiment 8) and the second -60° (the standard view of Experiment 7) from the line of sight (which coincides with the line of symmetry). Of the 84 objects, 30 objects were used as target objects, and the rest were used as filler items (i.e. in unprobed conditions). The target objects were placed into 5 subsets which appeared equally often across participants in all conditions (attended-identical, attended-rotated, ignored-identical, ignored-rotated, and unprimed). The filler items were randomly assigned for each subject. There were two standard views counterbalanced across 2 groups: objects were shown in an orientation rotated -60° off the line of sight in group 1, or rotated 30° off the line of sight in group 2 (corresponding to one of the views used in Experiment 8). The prime and probe objects were displayed in either one of the two views depending on the experimental condition. The two views were counterbalanced across participants so that each view would serve in each condition equally often.

Procedure

The procedure was the same as in Experiment 8, except that the probe view in each group of participants was always depicted in the same general orientation and the prime view was either identical to the probe view or rotated.

3.4.4.3 Results

Trials on which either the prime or probe responses were incorrect were excluded from statistical analysis (7.92 %), as were voice key errors (3.83 %). The group means for the baseline probe views were 770 ms (SE 31.8; for the -60° orientation) and 803 ms (SE 29.7, for 30° orientation), a non-significant difference, $t\ (38) < 1$. Both views elicited similar latencies in the baseline conditions. Figure 24 shows the priming results of Experiment 9 as savings in response times relative to the baseline (unprimed) condition.

A 2 (Group: probe view 30° vs. -60°) x 2 (Attention: attended vs. ignored) × 2 (View: same vs. rotated) mixed analysis of variance (ANOVA) revealed no reliable effect of Group (i.e. priming patterns in the two probe orientation groups did not differ), $F\ (1, 38) < 1$, a reliable main effect of attention, $F\ (1, 38) = 105.13$, $p < .001$, and View, $F\ (1, 38) = 10.79$, $p < .01$. There was no statistically reliable interaction. A Friedman ANOVA on probe errors revealed no significant effects, Chi Sqr. (3) = 1.50, $p > .68$.

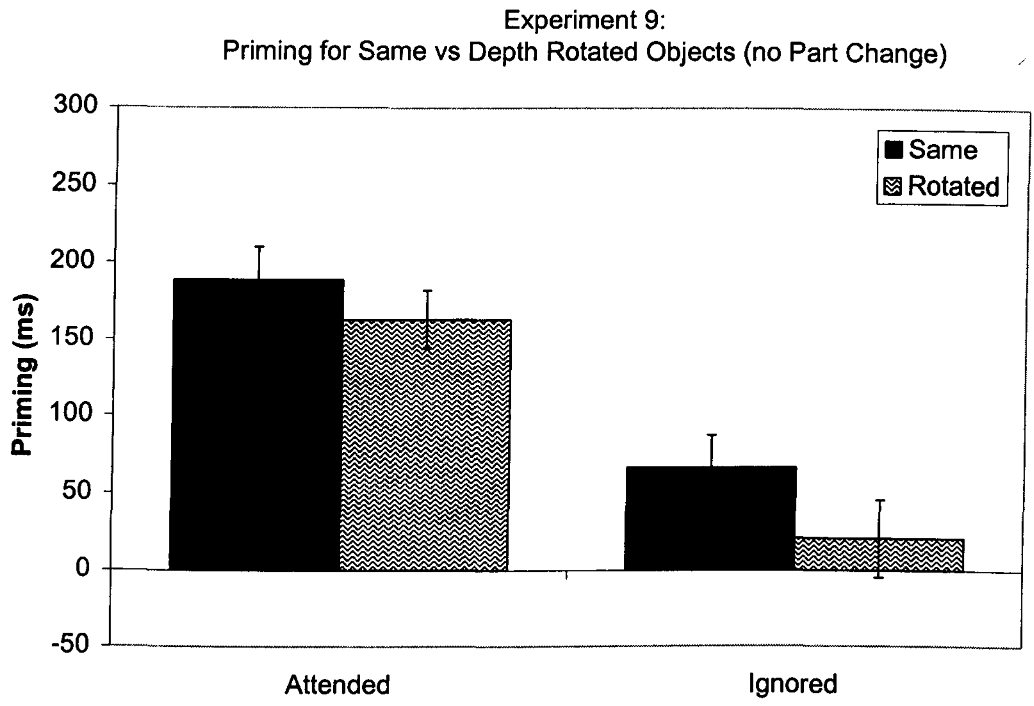


Figure 24: Priming means in ms and standard errors Experiment 9 as a function of whether the object was attended or ignored in the prime display and whether the probe objects were presented in the same orientation or rotated in depth (n = 40).

	Attended Same	Attend. Rotated	Ignored Same	Ignored Rotated	Unprimed Same
% Errors (SE)	8.74 (2.15)	8.33 (1.89)	9.17 (1.46)	6.25 (1.54)	7.08 (1.45)

Table 11: Mean error rates for Experiment 9 (standard error in parantheses).

Matched pairs t tests were conducted on each priming condition (collapsed over groups) to determine reliable savings in response time compared to the baseline. Priming was reliably greater than zero in the attended-same condition, $t(39) = 8.61$, $p < .001$; attended-rotated condition, $t(39) = 7.93$, $p < .001$; and ignored-same condition, $t(39) = 3.47$, $p < .01$, but not in the ignored-rotated condition, $t(39) < 1$. Probe images were successfully primed by attended images in the prime display shown in both the same and rotated view, but ignored images primed the probe image only when presented in the same view.

Of critical interest in this experiment was whether there was a difference in view conditions after increasing rotation in depth compared to Experiment 8. The difference between the attended-same and attended-rotated conditions was statistically reliable, $t(39) = 2.41$, $p < .05$, as was the difference between the ignored-same and ignored-rotated conditions, $t(39) = 2.30$, $p < .05$.

3.4.4.4 Discussion

The results of Experiment 9 replicate earlier findings obtained with mirror-images and plane-rotations for no-base objects: Probe objects were primed by previously attended images presented in the same view as well as in a changed (here: depth-rotated) view, whereas a probe image was only primed by an ignored prime if it was presented in the same view. The effects of attention and view were additive. Images primed themselves more in the same orientation than in a rotated view; this was found for both attended and ignored objects.

The priming pattern shown in Figure 24 is clearly predicted by the hybrid model of object recognition because the views were chosen in such a way that they reveal generally the same part-structure and should therefore produce additive priming effects for attention and viewpoint. The holistic surface map is activated in parallel with the independent geon array when presented with the prime object. On subsequent presentation of the same object in the same view, both the IGA units and the HSM units benefit from the previous presentation resulting in faster recognition. Presented with the identical but depth-rotated object, the same parts (and relations) are presented to the IGA, which produces faster recognition in this route, resulting in an activation (i.e. analytic priming component) that is equivalent to that of the identical view. In contrast, the locations of surfaces (if projected on a 2D grid) have changed considerably after depth-rotations. Therefore, the activation pattern of the HSM units is very different between prime and probe trial and no priming from the holistic component is predicted.

The pattern of results found here would have clearly not been predicted by geon theory for the same reasons as in Experiment 7. First, priming for ignored identical images would have not been predicted by structural description accounts that stress the role of attention for dynamic binding. Second, geon theory would have arguably not predicted a reduction in priming for depth-rotated objects in the attended conditions. Both views showed roughly the same part structure and were far from accidental views. Of course it could be argued that every depth-rotation changes the visibility of some parts, and may produce spurious effects of view-dependence. However, the fact that this experiment obtained the equivalent priming pattern as Experiment 7 with mirror-images (which by definition show exactly the same parts) make this counterargument less convincing.

View-based theories could explain the viewpoint effects in the attended conditions with performance costs due to mental rotation or other time-consuming matching procedures such view interpolation. However, the task employed was a not a matching task but a naming task. There were no performance costs for one view relative to the other as measured by the response to unprimed conditions. If both views seem to be equally similar to a stored template, there is no need to assume normalisation processes with a higher cost for a rotated view. Again, the argument that identical images prime themselves more than non-identical seems an unlikely explanation in terms of view-based accounts given the results of Experiments 3 and 6.

The results are also in line with other studies employing depth-rotations that are similar to the ones presented here. For example, Lawson and Humphreys (1998) studied effects of long-term (in the order of several minutes) priming for line-drawings of common objects rotated in depth. Although the interest of these studies was in priming for foreshortened views, inspection of their results for views almost identical to those employed in Experiment 9 (termed 60° and 150° or 30° and 120° in Lawson & Humphreys, 1998) show that responses in the probe block were roughly equivalent for these (but not foreshortened) views. Biederman and Gerhardstein (1993) used somewhat different orientations in depth in their Experiments 1 and 2. They found no differences in long-term priming effects over rotations ranging from 33.75° to 135° for common objects if there were no part changes, but view-dependent effects if different parts were visible between the study and test view.

3.4.5 General Discussion of Experiments 7-9

Experiments 7 to 9 used the short term priming paradigm for attended and ignored objects shown in identical views or rotated in depth. The results agreed with the predictions of the hybrid model. Large view changes caused an equivalent reduction in priming for both attended and ignored images, provided these view-changes exhibit roughly the same visible parts. However, if considerable part changes occur between prime and probe view, then a significant reduction in priming is observed only for the attended view. These results indicate that attended objects are treated qualitatively different from ignored objects which is in line with the notion of an analytic representation for attended and a holistic representation for ignored images.

A possible limitation of both Experiments 7 and 8 was that objects were depicted in pairs of orientations whose co-occurrence may be considered unusual; mirror-reflections of canonical views and accidental (planar) views. The responses to these view-pairs may be considered as a rather rare and special cases of depth-rotations. The results of Experiment 9, however, show that the priming pattern observed with mirror images in Experiment 8 were not accidental and can be extended to other situations in which depth-rotated objects have to be recognised. A further limitation may be seen in the indirect manipulation of same vs. different visible part structure over depth-rotations for images of common objects. Here, the extent of part-changes were predicted from scrutinising the stimulus set, reviewing results from previous studies employing depth-rotated views, and observer ratings. There is a strong consensus from these sources that for many objects shown in view pairs including the 90° depth orientation some parts are occluded.

Many studies using common objects rotated in depth ignore possible part occlusions or assume that all views show all the critical parts. For example, Hayward (1998) used almost identical viewpoints as Experiment 8 (30° and 90° orientations) in a sequential matching task (view difference 0° and 60°, see Figure 25). He found that objects in planar views (60° view difference to the study view) were matched more slowly than a non-planar view rotated 180° from the standard view. On the assumption that all three viewpoints show the same parts (which on inspection of his stimuli they certainly did not; see Figure 25b), Hayward's conclusion was that common outline shape was the crucial factor for object constancy across view changes in 3D, not part structure. The present results in the attended conditions of Experiments 8 and 9 mirror those of Hayward's Experiment 3 (for +30° and +90° views), namely greater reduction in priming for rotations to an accidental (planar) than to other views, despite the fact that the angular difference between prime and probe view was smaller for accidental views. Hayward proposed that view-dependent priming effects were not due to part-changes but to the extent of similarity of global outline shape between views. However, our conclusions are different from those of Hayward's (1998). The lack of a significant difference in the ignored conditions between the two views in Experiment 8 clearly argues against a global, holistic difference due to outline shape. They fit with the assumption that the difference between those views observed in Hayward's experiments and in the attended conditions here (Experiment 8) are due to changes in visible parts (see Figure 25).

Moreover, similar to Hayward's (1998) study, the priming difference between attended objects in the same and depth-rotated view in Experiments 8 and 9 were not predicted by the angular degree of rotation. Thus, even if we ignore the part changes between the different viewpoints in the present (Experiments 8 and 9) and in Hayward's (1998) study, the results would not be in line with view-based accounts which assume a linear increase (e.g., Tarr & Pinker, 1989) or even an accelerated increase (e.g., Poggio & Edelman, 1990) in latencies after orientation changes.

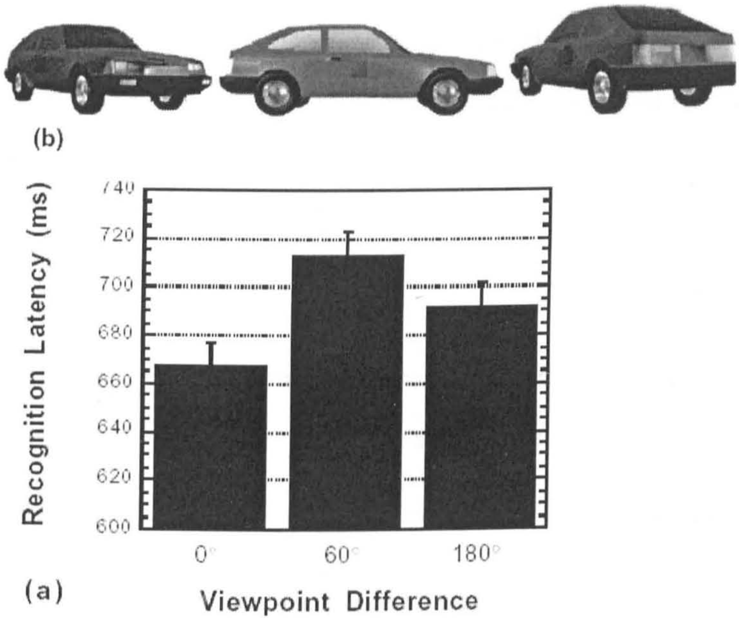


Figure 25: Example of an object shown in three viewpoints used in Hayward's (1998) Experiment 3 (b) and matching performance in ms (a). The 0° and 60° viewpoints in which the common objects are shown (b) correspond almost exactly to viewpoints defined as 30° and 90° views in Experiment 8 of this investigation.

Although the data reported here clearly demonstrate a qualitative difference between attended and ignored objects after depth-rotations, it is not necessarily the case that this difference is due to parts defined as geons. Alternatively, observers may have noted the appearance and disappearance of surfaces or certain vertices after depth-rotations. Future research needs to employ a stricter criterion for part changes in natural objects to establish whether the effects here are due to representations resembling structural descriptions. Nevertheless, the current results are clearly in line with the predictions derived from the hybrid model of object recognition, and gives further evidence for its generality.

4. Chapter 4: General Discussion

4.1 Summary of Results

The priming studies described above have shown considerable evidence for the role of attention in object constancy. The data are also consistent with the hybrid model of object recognition (Hummel, 2001; Stankiewicz et al., 1998) which postulates two qualitatively different representations in object recognition. While some data taken by itself could be explained by other theories, the fact that all were derived from and tested the hybrid model of object recognition tells a compelling story. There are several main findings:

1. The format of representation mediating object recognition depends on attention. Attended and ignored images prime their identical versions, but only attended images also prime their split, plane-rotated, and depth-rotated versions.
2. The priming obtained for attended objects has analytic properties. Split images show visual priming only when attended, and part changes affect attended images more than ignored images.
3. Priming for ignored images is shape-specific, because ignored objects do not prime an image of another object with the same name but different shape.
4. Ignored images make automatic contact with stored holistic memory representations. Ignored images do not prime themselves in unfamiliar views nor do ignored split images prime themselves.
5. The two representations - analytic and holistic - seem to be independent of each other and can work in parallel. The priming effects of attention and change of holistic properties (configuration, viewpoint) are additive. Priming effects are non-additive if qualitative properties (i.e. parts) of an object's structural description change across views.

Together, these findings suggest that priming differences between attended and ignored objects are a result of two qualitatively different representations. The importance of these data for the hybrid model of object recognition will be discussed in the next section. The following sections consider the wider implications of the results for other models of object recognition, for theories of attention, and for research on brain functions. The chapter will close with an evaluation of the hybrid model.

4.2 Implications of the Results

The priming results for plane and depth-rotated objects confirm the findings of Stankiewicz et al. (1998) that attended objects prime themselves in the same and a different view, but ignored images prime themselves only in the same view. Therefore, the predictions for view-dependence in the ignored route are not limited to mirror-reflections (Stankiewicz et al., 1998) but also hold for plane- and depth-rotations. The model can therefore potentially account for a variety of performance patterns after changes in orientation. The fact that these priming effects were the same across plane and depth-rotation as well as configural changes suggests that the same mechanisms for achieving view-invariance may be used. Object constancy across view changes and configural changes can only be achieved by attending to an object. However, the visual system can still recognise objects without attention if no compensation for orientation or configural change is needed. Therefore, the results clearly indicate that there are at least two routes to object recognition depending on attention.

The results from Experiments 1 to 3 confirm the prediction of the hybrid model that these two routes have separate analytic and holistic properties. The fact that split objects visually prime their identical versions only when they are attended but not when ignored shows that the representation generated in response to attended images can compensate for configural changes (such as splitting an image), whereas the representation for ignored objects cannot. Moreover, part-changes after depth-rotation affect priming for attended images more than for ignored images (Experiment 8). Thus, representations mediating object recognition for attended images have analytic (i.e. non-holistic) properties (e.g., as in a structural description), whereas the representation mediating recognition of ignored images are strictly holistic (e.g., as in a view-based representation).

The priming found for ignored identical images cannot be explained by a simple "identity benefit" due to low-level processing, that is, they cannot be attributed to the hypothesis that identical images always prime themselves more than non-identical ones. Split images prime themselves when attended, but not when ignored, just as images of objects that are rotated from their upright (canonical) orientation prime themselves only when attended. These findings suggest that only previously stored representations prime themselves in the ignored route. This is corroborated by Stankiewicz and Hummel (2002) who showed that non-identical images of an object (translated and scaled version) can prime themselves as much as identical images prime themselves. Thus, priming for identical images does not rely solely on low-level processes, but rather on whether this view has been stored in memory.

The priming patterns obtained with plane-rotated images in Experiments 5 and 6 also indicate that the reduction of priming for split images found in Experiment 1, 2 and 3 are unlikely to be due to differences in difficulty levels. The results cannot be explained by the fact that split images are simply harder to identify than intact images because for example, they may be initially perceived as two items rather than one. No-base objects have no preferred familiar view and are recognised equivalently well from virtually any orientation in the picture plane. In Experiment 5 rotated no-base objects showed exactly the same reduction in priming as did split images. Together with the results on priming for identical versus mirror-reflected images (Stankiewicz et al., 1998) these data provide converging evidence that the visual system employs two qualitatively different representations to deal with the described changes in retinal input.

The fact that priming is not due to low-level-processes also indicates that the holistic route requires access to LTM. In addition, ignored images do not prime an image of an object with the same name but a different shape. Therefore, the priming observed in the ignored route does not extend to semantic representations (Experiment 2). The lack of semantic priming between different exemplars of an object and the fact that images in unfamiliar views or configurations do not prime themselves indicate that priming manifests itself between perception and memory, that is priming for ignored images is mediated by a holistic visual representation of object shape (Experiments 3 and 6). This is consistent with simulation studies (Hummel, 2001) in which priming was implemented between layers 5 and 6 in the hybrid model.

The priming effects of attention and viewpoint were additive if only holistic properties were changed from prime to probe display. The priming effects were additive for attention and configural (split images) change as well as for changes in orientation. This, again, is in accordance with the prediction of the hybrid model that two qualitatively different representations are generated in parallel. Attended images generate both an analytic as well as a holistic representation, whereas ignored images only generate a holistic and view-dependent representation.

The priming effects of attention and rotation in depth were non-additive if analytic properties of the object changed from prime to probe display. The reduction in priming for depth-rotated compared to identical views was significantly higher for attended than ignored images, but only if the rotated views differed in terms of their visible parts. Depth-rotated views that did not differ significantly in their part structure showed similar reductions in priming across view-changes compared to mirror reflections. This result supports the conclusion derived from priming with split images that attended images are mediated by part-based representations.

Together, the results of Experiments 1 to 9 suggest that visual attention affects the qualitative nature of the visual representation generated from an object's image. A possible alternative account for the original findings of Stankiewicz et al. (1998) and of some studies described here (Experiments 1, 4, 5, 7, and 9) is that the visual system represents shape in only one format, but that attention modulates the efficiency of the processes that generate an object representation from its image or match it to memory. According to this account, priming is simply greater for identical images than for non-identical images, and greater for attended images than for ignored images. In short, differences in priming between conditions would not be due to qualitatively different representations but would reflect a quantitative drop in priming when objects are unattended and presented as a different image. This explanation of the priming data would in principle fit in the framework of models of object recognition which assume only a single representation of object shape, such as structural descriptions (e.g., Biederman, 1987; Hummel & Biederman, 1992; Marr, 1982) or view-based templates (e.g., Edelman, 1998; Olshausen et al., 1993; Poggio & Edelman, 1990; Tarr & Pinker, 1989; Ullman, 1989, 1998). Such an alternative account of the priming data would predict uniform effects of attention and view change on patterns in visual priming. However, Experiments 3, 6, and 8 showed differential effects of attention on patterns of view invariance in visual priming as predicted by the hybrid model and its postulated multiple representations. Although attended images in general prime more than ignored images, identical images do not necessarily prime more than non-identical (i.e., rotated and split) images. In addition, some changes from identical to nonidentical images (such as depth-rotation with part changes) elicit different changes in priming for attended and ignored objects. Thus, the results challenge any model of object recognition that relies on a single type of shape representation.

There are further reasons to question the alternative interpretation of the obtained priming effects as no model based on a single format of representation seems able to account for the range of different priming patterns observed in this investigation. Object shape representations that are solely based on parts and relations require attention for dynamic binding and therefore would not predict priming in the ignored conditions. These accounts would also not predict the differences in priming found for attended mirror-images and rotated images of no-base objects because the structural descriptions would be the same for all views. Similarly, theories that propose that shape representation is completely image-based would have difficulties in explaining the priming patterns obtained in the present investigation. View-based theories would not predict the visual priming effects observed for split images (Experiments 1 to 3) because a view-based representation is by definition a holistic (i.e. indivisible) image. View-based theories would also have problems to account for the differences found for priming of attended and ignored depth-rotated images when parts change across views.

In Table 12, the results of the experiments in the present investigations are summarised qualitatively by assessing whether or not they can in principle be accounted for by predictions from geon theory (Biederman, 1987; Hummel & Biederman, 1992), multiple views theory (Tarr & Bulthoff, 1995; Tarr & Pinker, 1989), and the hybrid model (Hummel, 2001; Hummel & Stankiewicz, 1996a). The results summary table was made under the assumption that priming for ignored images is not due to a simple identity advantage on the basis of low-level matching, as ruled out by the findings in Experiments 3 and 6 and by Stankiewicz and Hummel (2002).

Theory						
	Recognition by Components (RBC)		Multiple Views Theory		Hybrid Model	
Experiment	Attended	Ignored	Attended	Ignored	Attended	Ignored
1 Split vs. Intact Intact probe	?	-	✓	-	✓	✓
2 Split vs. SNDE Intact probe	✓	✓	-	✓	✓	✓
3 Split primes Split Intact primes Intact	✓	-	-	-	?	✓
4 Naming rotated objects	✓	N/A	✓	N/A	✓	N/A
5 Same vs. rotated No-base Objects	-	-	✓	-	✓	✓
6 Upright vs. Rotated Base objects	✓	-	?	-	?	✓
7 Identical vs. mirror reflection	✓	-	✓	-	✓	✓
8 Identical vs. depth rotated (part change)	✓	-	?	?	✓	?
9 Identical vs. depth rotated (same parts)	✓	-	✓	-	✓	✓

Table 12: Summary of experiments and how they comply with predictions of geon theory (Hummel & Biederman, 1992), multiple views theory (Tarr & Pinker, 1989), and the hybrid theory (Hummel, 2001). Key: "✓" = compatible with predictions; "-" = incompatible with predictions; "?" = predictions or results not clear. Note: Compatibility was assessed under the assumption that priming for ignored images cannot be entirely attributed to low-level perceptual processes.

In general, the results fit well with the predictions of the hybrid model of Hummel (2001; Hummel & Stankiewicz, 1996a) that object recognition is mediated by two relatively independent representations of object shape. One is a structural description specifying an object's parts and their spatial relations (Hummel & Biederman, 1992). This representation requires attention to bind local features in the image (contours, vertices, etc.) into an object's parts and code these parts independently of one another and of their relations. Accordingly, it is invariant with translation, scale, left–right reflection, and other changes in viewpoint that do not alter the structural description. The second component of the hybrid model is an holistic representation that codes an object's parts in respect to their co-ordinates in a coarse reference frame that is orientation-sensitive. This representation does not depend on visual attention because it separates an object's parts according to their locations in the reference frame. Thus, visual attention for dynamic binding is not necessary to maintain the independence of an object's parts. However, coding an object's features or parts separately at each location in the reference frame makes this holistic representation more sensitive to variations in viewpoint than the propositional representation (Hummel, 2001; Stankiewicz & Hummel, 2002).

The experiments presented here have extended previous tests of the hybrid model in a number of ways. First, the experiments further support the hybrid model's prediction that two qualitatively different representations work in parallel rather than in serial. Second, they show that the predictions for view-dependence in the ignored route are not limited to mirror-reflections but also hold for plane- and depth-rotations. The model can therefore potentially account for a variety of performance patterns after changes in orientation. Third, the priming pattern for split images indicates very strongly that a non-holistic representation (presumably a structural description) is involved in object recognition. Visual priming for attended but not for ignored split versions of an image is a clear indication for a qualitatively different representation of object shape that depends on visual attention. The implications of these findings for key issues in object recognition theories will be discussed in the next section.

4.3 Implications for Issues in Object Recognition

4.3.1 Parallel vs. Serial Processing

Hummel (2001; Hummel and Stankiewicz, 1996a) claimed that, in the hybrid model, analytic and holistic representations work in parallel. However, Stankiewicz et al. (1998) admitted that their results cannot completely exclude the possibility of a serial model. According to a serial model, the representation that gets primed without attention (the holistic representation) resides in an early part of the processing stream, whereas the representation that gets primed only with attention (the analytic representation) resides at a later part of the stream. Serial processing is inherent in Marr's (1982) computational model, in which an early view-based description is extracted from the image and serves as the basis for generating a view-invariant structural description model. Two aspects of the current investigation argue against the hypothesis that holistic and analytic representations work in a serial rather than in a parallel manner.

First, the observed priming for identical ignored images is not due to simple low-level priming which would be predicted by a serial account. Such an account could not explain the failure of rotated objects with a definite base and split objects to prime themselves when ignored. Rather, these results suggest that the observed priming for ignored identical images must be due to a representation that makes contact with a stored holistic representation in LTM. This representation is accessed in parallel and somewhat independent of an attention-consuming analytic representation.

Second, the fact that mirror images, rotated no-base objects, and split images prime their corresponding standard image (unrotated or intact view) equivalently also argues against a stage model. All these manipulations produced an equivalent reduction in priming (about 50 ms). In a serial model, a split image would have to be transformed or matched to an intact (stored) holistic representation, which in turn serves as a basis for extracting a more abstract representation. This extra processing should produce an additional reduction in priming compared to mirror images. In contrast, a parallel model predicts that these manipulations disrupt priming from the holistic component equally. Therefore, only the analytic component visually primes the standard version of an image, and the reduction in priming for mirror-reflected, plane-rotated and split images should be the same.

The evidence for parallel rather than serial processing of a view-dependent and a more abstract representation supports data by Marsolek (1999). He found that exemplars with the same and different shape as the probe objects primed equally well when presented to the left hemisphere. Thus, processing of a particular shape was not a necessary first step, as would have been predicted by a serial model. Consequently, Marsolek assumes two systems working in parallel on qualitatively different representations. Further experiments will have to establish the time course of generating these representation and their exact nature.

4.3.2 View-Dependency and Plane-Rotation

Studies on object recognition often focus on the performance with plane-rotated objects to investigate the question of view invariance. The main difference between the hybrid model and Tarr's (1995) multiple views model is not the existence of view-based representations (which both theories assume) but rather the role of these representations and mental rotation to achieve object invariance. The hybrid model predicts no role of mental rotation, but rather assumes that object invariance is achieved via its analytic route involving structural descriptions. In contrast, the multiple views model proposes that analogue transformation processes are crucial in matching the input image with a limited number of stored views (Tarr & Pinker, 1989, 1990, 1991).

As described in earlier sections (1.4.3.3), structural description theories attribute the performance costs after plane rotation of base objects to be a result of a mismatch between perceived and stored part structure (Hummel & Biederman,1992). According to this prediction, spatial relations in a structural description should be maximally perturbed for plane rotations of about 135°. Indeed, response times increase linearly to about 120° whereas rotations of 180°, however, result in faster naming times than would be expected by a linear normalisation process (Jolicoeur, 1985, 1988, 1990; McMullen & Jolicoeur, 1992). Also, top-bottom discrimination tasks (McMullen & Jolicoeur, 1992) but not handedness discrimination (Shepard & Cooper, 1982) revealed similar patterns of performance costs after plane rotation than naming tasks. This finding seems to further support accounts (e.g., Biederman, 1987) which stress that explicit spatial relations for "top-bottom" but not "left-right" relations are encoded in object representations (McMullen & Jolicoeur, 1992).

Experiments 4 through 6 tested and confirmed the hybrid model's prediction that plane-rotated images prime themselves when attended but not when ignored, and that ignored identical images prime themselves only when presented in a familiar view. This corroborates the hybrid model's prediction that attention plays an important role in object invariance. Plane rotations are compensated for by establishing the part structure of the image and matching it to a similar part structure in memory. View-invariance is achieved via a part-based analytic representation that depends on attention. Thus, our results are in line with studies by Murray (1995c) who found that orientation effects for plane-rotated objects decreased when they had been attended rather than ignored in an earlier block, suggesting that attentional resources are necessary to achieve object invariance.

The hybrid model also fits with the finding that effects of viewpoint differences attenuate in subsequent presentations. For example, plane rotation effects have been found to reduce with practice in naming tasks (Jolicoeur, 1985, 1988; Lawson & Jolicoeur, 1999; McMullen & Jolicoeur, 1992) but not in mirror discrimination tasks (Jolicoeur, 1988). Plane-rotated objects may be recognised more efficiently after repeated naming because of long-term priming of the structural description component or because of facilitated feature extraction. However, since a structural description does not code left/right orientation explicitly (Biederman, 1987; Hummel & Biederman, 1992) mirror discrimination tasks may require that observers refer to a viewer-centred frame of reference (McMullen & Jolicoeur, 1990). Therefore, accounts incorporating structural descriptions would predict the dependence of attenuation effects on the nature of the task, whereas view-based models would not.

There is a further attenuation effect that argues against mental rotation as a mechanism to achieve viewpoint-invariance. Plane rotation effects were reduced after practice, but these practice effects did not transfer to other objects – they are object specific (Jolicoeur, 1985). Furthermore, the size of this attenuation effect is independent of whether the object has been seen in the same or different viewpoint (Murray et al., 1993). For example, if an object is shown in 60° and 120° in practice trials and then tested at 240° and 300° a reduced effect of orientation on performance is observed. The reduction is the same for the test and practice orientations suggesting that observers do not form view-specific representations with repeated exposure, but rather form some orientation-invariant representations.

As discussed in earlier section (1.4.2.2) there is evidence that mental rotation is an unlikely candidate mechanism for recognising plane-rotated views (Jolicoeur et al., 1998). That mental rotation plays a role in object recognition seems also unlikely for a further reason. It has been claimed that plane rotation effects in naming tasks are linear for views rotated between 0° and 120° (Murray, 1997). However, Lawson and Jolicoeur (1999) tested effects of plane rotation with more views than the usually employed 0°, 60°, and 120°. They found that certain views (e.g., 30°, 90°, 150° and 180°) were identified more efficiently than predicted by a linear mental rotation function. The findings suggest that the visual system does not employ an analog mental rotation process to compensate for plane rotation effects (Lawson, 1999). Together, these studies make other view-based mechanisms to achieve object recognition (Ullman, 1998; Poggio & Edelman, 1992) also less plausible.

4.3.3 View-Dependency and Depth-Rotation

As outlined in earlier sections (1.4.2.6) the focus of the debate on the role of view invariance in shape recognition theories has recently shifted to depth-rotations. Experiments 8 and 9 tested the hybrid model concerning its predictions on performance for depth-rotated objects. The priming difference between attended objects in the same and depth-rotated view in Experiments 8 and 9 was not predicted by the angular degree of rotation. Thus, normalisation procedures such as mental transformation seem unable to account for the obtained priming differences (see also Willems & Wagemans, 2001).

The results of Experiment 8 show that qualitative differences such as changes of part structure from study to test have a stronger effect on attended than on ignored images. In this case, the difference in priming between depth-rotated images and identical image pairs is greater for attended than for unattended images. These data corroborate results concerning the effect of part-changes (Hayward, 1998; Experiment 1 and 2). Views in which parts (similar to geons) of novel stimuli changed caused longer response times in sequential object matching tasks compared to equivalent rotations in which the parts remained the same (Figure 26). Interestingly, the difference in matching time between identical viewpoint condition and the depth-rotated condition (with the same parts visible in both views) is about 40 ms, a delay that is expected from the hybrid theory and the present studies (Experiment 8).

The hybrid model can account for the pattern of short-term-priming effects found for depth-rotated familiar objects obtained here (Experiments 8 and 9) and in sequential matching tasks elsewhere (Hayward, 1998; Srinivas, 1995; see section 3.4.5). Potentially, the model can also account for long-term priming effects over view changes found in other studies. In general, over long-term (several minutes) delays the short-lived (several seconds) view-dependent effects of the holistic component are predicted to disappear, and therefore, there should be no priming differences for part-equivalent views. Previous studies with familiar objects had shown that the short-term priming differences found between identical versus mirror-reflected image pairs disappeared when the images were probed after a delay of several minutes (Stankiewicz et al., 1998). Instead, long-term priming for mirror-reflected and identical image pairs was the same (Biederman & Cooper, 1991a). This was predicted by the hybrid model because only the view-independent analytic (structural description) representation should be tapped after such delays, whereas the activation of the holistic surface map declines quickly after a few seconds.

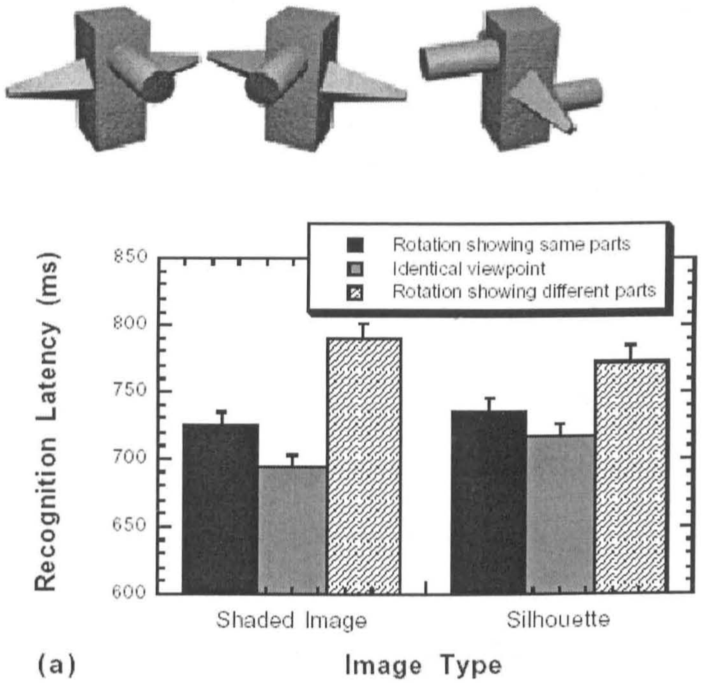


Figure 26: Example of a novel object shown in three viewpoints used in Hayward's (1998) Experiment 1 (top) and performance in a sequential matching task (bottom).

In contrast to the findings on view-invariance with long-term name priming for mirror images, other researchers have found persistent view-dependent effects after depth-rotations of primarily novel objects (Hayward & Tarr, 1997; Tarr, 1995; Tarr & Bulthoff, 1995). These studies were discussed in an earlier section (1.4.2.6) as were the responses from proponents of structural description accounts (Biederman, 2000; Biederman & Bar, 1999, 2000; Biederman & Gerhardstein, 1995). For example, in a critical review of studies reporting high rotation costs, Biederman and Bar (1999) found that distinctive geons in depth-rotated views of objects were often depicted in a low resolution. Biederman and Bar

(1998) showed that increasing the discriminability of distinguishing geons in rendered images for example, by avoiding near accidents or using increased exposure durations, effectively reduced rotation costs. Thus, the long-term priming differences obtained with depth-rotated objects could be an artefact of stimulus (viewpoint) generation. Other artefacts that are potentially responsible for viewpoint dependent effects are the involvement of task-specific non-recognition (episodic) memory systems, non-distinctive structural descriptions among stimuli, uncontrolled occlusion of parts, and time-consuming search for geons on a small scale (e.g., Biederman, 2000). However, there is evidence for view-dependent effects after depth-rotations even for qualitatively similar views (Lawson & Humphreys, 1998).

Lawson and Humphreys (1998) found that the view of a prime object strongly influenced the amount of long-term (several minutes) priming for a target view. Naming of a target view was sensitive to which prime view of that object was presented. Priming effects in these experiments were view-specific. For example, the initial disadvantage naming a foreshortened view in Lawson and Humphrey's (1998) study disappeared with a subsequent presentation of that foreshortened view. As such, this finding clearly contradicts the predictions derived from the hybrid model for long-term priming as well as those of geon theory. These accounts generally predict that after delays of several minutes view-dependent priming effects should disappear. However, in line with properties of structural descriptions observers benefited greatly when naming objects that were previously seen in a different view. There was a clear effect of view generalisation to unseen views. Several points should be noted when considering these results. First, there was no controlled manipulation of part change for specific objects across views. Second, the effects of view on priming were mainly due to foreshortened views. Foreshortened views are specifically disadvantaged in initial recognition tasks (Lawson & Humphreys, 1998, 1999). This is predicted by most object recognition theories because diagnostic features or parts are likely to be occluded in foreshortened views. Similarly, foreshortened views are less likely to be stored in a view-based model because they are unstable and less familiar (Lawson & Humphreys, 1998). Furthermore, unfamiliar material (e.g., unusual views) show greater priming effects (e.g., Srinivas, 1993) which may explain why foreshortened views primed themselves even more than usual views primed themselves. Nevertheless, their results clearly indicate a role for view-specific priming.

Lawson and Humphreys (1998) attributed their results to view-based representations, but it is not quite clear how image-based theories would explain attenuation of view effects from conventional to (presumably unfamiliar) foreshortened views. In principle, the findings can be accommodated within the framework of the hybrid model. For example, the initial presentation of a foreshortened view should be more problematic for both the analytic (part occlusion) and the holistic representation (uncommon view) than a canonical view. However, attending to a foreshortened view allows the encoding of this view into long-term memory. Thus, both the analytic and the holistic component would prime the foreshortened view on subsequent presentation. Other views (foreshortened or not) should mainly be primed by the analytic component. This could explain some of the observed attenuation effects with familiar and foreshortened views in subsequent presentation. It could be tested whether attending to a foreshortened view will encode it as a holistic representation. Previously attended or ignored foreshortened views could be used in the short-term priming paradigm. Previously unattended foreshortened views should not prime themselves when ignored, whereas foreshortened views that were previously attended should prime themselves.

Considering the difficult issue of interpreting depth-rotation effects, it is also noteworthy that not all models of object recognition which incorporate structural descriptions necessarily predict view-invariance after depth-rotation even for long-term priming. In an extension of the structural description representation based on geons, Hummel and Stankiewicz presented a model that codes metric properties in a categorical fashion (Hummel & Stankiewicz, 1998) for example, by coding whether a geon's "side-attached" relation appears closer or further away from the centre (e.g., depending on the orientation of the object in depth). Alternatively, the hybrid model may also account for the viewpoint dependent long-term priming patterns found for depth-rotated objects. Recall that the predictions for priming of ignored and attended images derived from the hybrid model were attributed to independent priming components from layers 5i (IGA) and 5s (HSM) to their correspondent representations in layer 6. This is also the locus of priming in the simulations of the hybrid model (Hummel, 2001; Hummel & Stankiewicz, 1996). However, it is conceivable that priming also manifests itself between layer 6 and 7, that is, between holistic surface map and stored object model. Therefore, an object seen previously in the identical view may receive long-term priming from a stored and previously activated holistic representation, whereas an object seen in a novel (or less familiar) view may only receive priming from the analytic component. This would not affect the general tenet of the hybrid model that object constancy across variations in view and shape is achieved via the analytic part-based representation, but would accommodate observed view-specific long-term priming effects.

In summary, the discussion in this section shows that the hybrid model (its structural description component in particular) does not necessarily predict complete (long-term) view-invariance over depth-rotations. However, this does not render the model unfalsifiable, as it would predict that qualitative changes (such as part changes between depth-rotated views) affect short-term priming more for rotated views of objects that were attended relative to ignored views. These predictions were tested and confirmed in Experiments 7 to 9.

4.3.4 Part-Based Representation

Experiments 1 to 3 established that only attended split images showed visual priming. This is predicted by the hybrid model because split images can only be recognised via a part-based representation that depends on attention. Split images of course do not resemble a part-based or even a geon based representation. Rather, splitting an image was intended to prevent access to a holistic (view-based) representation. The reason is that a view is a vector of spatial coordinates, and it is holistic in the sense that the various features in a view are not represented independently of their locations in the vector (Hummel, 2000). Therefore, split images should only activate an analytic (i.e. part-based) representation. A further piece of evidence for part-based representations comes from Experiments 8 and 9. If views were to be represented in their coordinate spatial relations then the amount of angular separation rather than part changes between depth-rotated views should have affected response performance. However, part changes affected priming more in the attended than in the ignored conditions; this clearly indicates the involvement of part-based representation.

The assumption that the visual system employs a part-based representation still leaves the question whether the representations are based on contours and vertices as proposed in the hybrid model and other structural description models. As described in earlier sections, Biederman and his colleagues provided evidence for his account of geon based representations derived from edge-extractions (Biederman, 1987; Biederman & Cooper, 1991b; Biederman & Ju, 1988). More recently, there is new evidence supporting the special role of convex contours in object recognition. In single-cell recordings with monkeys, Baylis and Driver (2001) used two-dimensional polygons that varied in their curved contour. They found that the shape preferences of IT cells generalised across contrast reversals of these contours and across mirror images of the stimuli, but not across figure – ground reversals. This finding is striking because the three transformations are very different manipulations of the critical curved contour. Contrast reversal changes the polarity of this critical contour, mirror reversal reflects this contour about a vertical axis, and only figure–ground reversal leaves the critical curved contour itself unchanged. The results demonstrate that the selectivity of IT responses is not solely determined by the distinctive contours in a display, contrary to simple edge-based models of shape recognition (e.g., Riesenhuber & Poggio, 2000). Thus, although the ground shares the same contour as the original figure, IT cells seem to generalise more strongly across mirror images than across figure–ground reversal. According to Baylis and Driver, the finding that IT neurons were sensitive to figural shapes defined by one-sided edge assignment, and not by the contours per se, is consistent with shape representation models that stress the role of such component parts (e.g., Biederman, 1987) within IT.

An alternative explanation for the evidence of part-based description obtained in the present experiments is theoretically possible. It is conceivable that the visual recognition system may rely on some sort of a part- or feature-based representation that does not code the spatial relations between them categorically and explicitly (as in a structural description). Recently, Poggio and Riesenhuber proposed a view-based model that claims to be able to generalise over translation, scaling, mirror-reversal, plane and depth-rotations, and even incomplete views of an object (Riesenhuber & Poggio, 1999, 2000). Partial views were instantiated as view-tuned cells. Thus, this model seems in principle capable of recognising split images without assuming a part-based representation that explicitly codes the relative categorical location of parts. Two arguments can be made against such an account. First, there is accumulating evidence for categorical spatial relations in object recognition (Hummel & Stankiewicz, 1996b; Kosslyn et al., 1989; Rosielle & Cooper, 2001; Rosielle, Crabb, & Cooper, 2002). Second, even if part-based recognition can be achieved within a view-based model, as claimed by Riesenhuber and Poggio, it is hard to see how such a system could effectively represent the similarity structure between the two simple shapes in Figure 27. According to Hummel (2000), a visual system relying on structured symbolic representations (such as GSDs) makes it possible to appreciate their obvious similarity while being able to specify how they differ. In contrast, a visual system relying on holistic representations (Poggio & Edelman, 1990) would classify them as completely different shapes. The Poggio and Riesenhuber model could implicitly detect the similarity. For example, view-tuned cells could respond to the "partial" views (which would have to be scale-invariant) corresponding to the square and the circle of the two shapes. However, the representations generated from these two units would be indistinguishable from each other. In contrast, a symbolic representation like the analytic route in the hybrid model that explicitly codes spatial relations between parts can represent the similarity between the two shapes while still keeping them separate. This property of analytic representations enables a visual system to generalise from one object shape to another and form object categories.

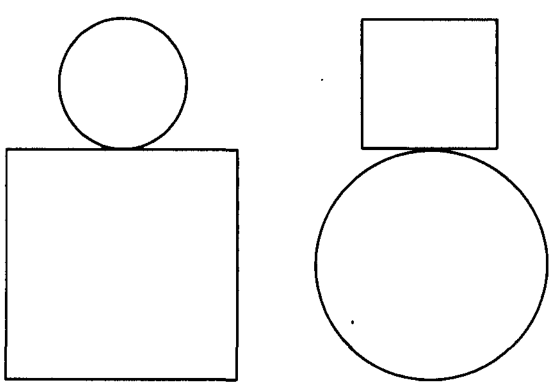


Figure 27: Two simple shapes with a similar structural description (adapted from Hummel, 2000).

4.3.5 Can the Priming Results be Explained by Other Hybrid Models?

In an earlier section, several accounts of object recognition were described that incorporated multiple representations of object shape. Similar to the hybrid model, Farah (1990) proposed that the visual system may employ both holistic and analytic shape representations depending on the nature of the object. The experiments reported here used images of common objects and therefore cannot answer the question whether faces are processed only holistically whereas letters are processed only analytically. The results do show, however, that common objects may be generally processed both holistically and analytically, as suggested by Farah.

Another early hybrid account of object recognition was based on Riddoch and Humphreys' (1984) finding of a double dissociation of processes for achieving object constancy. One group of their patients seemed to rely on global properties in object recognition tasks as they were selectively impaired in matching a foreshortened view to a more canonical view of a target object. At the same time they could match photographs of objects missing an important feature. In contrast, patient HJA had problems with local feature information (the patient showed impaired matching only when the saliency of the target object's primary distinctive feature was reduced) but was unimpaired in matching foreshortened views of a target object. This dissociation suggests a distinction in an independent global and local processing system which is in generally in line with the hybrid model. One problem in fully interpreting these results within the hybrid model framework is that it is not clear in how far foreshortened views can tap the analytic and the holistic representation. However, there are some indications that HJA's presumed impairment in feature processing is more detrimental for object recognition. Although HJA was considerably better in matching foreshortened views than right-hemispheric patients, he was equally impaired in naming those views. HJA's naming performance was in general worse than that of the other patients with more global deficits. This is in line with the assumption that a feature-based representation (which is presumably impaired in HJA) is essential for achieving object constancy.

Jolicoeur's (1990) dual route model proposed a mental rotation system as well as a feature-based representation that work in parallel. However, we found reductions in priming for mirror-images (Experiment 7) as well as for rotated no-base objects (Experiment 5) that usually do not exhibit any mental rotation costs in recognition tasks (Experiment 4). Moreover, the priming pattern for identical versus rotated no-base objects was similar to that of intact versus split images. Since split objects were shown in the same orientation as their intact counterparts, there was no need for a mental rotations system. Without a reason to mentally rotate an object it is not clear why on Jolicoeur's account there is a reduction in priming for split and rotated objects. The same features were visible in the identical relative to the split, mirror-reflected, or plane-rotated (no-base) images. Finally, Experiments 8 and 9 indicated that the amount of angular separation for depth-rotated views does not predict the reduction in priming for attended images. A view difference of 60° (rotation around the z-axis) reduced priming more (over 100 ms, Experiment 8) than a 90° rotation (ca. 40 ms, Experiment 9) in the attended conditions. Therefore, a linear analogue transformation like mental rotation - and in consequence the dual route account - seems unlikely to play a role in the observed priming pattern.

Similar reasons argue against Corballis' "double checking" account. It proposes that recognition costs in studies with rotated objects are due to a mental rotation process after the object has been recognised by a view-invariant representation system. It is conceivable that participants mentally rotated even no-base images in the probe trial to match it with a rotated prime image to verify the identity. However, a different verification process would have to be assumed for split images. Finally, both Corballis' and Jolicoeur's accounts do not predict the observed differences in priming for attended and ignored images.

The current experiments used only one task - basic level naming. It is therefore difficult to assess for the priming studies described here whether task demands played a role in the employment of view-dependent versus view-invariant representations as proposed by Tarr and Bulthoff (1995). However, since the task demands did not change across experiments, their account seems unable to explain the differences in view-dependent priming between Experiment 8 and 9. Nevertheless, it is quite likely that changes in task demands such as subordinate rather than basic-level recognition may produce different priming patterns that indicate a higher degree of view-dependence. But this does not necessarily mean that structural descriptions are not employed. According to the structural description account, it is possible that subordinate-level recognition requires the parsing of geons on a finer scale

(Biederman, Subramaniam, Bar et al., 1999). Therefore, geon theory would also predict that subordinate-level recognition should affect priming differences after view-point changes more than basic-level identification for attended objects, simply because more geons (extracted on a finer scale) and relations mismatch between perceived and stored shape description.

The experiments described here are generally in line with Marsolek's (1999) theory of different visual subsystems in the brain. The abstract-category recognition system, which is associated with the left brain hemisphere has the ability to map different input shapes to the same output representation. The processing of shape information is assumed to be "feature-based" and thought to include non-accidental properties (Lowe, 1985). It is therefore similar to the structural description component of the hybrid model. The second subsystem, associated with "whole-based" shape processing predominantly in the right hemisphere (Marsolek, 1999), is very sensitive to changes in object shape. Marsolek's notion of two hemispherically dissociated subsystems is supported by recent event-related imaging studies (Vuilleumier et al., 2002). The current priming-paradigm was not designed to distinguish between hemispheric processes, and the hybrid model does not assume (nor exclude) preferred hemispheric localisation of the proposed types of representation. Marsolek's subsystem theory does not predict the rapid dissipation of view-specific effects as predicted by Hummel's (2001) model. A further difference between the theories is that the dissociable neural subsystems model does not make any predictions concerning the role of attention in object recognition. The combination of these two lines of research seems a promising direction for further investigations concerning the neural bases of object recognition.

4.4 Implications for Issues in Visual Attention

The hybrid model is not a general theory of attention. For example, it does not address the question of location-based versus feature based visual search, visual neglect, inhibition of return, among others. Rather, the model focuses on the role of attention for representation and basic-level classification of object shape. However, the findings reported here and elsewhere (Stankiewicz & Hummel, 2002; Stankiewicz et al., 1998) have important implications for the question of early versus late selection of visually attended stimuli, as well as for the question of the connection between attention and awareness. Furthermore, the results of the priming paradigm concern the nature of automatic vs. non-automatic processes. These issues will be discussed in the next sections.

4.4.1 Early vs. late Selection

As described in a previous section (see 2.4.4), 'early selection' approaches argued that the treatment received by attended vs. unattended information differs early in perceptual processing. Unattended information was thought to be blocked completely once a fixed bottleneck was reached, with only simple 'physical' properties being extracted prior to that. 'Late selectionists' proposed that the limited awareness of unattended stimuli might not be attributed to less than full perceptual processing, but rather with prohibiting entry of unattended information into memory or into the response control (Duncan, 1980). An initial parallel, unlimited stage of perceptual processing is followed by a second serial, limited-capacity stage concerning selection for awareness, response and memory.

The current investigations was concerned with the formats of shape representation rather than selection processes, but the results for priming of attended and ignored stimuli indicate possible differences in the visual processing stages. On the one hand, the positive priming observed with ignored images suggests that humans can recognise objects without attending to them. This is consistent with the late selection view. On the other hand, the fact that attention plays an important role in object representation is in line with the early selection view. Although attention per se does not determine whether an object will be recognised, it seems that it determines how it will be represented for recognition.

Consistent with a late selection interpretation of the priming effects for ignored images, other researchers such as Tipper (1985) and Treisman and DeSchepper (1996) demonstrated negative priming for ignored stimuli. However, there are three main differences between those studies obtaining recognition for ignored stimuli and the ones reported here. First, the paradigm used here produced positive priming rather than negative priming. Second, negative priming has been observed previously for semantically related objects (Tipper & Driver, 1988), whereas in Experiment 2 (similar to Stankiewicz et al., 1998) there was no priming for ignored objects with the same name but different shape. Finally, some researchers have found priming from unattended novel shapes (DeSchepper & Treisman, 1996), whereas Experiments 3 and 6 demonstrated that priming for ignored images depends on previously stored representations.

Concerning the finding of positive instead of negative priming, the main difference between the two priming paradigms is that, in the former, attention was directed to the target object in the prime display by a spatial cue (a square box) whereas, in the latter, participants had to select one object of two overlapping stimuli defined by a particular colour in both the prime and probe trial. According to Stankiewicz and Hummel (2002), the participants in a negative priming paradigm had to "actively ignore" (or select against) the unattended stimulus. Therefore, processing of the ignored image may proceed up to a semantic and even response level in these experiments which could explain the negative (instead of positive) priming for both identical and semantically related objects. The activated response (naming of the ignored image) had to be suppressed. In contrast, in the positive priming paradigm spatially separated uncued images do not have to be inhibited when observers select the cued stimulus and therefore may not be processed at a semantic level. Indeed, ignored images can facilitate naming if their holistic shape corresponds with the probe. Support for this interpretation of priming based on segmentation processes comes from studies showing that distracter interference could be reduced if the distracters were made more physically distinct from targets (Francolini & Egeth, 1980), or placed further away (Eriksen & Eriksen, 1974). In fact, Tipper and Cranston (1985) found no negative priming when the probe was displayed alone (not overlapping with another stimuli; see also Stankiewicz et al.'s, 1998, Experiment 3) and concluded that inhibition may have rapidly decayed.

Concerning the question of semantic priming from pictures, in contrast to the current paradigm Tipper and Driver (1988) used a picture categorisation task presumably to maximise semantic processing of the stimuli. They found negative priming from ignored pictures onto semantically related words, indicating a central semantic locus of the negative priming effect. However, their research was criticised because their task afforded an identical response in the semantically related condition, which had to be suppressed in the prime trial (Fox, 1995). Thus, "response repetition" rather than attentional inhibition of semantic representations could account for slower response times in the semantically related condition. Similar conclusions were drawn by Damian (2000) who sought to establish whether semantic representations were involved in negative priming by eliminating potential confounds. He concluded that the semantic negative priming found earlier (Tipper & Driver, 1988) was due to an artefact of the categorisation task and that the crucial test of the locus of this effect was the failure to obtain semantic negative priming in a naming task. In general, the finding that negative priming with a naming task does not necessarily extend to categorically related stimuli is in line with the results obtained with positive priming in Experiments 1 and 2. Experiment 2 showed no positive priming for ignored shapes that shared the same name with the test exemplar in the probe trial. An interesting question for further research is whether structurally similar exemplars prime each other more than structurally dissimilar exemplars, and if so whether such differences are the same in attended and ignored conditions.

Experiments 3 and 6 established that unattended images in unfamiliar configurations or orientation do not prime themselves. The latter findings support the hybrid model's prediction that only already encoded representations can be primed in the ignored route. In apparent contradiction to this conclusion, DeSchepper and Treisman found negative priming for novel shapes (DeSchepper & Treisman, 1996). However, their paradigm and task varied considerably from the short term paradigm in this investigation. Most notably, prime and probe displays remained visible until observers responded, potentially giving them the opportunity to allocate attention to the to-be-ignored shape. In addition, the novel (overlapping) shapes were used in more than one trial-pair. Thus, the results of this study do not seem to be directly comparable to the findings reported here.

The current debate about the nature of negative priming (see Tipper, 2001; and Fox, 1995, for reviews) is beyond the scope of this thesis; rather, the aim was to investigate the format of representations used in object recognition with and without attention rather than the selection process. However, the discussion above shows that negative and positive priming effects are non-exclusive and strongly depend on the paradigm. It has also been shown that there is growing evidence for visual processing of ignored objects at least to the level of shape representation, which must be taken into account by theories of visual selection.

In general, our results are in line with theories that predict a role of attention for feature binding because only attended split and plane-rotated objects were primed. For example, the

Feature Integration Theory of Treisman and Gelade (1980) and the Guided Search model of Wolfe, Cave and Franzel (1989) have two processing stages, the first parallel (preattentive) and the second serial (requiring focused attention). In the first stage simple features are computed in parallel across the visual scene. In the second stage, attention is focused on the location in the visual field containing the target defined by a conjunction of attributes and a (temporary object) representation is made. However, although both models account for a wide variety of results from the literature on visual search they provide no straightforward explanation for the findings of Shiffrin and Schneider (1977). They showed that complex stimulus characteristics such as shapes of whole letters seemed to be matched against long-term memory representations before attentional selection occurs. Similar to these results that imply late selection, the priming experiments presented here show that even complex stimuli such as whole shapes of common objects can be processed without attention.

How can the present findings be integrated in late vs. early selection theories? The predictions of the hybrid model and the results of the experiments targeted to test them are broadly reminiscent of Treisman's (1960) attenuation theory. It states that ignored stimuli can sometimes be processed passed basic physical stages and "breakthrough" to awareness (e.g., words presented consistent with expectations). Applied to the present findings, attenuation or the flexibility of this "filter" for ignored information seems to depend on the familiarity with the ignored stimulus, as only holistic representations in familiar views seem to prime in the ignored conditions. Further factors may influence selection which resemble explanations in terms of "capacity limits". Lavie (1995; Lavie & Fox, 2000) recently proposed an account targeted to explain findings for both late and early selection views. In an extensive review, Lavie and Tsal (Lavie & Tsal, 1994) observed that results supporting the late selection view had often been obtained in situations of low 'perceptual load' (e.g., with just a single target and single distracter, or an undemanding task for the target). But there was also evidence for early selection when perceptual load was higher (e.g., more stimuli presented, and/or a more demanding task for target detection). Adapting the perceptual load account to the short-term priming paradigm, the prediction of the hybrid model would be that increasing the number of different ignored images should reduce the amount of visual priming for each of them.

In summary, the priming results of the experiments reported here are generally in line with other studies demonstrating priming for unattended information. Our data also seem to confirm and extend previous views that interpret early vs. late selection phenomena as the result of a complex interaction between capacity limits and stimulus properties.

4.4.2 Attention, Binding and Awareness

The results of the experiments reported here showed that manipulations of the image structure affected attended and ignored objects differently. Priming for attended images is more sensitive to part changes than priming for ignored images (Experiment 8), whereas manipulations that affect the holistic but not the local (part) structure of an image disrupt priming for unattended objects (Experiments 1 to 3 and 5 to 6). These results support the notion that attention is needed to bind object attributes (Treisman, 1998).

Evidence for the role of attention in binding of object properties was discussed in an earlier section (2.4.4). A further line of evidence that the binding of visual aspects of an object requires attention comes from phenomena known as the 'attentional blink'. In this paradigm, participants must attend to and report two targets in a rapidly presented stream of items. Identification of Target 2 is severely impaired if it is presented 100—300 ms (or more) after Target 1 (Chun & Potter, 1995; Shapiro, 1994). Attentional blink is reduced by increasing target-distractor discriminability and is thought to reflect a general limit of the speed with which succesive stimuli can be processed. According to Treisman and Kanwisher (1998), the results of this paradigm can be interpreted as a failure to establish a separate object representation as a consequence of attentional capacity limitations. Interestingly, Shapiro, Driver, Ward, and Sorenson (1997) found priming produced by "blinked" stimuli, which is also consistent with the present findings that ignored stimuli were processed (primed in same view) but were not available for report (Experiment 7). These results conform with the hybrid model's assumption that conscious recognition of objects requires attention, which is necessary to bind object attributes together.

Another demonstration of the special role of attention for binding and awareness in object recognition comes from the work of Rensink (2000; 2002) on change blindness. In a number of paradigms, he observed the striking inability to identify a change for example, in rapidly presented scenes. At least two findings suggest that change blindness depends on attention. First, it seems that a change will be noticed if it produces a visual transient that attracts attention (Jonides & Yantis, 1988; Yantis, 1993; Yantis & Johnson, 1990). Second, a change will be noticed if it occurs at the current locus of attention (Rensink, 2002; Rensink, Oregan, & Clark, 1997). It appears that the change is found when attention rests on the correct object while that object changes. Recently, Rosielle, Crabb, and Cooper (2002) found that participants were faster at detecting positional changes in alternating versions of a scene if categorical relations rather than only metric relationship between objects in a scene changed. This finding is supported by the results of this thesis indicating that a representation based on the categorical coding of parts (Experiment 8) rather than on metric properties has preferred access to awareness.

The crucial assumption of the hybrid model is that attention binds object attributes dynamically. Whether dynamic binding for object attributes is established by temporal synchrony (Hummel, 2001; Hummel & Biederman, 1992) or some other means is obviously beyond the scope of this thesis (von der Malsburg, 1995). There is evidence that synchrony of firing may be the (or one) basis for dynamic binding (Castelo-Branco, Goebel, Neuenschwander, & Singer, 2000; Gray, Engel, Konig, & Singer, 1992; Gray, Konig, Engel, & Singer, 1989; Gray & Singer, 1989). However, Hummel (1997) pointed out that - from a computational standpoint - the question of how dynamic binding is represented is not necessarily important for structural description models. Therefore, in summary, the results presented here simply support the notion that attention is needed for the dynamic binding of object attributes and for the conscious (reportable) recognition of an object.

A further issue on the matter of awareness and attention arising from the experiments using the priming paradigm concerns the issue of whether unattended images are completely ignored. If attention is necessary for binding of shape attributes, but at the same time unattended objects show priming, it could be argued that the reduced priming in the ignored conditions was due to residual attentional processing. It is conceivable that the ignored objects in the short-term priming paradigm were not completely ignored and therefore some attention was allocated to the uncued objects. If this was the case the model's prediction and its special emphasis on the role of attention would be less convincing. This criticism of the paradigm seems very implausible. First, observers in Experiment 7 were unable to report the correct identity of the ignored image in the last trial, similar to Rock and Guttman's (1981) observers who were unable to report the formerly unattended object in a superimposed object pair. However, it may be argued that observers simply forgot the ignored object quickly, or that some attentional resources were directed towards the ignored stimulus but not enough to establish a conscious representation. But even if participants allocated some attention to the ignored stimulus, one would expect at least some priming for example, for ignored mirror images or rotated no-base objects. Here, no evidence for priming for was found for ignored objects (unless there was a metric overlap as in Experiment 8). Moreover, previous research has shown how important location is for allocation of attention (see section 2.4.4). For example, in paradigms using inattentional blindness, observers who perform a single task (e.g., a line length judgement) cannot report a stimulus flashed near to the task-relevant target when queried after the trial (Mack, Tang, Tuma, & Kahn, 1992; Moore & Egeth, 1997). Finally, as the observers in the present experiments certainly did not perform at ceiling, correct prime-probe trials with the short duration of image presentation hardly allowed enough time to pay attention to both cued and uncued locations.

There are further reasons to assume that unattended objects are really ignored while at the same time some shape properties can be processed. This evidence comes from patients with visual neglect. Visual neglect affects visual attention and awareness. Although their visual cortex and its initial afferent inputs are intact, patients with neglect after right-parietal injury have deficient awareness for visual stimuli at the contralesional side of space. However, if their attention is drawn to that side, the visual stimuli are brought into awareness (Driver & Mattingley, 1998). At the same time, recent evidence suggests that attributes of neglected stimuli such as colour and shape still get encoded by the neglect patient's visual system, despite the loss of awareness (Cohen, Ivry, Rafal, & Kohn, 1995; Mattingley, Bradshaw, & Bradshaw, 1995). Thus, we conclude that the evidence is against the hypothesis that the priming for ignored uncued images in this paradigm was due to (residual) attentional processing.

4.4.3 Automatic Processing vs. Controlled Processing

The fact that ignored images can prime a representation of an objects' shape (and identity) is in line with the hybrid model's prediction that object shape can be processed automatically, that is without attention (Hummel & Stankiewicz, 1996b; Stankiewicz & Hummel, 2002; Stankiewicz et al., 1998). The results of the current experiments (in particular Experiments 3 and 6) suggest that automatic processing of ignored images is view-sensitive and only occurs with familiar views. In contrast, attention is necessary to process the analytic properties of an object (such as its parts and their spatial relations to compensate for view changes). The additive effects of view (or configuration) and attention found in Experiments 1, 5, 7, and 9 were attributed to the automatic holistic representation contributing to the priming from the analytic component.

According to JIM.3 and its predecessors (Hummel & Biederman, 1992), processing of a geon structural description proceeds in a bottom-up fashion and is therefore expected to be initiated automatically as soon as attention is allocated to an object. This issue is connected with the concept of attentional control. Posner (1980) distinguished between exogenous and endogenous attention control, whereas Jonides (1981) proposed a dichotomy between automatic and nonautomatic attention control. Both the concepts of exogenous and automatic attention stress factors outside the organism in attentional processing. The response to a stimulus is determined by stimulus characteristics. In consequence, exogenous control has often been associated with stimulus-driven or bottom-up control. In contrast, the endogenous/nonautomatic mode of attention control is associated with the initiation of overt voluntary action for example, to direct one's attention to a specific location or attribute.

Although the question of automatic vs. non-automatic attentional control (not to be confused with Hummel's terminology of automatic processing of holistic representations) was not directly addressed in this investigation, it is noteworthy that paying attention to a cued location may automatically activate both the holistic and the analytic representation (as predicted by the hybrid theory). There is evidence that automatic processing may be involved for attended objects up to a semantic level. Boucart and Humphreys (1992) showed that observers cannot selectively process global shape information without accessing semantic or name information. Humphreys and Boucart (1997) have shown that when processing of local form is required for response, surrounding global information is automatically processed to a semantic level. Whether the priming obtained with these tasks is due to the activation of both an analytic and holistic representation or only due to the latter cannot be decided on the basis of their results. Further experiments will have to establish whether attending to an object without naming it automatically triggers a part-based representation. This should shed more light on the extent of bottom-up processing in object recognition.

4.5 Multiple Representations in the Brain?

This research presented behavioral evidence in support of two qualitatively different representations of shape. However, as many models of object recognition are increasingly driven by neuropsychological and neurophysiological observations, the hybrid model would be more convincing if it could account for these findings as well. The introductory chapters described evidence for multiple (presumably view-based and part-based) representations in the brain (2.4.3). This section discusses this evidence in relation to the findings presented here.

On a general level, the findings that object recognition is mediated by two qualitatively different systems supports neurophysiological and imaging studies showing that neuronal areas in the ventral pathway differ in their response to changes in view conditions (see section 2.4.3). In line with the hybrid model, Experiments 5 to 9 indicate that an analytical representation generalises across differences in orientation of objects, whereas a holistic representation is highly sensitive to changes in picture plane and depth-orientation. This fits with recent evidence that some neurons in IT code for complete objects whereas others are selective for individual views (Booth & Rolls, 1998). Janssen, Vogels and Orban (2000) showed that different areas in the macaque IT respond selectively for 3D and 2D shape. Neurons in the superior temporal sulcus were selective for three-dimensional shape whereas neurons in lateral TE were generally unselective for 3D shape, though equally selective for 2D shape. These findings strongly suggest that IT (or macaque TE) consists of at least two distinct areas with different sensitivity to shape properties.

The results from the fMRI study of Vuilleumier et al. (2002) are even more directly relevant for the studies presented here. They showed in a priming paradigm that repetition of images of common objects decreased activity (i.e. primed) in the left fusiform area independent of viewpoint (and size), whereas a viewpoint-dependent decrease in activation was found in the right fusiform cortex. Interestingly, the latter area was sensitive to changes in orientation but not in size, properties directly predicted from the holistic representation (Hummel, 2001). Moreover, Henson, Shallice and Dolan (2000) found that repetition priming resulted in different patterns of attenuation in the right fusiform area depending on whether the repeated stimuli were familiar or unfamiliar, supporting the current results of Experiments 3 and 6 that only stored holistic views prime themselves.

The priming results obtained with split images (Experiments 1 to 3) also fit with observations on patient RK (Davidoff & Warrington, 1999) who could not recognise parts of objects although he could recognise complete objects in a whole conventional view. In particular, RK was markedly impaired in recognising “exploded” versions of a Snodgrass and Vanderwart (1980) image. He was also impaired in detecting part changes when presented with different alternatives of objects (e.g., a donkey with original ears or ears from another animal). These data can be interpreted within the hybrid model to indicate that RK's analytic route is impaired, as he was poor with unconventional views, exploded views, and detecting part changes. At the same time, his holistic representation may be intact allowing automatic recognition of conventional views. Furthermore, although he was unable to discriminate between mirror images and rotated versions of no-base objects, RK was good at discriminating between upright base objects and their plane-rotated versions. This, again, is in line with the data presented here (Experiments 4 to 7), indicating that there is a qualitative difference in holistic properties for these manipulations in viewpoint. The question arises why RK could recognise but not discriminate between familiar views of no-base objects. According to the hybrid model, the automatic holistic route does not allow conscious access to shape representation, but provides a fast direct route to object identity units. It is possible that RK was incapable to discriminate between mirror-images and rotational changes with no-base objects because familiar orientations access the same semantic unit for a given object. Interestingly, RK was able to distinguish between “right shoe” and “left shoe”, indicating that different shape representations access different semantic units in cases where orientation matters for identification. This interpretation is supported by another patient who was impaired at recognising objects from unconventional views (Warrington & Davidoff, 2000). JBA was unable to perform mirror discriminations with line drawings of common objects, but was able to discriminate mirror reflections of novel meaningless objects. Thus, it is possible that - because there was no automatic recognition of novel meaningless objects - stages of visual processing were engaged that allowed the discrimination of mirror-reflected shapes.

In conclusion, the results obtained in Experiments 1 to 9 seem to complement neuropsychological and neurophysiological findings that indicate formats of representation similar to the hybrid model. Further research may seek to explore the neural basis of these representations in dependence of attention.

4.6 Evaluating the Hybrid Model

In general, the results of the experiments presented support the hybrid model and its prediction of qualitatively different representations for ignored and attended images. Furthermore, the discussion above shows that the hybrid model seems to be generally in line with findings from other behavioral and even neurophysiological studies. At a theoretical level, the question arises about the need for the visual system to employ a view-dependent holistic representation in addition to a structural description component. The answer could come from the multiple roles served by object representations. Indeed, the hybrid model of Hummel (2001) was motivated by the observation that object recognition can be fast (Intraub, 1981) and operate without attention (Tipper, 1985). On the hybrid model, rapid object recognition would be achieved by holistic representations of stored views but these representations would be insufficient for all tasks requiring object recognition. In particular, they would be inadequate to decide that objects belong to a category when they do not look closely similar to each other. For example, a Land Rover is readily classified as a car, as is a Formula One Ferrari, although the 2D projections of the two object images are very different. A holistic representation strictly based on the laws of projective geometry (converting retinal images into stored object-centred representations) would not tolerate such variations in an object's shape (for discussions, see Edelman & Intrator, 2000; Hummel, 2000, 2001). In contrast, a structural description system that describes objects in terms of abstract generic parts and their interrelations would solve the problem rather easily. The prediction for the hybrid model would be that only attended images generalise (show priming) over changes in an object's shape. In Experiment 2, there was priming for objects with the same name but were a different exemplar only in the attended condition. However, since the obtained priming for a different exemplar could be entirely due to repeated naming, it is not certain that priming in this condition was due to the activation of a different instance of that object class. Evidence that attending to an object may indeed activate the representation of other exemplars comes from Marsolek (1999) who found that an image of an object receives more priming from a different exemplar of that object (with the same name) than from the visually presented name of that object.

The present experiments with rotated objects confirmed the model's general predictions for priming of attended and ignored views. This is important because the hybrid model is able to account for an interesting finding in the object recognition literature. In simulations with plane-rotated objects, JIM.3 - like human observers and like JIM (see Hummel & Biederman, 1992) - is faster and more accurate in recognising images that are completely upside-down (i.e., 180° off upright) than images that are slightly less than perfectly upside-down (e.g., 135° off upright). Thus, the model appears to be able to account for a large number of findings describing the relationship between recognition performance and angular degree of plane rotation. As discussed above, the model seems also to be able to account for findings with depth-rotated objects, at least in short-term priming paradigms or matching tasks. Future research can further test the models prediction with more finely tuned view differences.

Not surprisingly, JIM.3 still has some shortcomings and problems such as the coding of axes (which cannot be extracted from line drawings). Other issues that have to be addressed in future research have been discussed in earlier sections. They include the role of hemispheric differences, the time course of holistic and analytic priming, the degree of view-sensitivity of the holistic representation, and the role of attention in encoding new views. A particularly important limitation of JIM.3 - and other computational models of object recognition - is its inability to solve the figure-ground segmentation problem. This means that JIM.3 can "view" only one object at a time. But what happens if more than one (ignored) object is in a scene - will they all be recognised in their familiar views? According to Hummel (2001), the assumption is that features of ignored objects may fire occasionally, but only in synchrony with one another. Computationally speaking, all ignored objects are forced to share a single "time slice" in the oscillatory firing. The more ignored objects are present in the visual field, the smaller the probability that any given one of them fires by itself within any fixed amount of time. Moreover, with multiple ignored objects in the visual field, the probability that binding errors occur increases, not only for dynamic binding in the IGA, but also in the surface map, if two or more objects happen to fire at the same time. Thus, JIM.3 would predict that the relationship between attention and patterns of priming should vary as a function of how many ignored objects there are in the visual field. This can be tested by increasing the number of ignored images in the priming paradigm. In fact, there is evidence that negative priming is attenuated when the number of distracters in a prime display increases (Neumann & Deschepper, 1992).

A natural limitation in interpreting the present (and many other) experiments of object recognition is that observers were limited in their viewing conditions: they had to gaze at a particular point of space and saw only isolated objects. It may be tempting to question the ecological validity of such experiments when trying to extend their findings to real world conditions. Natural vision differs from most controlled viewing studies. Stimuli in the real world tend to be complex and stimulus-directed eye movements occur often during natural viewing. In single-cell studies with monkeys DiCarlo and Maunsell (2000) examined the effect of natural viewing conditions while controlling scene complexity. For nearly 90% of IT neurons, the reponses were equivalent to those under restricted viewing conditions. These results seem to indicate that activity in IT in response to object identity can be independent of how that image was brought to the retina.

A further general criticism of the hybrid model could address the notion of multiple views in the holistic route (Experiment 5). This component of the model might appear to have become too powerful and implicitly resemble a multiple views approach. However, the model clearly predicts that view-constancy is achieved via the attention consuming analytic representation, a prediction that seems increasingly corroborated (e.g., Murray, 1995c). The hybrid model does not have to assume an alignment or transformation process in order to generalise across viewpoint changes in the holistic route. Rather, the hybrid model assumes that although holistic 2D representations of views can be stored in memory, these representations are not used to achieve object constancy.

A general advantage of the hybrid model over many other models of object recognition is that it makes relative clear predictions about priming patterns for attended and ignored images. Thus, priming in the ignored and analytic route can help to tap different processes and representation formats, and therefore make more powerful predictions than models that propose general view-specificity in dependence of view familiarity (e.g., Tarr & Pinker, 1989) as well as models that solely rely on structural descriptions (Biederman, 1987).

4.7 Conclusion

In conclusion, the experiments reported here demonstrate that the visual representation generated in response to an attended image is qualitatively different from that generated in response to an ignored image. Although recognition takes place in both cases, it is mediated by one kind of representation when the image is attended and a different kind when the image is ignored. The findings presented here suggest a very specific answer to the question "How can the visual representation of shape have some properties that demand explanation in terms of analytic representations, and simultaneously have other properties that are strictly inconsistent with analytic representations?" It appears that the visual system represents shape analytically when it can, but represents it holistically when it must.

5. References

Abu-Mostafa, Y. S., & Psaltis, D. (1987). Optical neural computing. *Scientific American, 256*, 66–73.

Adini, Y., Moses, Y., & Ullman, S. (1997). *Face recognition: the problem of compensating for illumination changes.* Paper presented at the IEEE Transactions on Pattern Analysis and Machine Intelligence.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230*(4724), 456-458.

Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia, 35*, 1121-1131.

Bartram, D. J. (1976). Levels of coding in picture-picture comparison tasks. *Memory & Cognition, 4*, 593-602.

Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition - the effect of grouping factors. *Perception & Psychophysics, 51*(2), 145-162.

Baylis, G. C., & Driver, J. (2001). Shape-coding in IT cells generalizes over contrast and mirror reversal but not figure-ground reversal. *Nature Neuroscience, 4*(9), 937-942.

Bergevin, R., & Levine, M. D. (1993). Generic object recognition - building and matching coarse descriptions from line drawings. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 15*(1), 19-36.

Biederman, I. (1987). Recognition-by-components - a theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Biederman, I. (1995). Visual object recognition. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual Cognition* (2 ed., Vol. 2, pp. 121-165). Cambridge, MA: MIT Press.

Biederman, I. (2000). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision, 13*(2-3), 241-253.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39*(17), 2885-2899.

Biederman, I., & Bar, M. (2000). Differing views on views: response to Hayward and Tarr (2000). *Vision Research, 40*(28), 3901-3905.

Biederman, I., & Cooper, E. E. (1991a). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20*(5), 585-593.

Biederman, I., & Cooper, E. E. (1991b). Priming contour-deleted images - evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23*(3), 393-419.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology-Human Perception and Performance, 18*(1), 121-133.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects - evidence and conditions for 3-dimensional viewpoint invariance. *Journal of Experimental Psychology-Human Perception and Performance, 19*(6), 1162-1182.

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition - reply to Tarr and Bulthoff (1995). *Journal of Experimental Psychology-Human Perception and Performance, 21*(6), 1506-1514.

Biederman, I., Gerhardstein, P. C., Cooper, E. E., & Nelson, C. A. (1997). High level object recognition without an anterior inferior temporal lobe. *Neuropsychologia, 35*(3), 271-287.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*(1), 38-64.

Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., & Fiser, J. (1999). Subordinate-level object classification reexamined. *Psychological Research-Psychologische Forschung, 62*(2-3), 131-153.

Biederman, I., Subramaniam, S., Kalocsai, P., & Bar, M. (1999). Viewpoint-invariant information in subordinate-level object classification. *Attention and Performance Xvii, 17*, 91-111.

Binford, T. O. (1971). *Visual perception by computer.* Unpublished manuscript, Miami, FL.

Blanz, V., Tarr, M. J., & Bulthoff, H. H. (1999). What object attributes determine canonical views? *Perception, 28*(5), 575-599.

Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex, 8*(6), 510-523.

Boucart, M., & Humphreys, G. W. (1992). Global shape cannot be attended without object identification. *Journal of Experimental Psychology-Human Perception and Performance, 18*(3), 785-806.

Boucart, M., & Humphreys, G. W. (1997). Integration of physical and semantic information in object processing. *Perception, 26*(9), 1197-1209.

Boutsen, L., Lamberts, K., & Verfaillie, K. (1998). Recognition times of different views of 56 depth-rotated objects: A note concerning Verfaillie and Boutsen (1995). *Perception & Psychophysics, 60*(5), 900-907.

Boutsen, L., & Marendaz, C. (2001). Detection of shape orientation depends on salient axes of symmetry and elongation: Evidence from visual search. *Perception & Psychophysics, 63*(3), 404-422.

Bruce, V., Carson, D., Burton, M. A., & Ellis, A. W. (2000). Perceptual priming is not a necessary consequence of semantic classification of pictures. *The Quarterly Journal of Experimental Psychogy, 53A*(2), 289-323.

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89*(1), 60-64.

Bulthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are 3-dimensional objects represented in the brain. *Cerebral Cortex, 5*(3), 247-260.

Burgund, E. D., & Marsolek, C. J. (2000). Viewpoint-invariant and viewpoint-dependent object recognition in dissociable neural subsystems. *Psychonomic Bulletin & Review, 7*(3), 480-489.

Cabeza, R., & Nyberg, L. (2000). Imaging Cognition II: An Empirical Review of 275 PET and fMRI Studies. *Journal of Cognitive Neuroscience, 12*(1), 1-47.

Castelo-Branco, M., Goebel, R., Neuenschwander, S., & Singer, W. (2000). Neural synchrony correlates with surface segregation rules. *Nature, 405*(6787), 685-689.

Cave, C. B., & Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. *Perception, 22*(2), 229-248.

Cave, K. R., & Bichot, N. P. (1999). Visuo-spatia l attention: Beyond a spotlight model. *Psychonomic Bulletin and Review, 6*, 204–223.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in RSVP. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 109-127.

Cohen, A., Ivry, R. B., Rafal, R. D., & Kohn, C. (1995). Activating response codes by stimuli in the neglected visual field. *Neuropsychology, 9*, 165–173.

Cooper, E. E., & Biederman, I. (1993). Metric versus viewpoint invariant shape differences in visual object recognition. *Investigative Ophthalmology & Visual Science, 34*(4), 1080-1080.

Cooper, E. E., Biederman, I., & Hummel, J. E. (1992). Metric invariance in object recognition - a review and further evidence. *Canadian Journal of Psychology-Revue Canadienne De Psychologie, 46*(2), 191-214.

Cooper, L. A., Schacter, D., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(1), 43-57.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review, 95*(1), 115-123.

Corballis, M. C., Zbrodoff, N. J., Shetzer, L. I., & Butler, P. B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory & Cognition, 6*, 98-107.

Corbetta, M., Shulman, G. L., Miezin, F. M., & Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science, 270*(5237), 802-805.

Damian, M. F. (2000). Semantic negative priming in picture categorization and naming. *Cognition, 76*, B45-B55.

Davidoff, J., & Warrington, E. K. (1999). The bare bones of object recognition: Implications from a case of object recognition impairment. *Neuropsychologia, 37*(3), 279-292.

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(1), 27-47.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3*(1), 1-8.

DiCarlo, J. J., & Maunsell, J. R. H. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature Neuroscience, 3*(8), 814-821.

Dickerson, J., & Humphreys, G. W. (1999). On the identification of misoriented objects: Effects of task and level of stimulus description. *European Journal of Cognitive Psychology, 11*(2), 145-166.

Dickinson, S. J., Bergevin, R., Biederman, I., Eklundh, J. O., Munck-Fairwood, R., Jain, A. K., & Pentland, A. (1997). Panel report: The potential of geons for generic 3-D object recognition. *Image and Vision Computing, 15*(4), 277-292.

Dickinson, S. J., Pentland, A. P., & Rosenfeld, A. (1992). From volumes to views - an approach to 3-D object recognition. *Cvgip-Image Understanding, 55*(2), 130-154.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention - the spotlight metaphor breaks down. *Journal of Experimental Psychology-Human Perception and Performance, 15*(3), 448-456.

Driver, J., & Baylis, G. C. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The attentive brain* (pp. 299–326). Cambridge, MA: MIT Press.

Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience, 1*(1), 17-22.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*, 272–300.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501–517.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences, 1*(8), 296-304.

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences, 21*(4), 449-498.

Edelman, S., & Intrator, N. (2000). (Coarse coding of shape fragments) plus (retinotopy) approximate to representation of structure. *Spatial Vision, 13*(2-3), 255-264.

Edelman, S., & Weinshall, D. (1998). Computational approaches to shape constancy. In V. Walsh & J. Kulikowsky (Eds.), *Perceptual constancy: Why things look as they do* (pp. 124-143). New York, NY: Cambridge University Press.

Ellis, R., & Allport, D. A. (1986). Multiple levels of representation for visual objects: A behavioral study. In A. G. Cohn & J. R. Thomas (Eds.), *Artificial intelligence and its applications* (pp. 245-247). Chichester, England: Wiley.

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of object constancy. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 41A*, 775-796.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non search task. *Perception & Psychophysics, 16*, 143-149.

Eriksen, C. W., & Hoffman, J. E. (1974). Selective attention: Noise suppression or signal enhancement? *Bulletin of the Psychonomic Society, 4*, 587-589.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention - a zoom lens model. *Perception & Psychophysics, 40*(4), 225-240.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press/Bradford Books.

Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. *Cognitive Neuropsychology, 8*, 1-19.

Farah, M. J., & Aguirre, G. K. (1999). Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition. *Trends in Cognitive Sciences, 3*(5), 179-186.

Farah, M. J., & Hammond, K. M. (1988). Mental rotation and orientation invariant object recognition: dissociable processes. *Cognition, 29*(29-46).

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482-498.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*, 1–47.

Fiser, J., & Biederman, I. (1995). Size invariance in visual object priming of gray-scale images. *Perception, 24*(7), 741-748.

Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Research, 41*(2), 221-234.

Fox, E. (1995). Pre-cueing target location reduces interference but not negative priming from visual distracters. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 48*(1), 26-40.

Francolini, C. M., & Egeth, H. E. (1980). On the non-automaticity of 'automatic' activation: Evidence of selecting seeing. *Perception & Psychophysics, 27*, 331–342.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.

Gerlach, C., Aaside, C. T., Humphreys, G. W., Gadeb, A., Paulson, O. B., & Lawa, I. (2002). Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. *Neuropsychologia, 40*, 1254–1267.

Goodale, M. A. (1995). The cortical organization of visual perception. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual Cognition* (2 ed., Vol. 2, pp. 167-213). Cambridge, MA: MIT Press.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15*, 20-25.

Gray, C. M., Engel, A. K., Konig, P., & Singer, W. (1992). Synchronization of oscillatory neuronal responses in cat striate cortex - temporal properties. *Visual Neuroscience, 8*(4), 337-347.

Gray, C. M., Konig, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual-cortex exhibit inter- columnar synchronization which reflects global stimulus properties. *Nature, 338*(6213), 334-337.

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual-cortex. *Proceedings of the National Academy of Sciences of the United States of America, 86*(5), 1698-1702.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*(1), 187-203.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping, 6*(4), 316-328.

Hamm, J. P., & McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology-Human Perception and Performance, 24*(2), 413-426.

Hayward, W. G. (1998). Effects of outline shape in object recognition. *Journal of Experimental Psychology-Human Perception and Performance, 24*(2), 427-440.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology-Human Perception and Performance, 23*(5), 1511-1521.

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science, 11*(1), 7-12.

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science, 287*(5456), 1269-1272.

Hinton, G. E., & Parsons, L. M. (1981). Frames of reference and mental imagery. In J. Long & A. Baddeley (Eds.), *Attention and Performance IX.* Hillsdale, NJ: Erlbaum.

Holmes, E. J., & Gross, C. G. (1984a). Effects of inferior temporal lesions on discrimination of stimuli differing in orientation. *Journal of Neuroscience, 4*, 3063-3068.

Holmes, E. J., & Gross, C. G. (1984b). Stimulus equivalence after inferior temporal lesions in monkeys. *Behavioral Neuroscience, 98*, 898-901.

Hopfield, J. J. (1982). *Neural networks and physical systems with emergent collective computational abilities.* Paper presented at the Proceedings of the National Academy of Sciences of the United States of America, USA.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology, 148*, 547-591.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) in the cat's visual cortex. *Journal of Neurophysiology, 148*, 229-289.

Hummel, J. E. (1997). Structure and binding in object perception. In J. W. Donahoe & V. P. Dorsel (Eds.), *Neural-network models of cognition: Biobehavioral foundations.* (pp. 203-219). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in human shape perception. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. pp. 157-185). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition, 8*(3-5), 489-517.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape-recognition. *Psychological Review, 99*(3), 480-517.

Hummel, J. E., & Stankiewicz, B. J. (1996a). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 93-121). Cambridge, MA: MIT Press.

Hummel, J. E., & Stankiewicz, B. J. (1996b). Categorical relations in shape perception. *Spatial Vision, 10*(3), 201-236.

Hummel, J. E., & Stankiewicz, B. J. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition, 5*(1-2), 49-79.

Humphrey, G. K., & Jolicoeur, P. (1993). An examination of the effects of axis foreshortening, monocular depth cues, and visual-field on object identification. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 46*(1), 137-159.

Humphreys, G. W., Price, C. J., & Riddoch, M. J. (1999). From objects to names: A cognitive neuroscience approach. *Psychological Research, 62*, 118-130.

Humphreys, G. W., & Riddoch, M. J. (1984). Routes to object constancy - implications from neurological impairments of object constancy. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 36*(3), 385-415.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*, 67-103.

Humphreys, G. W., & Rumiati, R. I. (1998). Agnosia without prosopagnosia or alexia: Evidence for stored visual memories specific to objects. *Cognitive Neuropsychology, 15*(3), 243-277.

Huttenlocher, D. P., & Ullman, S. (1987). *Object recognition using alignment.* Paper presented at the First International Conference on Computer Vision, London.

Intraub, H. (1981). Identification and processing of briefly glimpsed visual scenes. In D. Fisher (Ed.), *Eye movements: Cognition and visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Janssen, P., Vogels, R., & Orban, G. A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science, 288*, 2054-2056.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*(4), 289-303.

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology-Revue Canadienne De Psychologie, 42*(4), 461-478.

Jolicoeur, P. (1990). Identification of disoriented objects: A dual systems theory. *Mind and Language, 5*, 387-410.

Jolicoeur, P. (1992). Identification of disoriented objects. In G. W. Humphreys (Ed.), *Understanding Vision* (pp. 180-198). Cambridge, MA.

Jolicoeur, P., Corballis, M. C., & Lawson, R. (1998). The influence of perceived rotary motion on the recognition of rotated objects. *Psychonomic Bulletin & Review, 5*(1), 140-146.

Jolicoeur, P., & Milliken, B. (1989). Identification of disoriented objects - effects of context of prior presentation. *Journal of Experimental Psychology-Learning Memory and Cognition, 15*(2), 200-210.

Jolicoeur, P., Snow, D., & Murray, J. (1987). The time to identify disoriented letters - effects of practice and font. *Canadian Journal of Psychology-Revue Canadienne De Psychologie, 41*(3), 303-316.

Jolicouer, P., Gluck, M. A., & Kosslyn, S. M. (1981). From pictures to words: Making the connection. *Cognitive Psychology, 16*, 213-275.

Jonides, J. (1981). Voluntary versus automatic movements of the mind's eye. In J. Long & A. Baddeley (Eds.), *Attention and Performance IX* (pp. 197–203). Hillsdale, NJ: Lawrence Erlbaum.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43*(4), 346-354.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kanwisher, N., Chun, M. M., McDermott, J., & Ledden, P. J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research, 5*(1-2), 55-67.

Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience, 9*(1), 133-142.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71*, 856–867.

Kohonen, T. (1978). *Associative memories: A system theoretic approach*. Berlin: Springer.

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., & Koenig, O. (1992). Categorical versus coordinate spatial relations - computational analyses and computer-simulations. *Journal of Experimental Psychology-Human Perception and Performance, 18*(2), 562-577.

Kosslyn, S. M., Koenig, O., Barrett, A., Backer Cave, C., Tang, J., & Gabrieli, J. D. E. (1989). Evidence for two types of spatial representations hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 723-735.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience, 20*(9), 3310-3318.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*(5534), 1506-1509.

Lades, M., Vorbrueggen, J. C., Buhmann, J., von der Lange, J.; Malsburg, C., Wuertz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42*, 300-311.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology-Human Perception and Performance, 21*(3), 451-468.

Lavie, N., & Driver, J. (1996). On the spatial extent of attention in object-based visual selection. *Perception & Psychophysics, 58*(8), 1238-1251.

Lavie, N., & Fox, E. (2000). The role of perceptual load in negative priming. *Journal of Experimental Psychology-Human Perception and Performance, 26*(3), 1038-1052.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics, 56*(2), 183-197.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica, 102*(2-3), 221-245.

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology-Human Perception and Performance, 22*(2), 395-416.

Lawson, R., & Humphreys, G. W. (1998). View-specific effects of depth rotation and foreshortening on the initial recognition and priming of familiar objects. *Perception & Psychophysics, 60*(6), 1052-1066.

Lawson, R., & Humphreys, G. W. (1999). The effects of view in depth on the identification of line drawings and silhouettes of familiar objects: Normality and pathology. *Visual Cognition, 6*(2), 165-195.

Lawson, R., Humphreys, G. W., & Jolicoeur, P. (2000). The combined effects of plane disorientation and foreshortening on picture naming: One manipulation or two? *Journal of Experimental Psychology-Human Perception and Performance, 26*(2), 568-581.

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions - evidence for viewpoint-specific recognition procedures. *Perception, 23*(5), 595-614.

Lawson, R., & Jolicoeur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory & Cognition, 26*(4), 791-803.

Lawson, R., & Jolicoeur, P. (1999). The effect of prior experience on recognition thresholds for plane-disoriented pictures of familiar objects. *Memory & Cognition, 27*(4), 751-758.

Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology-Human Perception and Performance, 20*(5), 1015-1036.

Logothetis, N. K., Pauls, J., Bulthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology, 4*(5), 401-414.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in theinferior temporal cortex of monkeys. *Current Biology, 5*(5), 552-563.

Logothetis, N. K., & Sheinberg, D. (1996). Visual object recognition. *Annual Review of Neuroscience, 19*, 577–621.

Lowe, D. G. (1985). *Perceptual organization and visual recognition*. Boston: Kluwer.

Lowe, D. G. (1987). The viewpoint consistency constraint. *International Journal of Computer Vision, 1*, 57-72.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior Temporal Mechanisms for Invariant Object Recognition. *Cerebral Cortex, 4*(5), 523-531.

Mack, A., Tang, B., Tuma, R., & Kahn, S. (1992). Perceptual organization and attention. *Cognitive Psychology, 24*, 475-501.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978). *Representation and recognition of three dimensional shapes.* Paper presented at the Proceedings of the Royal Society of London.

Marsolek, C. J. (1995). Abstract visual-form representations in the left cerebral hemisphere. *Journal of Experimental Psychology-Human Perception and Performance, 21*(2), 375-386.

Marsolek, C. J. (1999). Dissociable neural, subsystems underlie abstract and specific object recognition. *Psychological Science, 10*(2), 111-118.

Marsolek, C. J., Kosslyn, S. M., & Squire, L. R. (1992). Form-specific visual priming in the right cerebral hemisphere. *Journal of Experimental Psychology-Learning Memory and Cognition, 18*(3), 492-508.

Marsolek, C. J., Schacter, D. L., & Nicholas, C. D. (1996). Form-specific visual priming for new associations in the right cerebral hemisphere. *Memory & Cognition, 24*(5), 539-556.

Mattingley, J. B., Bradshaw, J. L., & Bradshaw, J. A. (1995). The effects of unilateral visuospatial neglect on perception of Müller-Lyer illusory figures. *Perception, 24*, 415-433.

McMullen, P. A., & Jolicoeur, P. (1990). The spatial frame of reference in object naming and discrimination of left-right reflections. *Memory & Cognition, 18*(1), 99-115.

McMullen, P. A., & Jolicoeur, P. (1992). Reference frame and effects of orientation on finding the tops of rotated objects. *Journal of Experimental Psychology-Human Perception and Performance, 18*(3), 807-820.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review, 81*, 521-535.

Moore, C. J., & Price, C. J. (1999). Three distinct ventral occipitotemporal regions for reading and object naming. *Neuroimage, 10*(2), 181-192.

Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology-Human Perception and Performance, 23*(2), 339-352.

Murray, J. E. (1995a). Imagining and naming rotated natural objects. *Psychonomic Bulletin & Review, 2*(2), 239-243.

Murray, J. E. (1995b). Negative priming by rotated objects. *Psychonomic Bulletin & Review, 2*(4), 534-537.

Murray, J. E. (1995c). The role of attention in the shift from orientation-dependent to orientation-invariant identification of disoriented objects. *Memory & Cognition, 23*(1), 49-58.

Murray, J. E. (1997). Flipping and spinning: Spatial transformation procedures in the identification of rotated natural objects. *Memory & Cognition, 25*(1), 96-105.

Murray, J. E. (1998). Is entry-level recognition viewpoint invariant or viewpoint dependent? *Psychonomic Bulletin & Review, 5*(2), 300-304.

Murray, J. E. (1999). Orientation-specific effects in picture matching and naming. *Memory & Cognition, 27*(5), 878-889.

Murray, J. E., Jolicoeur, P., McMullen, P. A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory & Cognition, 21*(5), 604-610.

Neumann, E., & Deschepper, B. G. (1992). An inhibition-based fan effect - evidence for an active suppression mechanism in selective attention. *Canadian Journal of Psychology-Revue Canadienne De Psychologie, 46*(1), 1-40.

Nissen, M. J. (1985). Accessing features and objects: Is location special? In M. I. Posner & D. S. Marin (Eds.), *Attention and Performance XI* (pp. 205-219). Hillsdale, NJ: Erlbaum.

Olshausen, B., Anderson, C., & Van Essen, D. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience, 13*(11), 4700-4719.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68*(1), 70-84.

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135-151). Hillsdale, NJ: Erlbaum.

Palmer, S. E. (1978). Structural aspects of similarity. *Memory and Cognition, 6*, 91-97.

Pashler, H. (1995). Attention and visual perception: Analyzing divided attention. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual Cognition* (Vol. 2). Cambridge, MA: MIT Press.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centered and object-centered coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86*(1), 159-173.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3-dimensional objects. *Nature, 343*(6255), 263-266.

Posner, M. I. (1969). Abstraction and the process of recognition. In J. T. Spence & G. Bower (Eds.), *The psychology of learning and motivation* (pp. 43-100). New York: Academic Press.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Posner, M. I., & Keele, S. W. (1967). Decay of visual information from a single letter. *Science, 158*, 137-139.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160-174.

Price, C. J., Moore, C. J., Humphreys, G. W., Frackowiak, R. S. J., & Friston, K. J. (1996). The neural regions sustaining object recognition and naming. *Proceedings of the Royal Society of London Series B-Biological Sciences, 263*(1376), 1501-1507.

Raajimakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to Deal with "The Language-as-Fixed-Effect Fallacy": Common Misconceptions and Alternative Solutions. *Journal of Memory and Language, 41*, 416-426.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7*(1-3), 345-376.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*, 245-277.

Rensink, R. A., Oregan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368-373.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019-1025.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3*, 1199-1204.

Rock, I., & Guttman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 275–285.

Rollenhagen, J. E., & Olson, C. R. (2000). Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science, 287*(5457), 1506-1508.

Rosch, E., Mervis, C. B., Gray, W. D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Rosielle, L. J., & Cooper, E. E. (2001). Categorical perception of relative orientation in visual object recognition. *Memory & Cognition, 29*(1), 68-82.

Rosielle, L. J., Crabb, B. T., & Cooper, E. E. (2002). Attentional coding of categorical relations in scene perception: Evidence from the flicker paradigm. *Psychonomic Bulletin & Review, 9*(2), 319-326.

Saiki, J., & Hummel, J. E. (1998). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology-Human Perception and Performance, 24*(1), 227-251.

Sanocki, T., Bowyer, K. W., Heath, M. D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology-Human Perception and Performance, 24*(1), 340-349.

Sartori, G., Miozzo, M., & Job, R. (1993). Category-specific naming impairments ? Yes. *Quarterly Journal of Experimental Psychology-Animal Behavior Processes, 46*, 489-504.

Schiller, P. H. (1995). Effects of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature, 376*, 342–344.

Schiller, P. H., Logothetis, C., & Charles, E. R. (1990). Functions of the colour-opponent and broad-band channels of the visual system. *Nature, 343*, 68-70.

Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American, 203*, 60-68.

Shapiro, K., Driver, J., Ward, R., & Sorenson, R. E. (1997). Priming from the attentional blink. *Psychological Science, 8*, 95-100.

Shapiro, K. L. (1994). The attentional blink: The brain's eyeblink. *Current Directions in Psychological Science, 3*, 86-89.

Shepard, R. N., & Cooper, L. A. (1982). *Mental Images and their Transformations*. Cambridge: MIT Press/Bradford.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174-215.

Srinivas, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology-Learning Memory and Cognition, 19*(3), 582-602.

Srinivas, K. (1995). Representation of rotated objects in explicit and implicit memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*(4), 1019-1036.

Stankiewicz, B. J. (2002). Empirical Evidence for Independent Dimensions in the Visual Representation of Three-Dimensional Shape. *Journal of Experimental Psychology-Human Perception and Performance, 28*(4), 913–932.

Stankiewicz, B. J., & Hummel, J. E. (2002). Automatic priming for translation- and scale-invariant representations of object shape. *Visual Cognition, 9*(6), 719–739.

Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology-Human Perception and Performance, 24*(3), 732-744.

Stankiewicz, B. S. (1997). *The role of attention in viewpoint-invariant object recognition.* Unpublished Doctoral Thesis, University of Claifornia Los Angeles, Los Angeles.

Tanaka, J., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 146A*, 225-245.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262*(5134), 685-688.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109-139.

Tarr, M. (1999). News on views: Pandemonium revisited. *Nature Neuroscience, 2*(11).

Tarr, M. J. (1995). Rotating objects to recognize them - a case-study on the role of viewpoint dependency in the recognition of 3-dimensional objects. *Psychonomic Bulletin & Review, 2*(1), 55-82.

Tarr, M. J., & Bulthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views - Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology-Human Perception and Performance, 21*(6), 1494-1505.

Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*(1-2), 1-20.

Tarr, M. J., Bulthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8*(4), 282-289.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition, 67*(1-2), 73-110.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape-recognition. *Cognitive Psychology, 21*(2), 233-282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame. *Psychological Science, 1*(4), 253-256.

Tarr, M. J., & Pinker, S. (1991). Orientation-dependent mechanisms in shape-recognition - further issues. *Psychological Science, 2*(3), 207-209.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1*(4), 275-277.

Tipper, S. (1985). The negative priming effect: Inhibitory effects of ignored primes. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 37A*, 571-590.

Tipper, S. P., & Cranston, M. (1985). Selective attention and priming - inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 37*(4), 591-611.

Tipper, S. P., & Driver, J. (1988). Negative priming between pictures and words in a selective attention task - evidence for semantic processing of ignored stimuli. *Memory & Cognition, 16*(1), 64-70.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*, 242–248.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review, 76*, 282–299.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American, 255*, 106–115.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 353*(1373), 1295-1306.

Treisman, A., & DeSchepper, B. (1996). Visual memory, attention, and object tokens. *International Journal of Psychology, 31*(3-4), 5024-5024.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107–141.

Treisman, A. M., & Kanwisher, N. G. (1998). Perceiving visually presented objects: recognition, awareness, and modularity. *Current Opinion in Neurobiology, 8*(2), 218-226.

Tsal, Y., & Lavie, N. (1988). Attending to color and shape - the special role of location in selective visual processing. *Perception & Psychophysics, 44*(1), 15-21.

Tsal, Y., & Lavie, N. (1993). Location dominance in attending to color and shape. *Journal of Experimental Psychology-Human Perception and Performance, 19*(1), 131-139.

Turnbull, O. H. (1997). A double dissociation between knowledge of object identity and object orientation. *Neuropsychologia, 24*(3), 456-469.

Turnbull, O. H., Beschin, N., & Della Sala, S. (1997). Agnosia for object orientation: implications for theories of object recognition. *Neuropsychologia, 35*, 153-163.

Turnbull, O. H., Laws, K. R., & McCarthy, R. A. (1995). Object recognition without knowledge of object orientation. *Cortex, 31*, 387-395.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*, 169-193.

Tversky, B., & Hemenway, K. (1991). Parts and the basic level in natural categories and artificial stimuli: Comments on Murphy. *Memory & Cognition, 19*, 439-442.

Ullman, S. (1989). Aligning pictorial descriptions - an approach to object recognition. *Cognition, 32*(3), 193-254.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition, 67*(1-2), 21-44.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 13*(10), 992-1006.

Ungerleider, L. G., Ganz, L., & Pribram, K. H. (1977). Effects of pulvinar, prestriate and inferotemporal lesions. *Experimental Brain Research, 27*, 251–269.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle & M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549-586). Cambridge, MA: MIT Press.

Vannucci, M., & Viggiano, M. P. (2000). Category effects on the processing of plane-rotated objects. *Perception, 29*(3), 287-302.

Vecera, S. P., Behrmann, M., & Fliapek, J. C. (2001). Attending to the parts of a single object: Part-based selection limitations. *Perception & Psychophysics, 63*(2), 308-321.

Vecera, S. P., Behrmann, M., & McGoldrick, J. (2000). Selective attention to the parts of an object. *Psychonomic Bulletin & Review, 7*(2), 301-308.

Verfaillie, K., & Boutsen, L. (1995). A corpus of 714 full-color images of depth-rotated objects. *Perception & Psychophysics, 57*(7), 925-961.

Vogels, R., Biederman, I., Bar, M., & Leuven, K. U. (1999). Sensitivity of macaque inferior temporal neurons to differences in view invariant vs metric properties of depth rotated objects. *Investigative Ophthalmology & Visual Science, 40*(4), S776-S776.

Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *Journal of Cognitive Neuroscience, 13*(4), 444-453.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology, 5*, 520-526.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience, 5*(5), 491-499.

Warrington, E. K., & Davidoff, J. (2000). Failure at object identification improves mirror image matching. *Neuropsychologia, 38*(9), 1229-1234.

Warrington, E. K., & James, M. (1988). Visual apperceptive agnosia - a clinico-anatomical study of 3 cases. *Cortex, 24*(1), 13-32.

Warrington, E. K., & Taylor, A. M. (1973). The contribution of the right parietal lobe to object recognition. *Cortex, 9*(2), 152-164.

Warrington, E. K., & Taylor, A. M. (1978). Two categorical stages of object recognition. *Perception, 7*, 584-694.

Weiskrantz, L. (1990). Visual prototypes, memory and the inferotemporal lobe. In E. Iwai & M. Mishkin (Eds.), *Vision, Memory and the Temporal Lobe* (pp. 13–28). New York, NY: Elsevier.

Weiskrantz, L., & Saunders, R. C. (1984). Impairments of visual object transforms in monkeys. *Brain, 107*(DEC), 1033-1072.

Willems, B., & Wagemans, J. (2001). Matching multicomponent objects from different viewpoints: mental rotation as normalization? *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1090-1115.

Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature, 222*, 960–962.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24*(1), 11-17.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search - an Alternative to the Feature Integration Model for Visual-Search. *Journal of Experimental Psychology-Human Perception and Performance, 15*(3), 419-433.

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology-Human Perception and Performance, 19*(3), 676-681.

Yantis, S., & Johnson, D. N. (1990). Mechanisms of attentional priority. *Journal of Experimental Psychology-Human Perception and Performance, 16*(4), 812-825.

Appendix

Appendix 1: Descriptive Data and Additional Test Statistics

Experiment 1

Prime and Probe Trial

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTSPLIT	42	648.9764	10065.69	100.3279	15.48094
ATTSAME	42	595.2300	6717.65	81.9612	12.64689
IGNSPLI	42	816.7461	14750.38	121.4511	18.74032
IGNSAME	42	762.4381	9450.64	97.2144	15.00050
UNPRIMED	42	812,6607	11524,21	107,3509	16,56460

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTSPLIT	42	.154762	.030343	.174191	.026878
ATTSAME	42	.071429	.013744	.117234	.018090
IGNSPLI	42	.134921	.028778	.169641	.026176
IGNSAME	42	.111111	.018519	.136083	.020998
UNPRIMEP	42	.142857	.016357	.127894	.019735

Summary of all Effects Errors in Percent; Design: 1-Attention, 2-Configuration

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	.004134	41	.011925	.346635	.559254
2	1	.120536	41	.015861	7.599542	.008676
12	1	.037202	41	.021958	1.694215	.200313

Probe Trial

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTSPLIT	42	.031746	.005743	.075780	.011693
ATTINTAC	42	.031746	.008453	.091939	.014186
IGNSPLIT	42	.023810	.004839	.069565	.010734
IGNINTAC	42	.035714	.006146	.078396	.012097
UNPRIMED	42	.057540	.005924	.076969	.011876

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 42, df = 4) = 12.80000 p < .01231

Coeff. of Concordance = .07619 Aver. rank r = .05366

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTSPLIT	2.952381	124.0000	.190476	.454683
ATTINTAC	2.809524	118.0000	.190476	.551632
IGNSPLIT	2.761905	116.0000	.142857	.417392
IGNINTAC	3.000000	126.0000	.214286	.470377
UNPRIMED	3.476191	146.0000	.345238	.461811

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 42, df = 3) = 1.265625 p < .73731

Coeff. of Concordance = .01004 Aver. rank r = -.0141

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTSPLIT	2.535714	106.5000	.190476	.454683
ATTINTAC	2.440476	102.5000	.190476	.551632
IGNSPLIT	2.440476	102.5000	.142857	.417392
IGNINTAC	2.583333	108.5000	.214286	.470377

Experiment 2

Prime and Probe Trial

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTPART	42	614.8090	14919.69	122.1462	18.84756
ATTSNDE	42	688.9402	20562.79	143.3973	22.12669
IGNPART	42	748.0671	25974.32	161.1655	24.86838
IGNSNDE	42	735.1071	23518.91	153.3588	23.66377
UNPRIMED	42	749.3288	10712.69	103.5021	15.97073

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTPART	42	.230159	.062266	.249532	.038504
ATTSNDE	42	.182540	.049877	.223332	.034461
IGNPART	42	.142857	.054975	.234467	.036179
IGNSNDE	42	.206349	.064783	.254524	.039274
UNPRIMED	42	.234127	.036666	.191484	.029547

Summary of all Effects; Design: 1-Attention. 2-Primetype

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	.042328	41	.043683	.968981	.330708
2	1	.002646	41	.077171	.034281	.854023
12	1	.129630	41	.057814	2.242188	.141951

Probe Trial Errors

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTPART	42	.071429	.024584	.156792	.024194
ATTSNDE	42	.087302	.027423	.165599	.025552
IGNPART	42	.039683	.017357	.131746	.020329
IGNSNDE	42	.031746	.015228	.123401	.019041
UNPRIMED	42	.075397	.013824	.117577	.018143

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. ($N = 42$. $df = 3$) = 6.106383 $p < .10657$

Coeff. of Concordance = .04846 Aver. rank r = .02526

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTPART	2.583333	108.5000	.071429	.156792
ATTSNDE	2.690476	113.0000	.087302	.165599
IGNPART	2.380952	100.0000	.039683	.131746
IGNSNDE	2.345238	98.5000	.031746	.123401

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. ($N = 42$. $df = 4$) = 9.712610 $p < .04558$

Coeff. of Concordance = .05781 Aver. rank r = .03483

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTPART	3.035714	127.5000	.071429	.156792
ATTSNDE	3.154762	132.5000	.087302	.165599
IGNPART	2.785714	117.0000	.039683	.131746
IGNSNDE	2.726191	114.5000	.031746	.123401
UNPRIMED	3.297619	138.5000	.075397	.117577

Experiment 3

Prime and Probe Trial

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTINTAC	36	629.9697	9178.04	95.8021	15.96702
ATTSPLIT	36	661.6194	11473.56	107.1147	17.85245
IGNINTAC	36	767.6025	13111.70	114.5063	19.08439
IGNSPLIT	36	889.9533	26388.09	162.4441	27.07402
UNPRIINT	36	829.8281	22500.84	150.0028	25.00047
UNPRISPL	36	887.0442	20501.65	143.1840	23.86400

Descriptive Statistics all Erros in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTINTAC	36	.078704	.011883	.109008	.018168
ATTPART	36	.199074	.036221	.190319	.031720
IGNINTAC	36	.171296	.026168	.161767	.026961
IGNPART	36	.171296	.032518	.180326	.030054
UNPRIMED	36	.120370	.020018	.141484	.023581
UNPSPLIT	36	.199074	.023523	.153372	.025562

Summary of all Effects all Erros in Percent; Design: 1-Condition. 2-Configuration

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	.019419	70	.021800	.89078	.414935
2	1	.237783	35	.018471	12.87348	.001009
12	2	.067258	70	.021756	3.09154	.051687

Summary of all Effects all Erros in Percent; Design: 1-Attention, 2-Configuration

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	.037809	35	.016380	2.308210	.137675
2	1	.130401	35	.020084	6.492865	.015377
12	1	.130401	35	.019290	6.760000	.013562

Probe Trial Errors

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTINTAC	36	.023148	.005004	.070742	.011790
ATTSPLIT	36	.083333	.015079	.122798	.020466
IGNINTAC	36	.092593	.011817	.108704	.018117
IGNSPLIT	36	.092593	.016578	.128757	.021460
UNPRIMED	36	.050926	.007650	.087464	.014577
UNPRISPL	36	.106481	.020877	.144490	.024082

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 36, df = 5) = 14.24185 p < .01416

Coeff. of Concordance = .07912 Aver. rank r = .05281

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTSAME	2.791667	100.5000	.138889	.424451
ATTPART	3.638889	131.0000	.500000	.736788
IGNSAME	3.833333	138.0000	.555556	.652225
IGNPART	3.722222	134.0000	.555556	.772545
UNPRIMED	3.194444	115.0000	.305556	.524783
UNPRPART	3.819444	137.5000	.638889	.866941

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 36, df = 3) = 11.40278 p < .00974

Coeff. of Concordance = .10558 Aver. rank r = .08003

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTSAME	2.013889	72.50000	.138889	.424451
ATTPART	2.638889	95.00000	.500000	.736788
IGNSAME	2.694444	97.00000	.555556	.652225
IGNPART	2.652778	95.50000	.555556	.772545

Wilcoxon Matched Pairs Test

	N	T	Z	p-level
ATTINTAC & ATTSPLIT	36	24,0000	2,485251	,012951
ATTINTAC & IGNINTAC	36	7,0000	3,010198	,002613
ATTINTAC & IGNSPLIT	36	27,5000	2,319567	,020371
ATTSPLIT & ATTINTAC	36	24,0000	2,485251	,012951
ATTSPLIT & IGNINTAC	36	115,5000	,357122	,721003
ATTSPLIT & IGNSPLIT	36	66,0000	,497050	,619157
IGNINTAC & IGNSPLIT	36	134,0000	,121660	,903169

Experiment 4

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ANIMAL0	29	827.4738	18194.04	134.8853	25.04757
ANIMAL60	29	863.6086	22979.59	151.5902	28.14959
ANIMAL120	29	913.3959	17725.65	133.1377	24.72305
BASE0	29	760.8355	13462.40	116.0276	21.54578
BASE60	29	829.9748	10488.55	102.4136	19.01773
BASE120	29	885.8993	18523.32	136.1004	25.27321
NOBASE0	29	772.4410	8136.94	90.2050	16.75064
NOBASE60	29	766.1438	9627.83	98.1215	18.22071
NOBASE120	29	778.0083	8519.13	92.2991	17.13951

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ANIMAL0	29	.038793	.004580	.067674	.012567
ANIMAL60	29	.060345	.008505	.092224	.017126
ANIMAL12	29	.172414	.018319	.135348	.025133
BASE0	29	.025862	.003772	.061413	.011404
BASE60	29	.047414	.006042	.077731	.014434
BASE120	29	.081897	.011469	.107092	.019886
NOBASE0	29	.025862	.003772	.061413	.011404
NOBASE60	29	.025862	.002655	.051531	.009569
NOBASE12	29	.004310	.000539	.023212	.004310

Experiment 5

Prime and Probe Trial

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	30	708.4973	24945.47	157.9413	28.83601
ATTSAME	30	655.0750	23325.92	152.7283	27.88424
IGROT	30	828.4723	19436.83	139.4160	25.45377
IGSAME	30	758.7500	13392.17	115.7245	21.12831
UNPRROT	30	823.9667	35031.63	187.1674	34.17193
UPRIMEDS	30	814.4890	40505.98	201.2610	36.74506

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	30	.083333	.022989	.151620	.027682
ATTSAME	30	.116667	.024713	.157203	.028701
IGROT	30	.141667	.033118	.181983	.033225
IGSAME	30	.150000	.058621	.242117	.044204
ROTUNPRI	30	.141667	.028807	.169728	.030988
SAMUNPRI	30	.150000	.037069	.192533	.035152

Summary of all Effects Errors in Percent; Design: 1-Attention. 2-Rotation

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	.063021	29	.024228	2.601186	.117615
2	1	.013021	29	.041038	.317287	.577573
12	1	.004688	29	.032705	.143328	.707752

Summary of all Effects Errors in Percent; Design: 1-Condition. 2-View

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	.042014	58	.025491	1.648192	.201280
2	1	.012500	29	.038362	.325843	.572515
12	2	.003125	58	.037608	.083095	.920373

Probe Trial Errors

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	30	.025000	.005819	.076282	.013927
ATTSAME	30	.050000	.018966	.137715	.025143
IGROT	30	.033333	.011782	.108543	.019817
IGSAME	30	.041667	.013290	.115283	.021048
UPRIMEDR	30	.025000	.005819	.076282	.013927
UPRIMEDS	30	.050000	.010345	.101710	.018570

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 30, df = 5) = 1.952381 p < .85569

Coeff. of Concordance = .01302 Aver. rank r = -.0210

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	3.416667	102.5000	.100000	.305129
ATTSAME	3.516667	105.5000	.200000	.550861
IGROT	3.416667	102.5000	.133333	.434172
IGSAME	3.516667	105.5000	.166667	.461133
UNPRIMROT	3.416667	102.5000	.100000	.305129
UNPRIMED	3.716667	111.5000	.200000	.406838

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 30, df = 3) = .2857143 p < .96269

Coeff. of Concordance = .00317 Aver. rank r = -.0312

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	2.466667	74.00000	.100000	.305129
ATTSAME	2.533333	76.00000	.200000	.550861
IGROT	2.466667	74.00000	.133333	.434172
IGSAME	2.533333	76.00000	.166667	.461133

Experiment 6

Prime and Probe Trial

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	30	603.7693	8482.92	92.1027	16.81558
ATTSAME	30	584.8553	4730.53	68.7789	12.55725
IGNROT	30	809.6437	13366.51	115.6136	21.10806
IGNSAME	30	722.0067	10304.76	101.5124	18.53354
UNPRIMED	30	787.3253	29132.50	170.6824	31.16221
UNPRIROT	30	805.6413	11788.18	108.5734	19.82269

Summary of all Effects for Latencies in ms; Design: 1-CONDITIO, 2-VIEW

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	712296.9	58	11426.61	62.33666	.000000
2	1	77958.8	29	4311.50	18.08158	.000201
12	2	23821.5	58	5919.80	4.02404	.023092

Error Analysis for All Errors

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	30	.172222	.025830	.160718	.029343
ATTSAME	30	.055556	.008301	.091112	.016635
IGNROT	30	.050000	.006034	.077682	.014183
IGNSAME	30	.061111	.016252	.127482	.023275
UNPRROT	30	.138889	.019317	.138985	.025375
UNPRIMED	30	.044444	.007535	.086805	.015848

Summary of all Effects for Errors in Percent; Design: 1-Attention. 2-View

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	.102083	29	.015398	6.62986	.015397
2	1	.083565	29	.013163	6.34870	.017510
12	1	.122454	29	.010864	11.27186	.002213

Summary of all Effects Errors in Percent; Design: 1-Condition. 2-View

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	.052006	58	.012574	4.13584	.020936
2	1	.200000	29	.012261	16.31250	.000360
12	2	.069907	58	.011957	5.84646	.004863

Probe Trial Errors

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	31	.043011	.033274	.182411	.032762
ATTSAME	31	.037634	.032796	.181096	.032526
IGNROT	31	.086022	.062724	.250448	.044982
IGNSAME	31	.080645	.131243	.362274	.065066
UNPRROT	31	.080645	.062724	.250448	.044982
UNPRIMED	31	.306452	.246476	.496463	.089167

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 30, df = 5) = 9.005236 p < .10889

Coeff. of Concordance = .06003 Aver. rank r = .02762

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	3.600000	108.0000	.200000	.406838
ATTSAME	3.266667	98.0000	.100000	.305129
IGNROT	3.500000	105.0000	.166667	.379049
IGNSAME	3.333333	100.0000	.166667	.530669
UNPRROT	3.300000	99.0000	.100000	.305129
UNPRIMED	4.000000	120.0000	.400000	.621455

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 30, df = 3) = 1.813187 p < .61207

Coeff. of Concordance = .02015 Aver. rank r = -.0136

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	2.616667	78.50000	.200000	.406838
ATTSAME	2.400000	72.00000	.100000	.305129
IGNROT	2.550000	76.50000	.166667	.379049
IGNSAME	2.433333	73.00000	.166667	.530669

Experiment 7

Prime and Probe Trial

Summary of all Effects priming rt; Design: 1-Attention. 2-View

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	616054.6	27	14355.12	42.91532	.000001
2	1	47994.6	27	10057.66	4.77194	.037785
12	1	1439.2	27	6537.07	.22017	.642681

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTREFL	28	671.8868	15023.93	122.5721	23.16396
ATTSAME	28	629.3036	11134.83	105.5217	19.94173
IGNREFL	28	813.5118	14052.14	118.5417	22.40228
IGNSAME	28	768.8989	14968.42	122.3455	23.12112
UNPRIREF	28	799.4729	25178.68	158.6779	29.98731
UNPRIMED	28	834.7386	30273.62	173.9931	32.88161

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTREFL	28	.178571	.031746	.178174	.033672
ATTSAME	28	.160714	.024140	.155371	.029362
IGNREFL	28	.205357	.032655	.180708	.034151
IGNSAME	28	.250000	.041667	.204124	.038576
UNPRIMED	28	.169643	.041915	.204731	.038690
UNPRREFL	28	.151786	.033978	184332	.034835

Summary of all Effects Errors in Percent; Design: 1-Condition. 2-View

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	.077381	54	.036100	2.143512	.127113
2	1	.003348	27	.021095	.158720	.693471
12	2	.013393	54	.027668	.484064	.618922

Probe Trial Errors

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTREFL	28	.089286	.019511	.139680	.026397
ATTSAME	28	.053571	.010913	.104464	.019742
IGNREFL	28	.116071	.025380	.159312	.030107
IGNSAME	28	.107143	.025132	.158532	.029960
UNPRIREF	28	.044643	.009507	.097505	.018427
UNPRIMED	28	.044643	.009507	.097505	.018427

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. $(N = 28, df = 5) = 6.819086$ $p < .23448$

Coeff. of Concordance = .04871 Aver. rank r = .01347

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTREFL	3.660714	102.5000	.357143	.558721
ATTSAME	3.303571	92.5000	.214286	.417855
IGNREFL	3.857143	108.0000	.464286	.637248
IGNSAME	3.767857	105.5000	.428571	.634126
UNPRIREF	3.214286	90.0000	.178571	.390021
UNPRIMED	3.196429	89.5000	.178571	.390021

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. (N = 28, df = 3) = 2.562914 p < .46404

Coeff. of Concordance = .03051 Aver. rank r = -.0054

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTREFL	2.500000	70.00000	.357143	.558721
ATTSAME	2.267857	63.50000	.214286	.417855
IGNREFL	2.642857	74.00000	.464286	.637248
IGNSAME	2.589286	72.50000	.428571	.634126

Experiment 8

Prime and Probe Trial

Descriptive Statistics Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	28	708.8661	19112.64	138.2485	26.12650
ATTSAME	28	643.6846	13836.85	117.6302	22.23001
IGNROT	28	783.4825	20935.73	144.6918	27.34419
IGNSAME	28	810.7914	21979.04	148.2533	28.01724
UNPRROT	28	824.2143	20376.08	142.7448	26.97623
UNPRIMED	28	899.9143	39403.22	198.5025	37.51344
VIEW30	28	866.1314	41464.12	203.6274	38.48196
VIEW90	28	857.1311	21445.71	146.4435	27.67523

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	28	.205357	.041915	.204731	.038690
ATTSAME	28	.160714	.033399	.182755	.034537
IGNROT	28	.258929	.039269	.198165	.037450
IGNSAME	28	.214286	.035714	.188982	.035714
UNPRROT	28	.214286	.054233	.232879	.044010
UNPRIMED	28	.160714	.042659	.206540	.039032

Summary of all Effects Errors in Percent; Design: 1-Condition. 2-View

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	.049479	54	.035204	1.405479	.254076
2	1	.095238	27	.029652	3.211896	.084319
12	2	.000372	54	.042424	.008769	.991271

Probe Trial Errors

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	28	.053571	.010913	.104464	.019742
ATTSAME	28	.071429	.013228	.115011	.021735
IGNROT	28	.098214	.024719	.157223	.029712
IGNSAME	28	.080357	.014137	.118899	.022470
UNPRROT	28	.053571	.015542	.124669	.023560
UNPRSAME	28	.053571	.020172	.142028	.026841

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. $(N = 28, df = 5) = 3.795455$ $p < .57923$

Coeff. of Concordance = .02711 Aver. rank r = -.0089

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	3.392857	95.0000	.214286	.417855
ATTSAME	3.553571	99.5000	.285714	.460044
IGNROT	3.767857	105.5000	.392857	.628890
IGNSAME	3.696429	103.5000	.321429	.475595
UNPRROT	3.339286	93.5000	.214286	.498675
UNPRSAME	3.250000	91.0000	.214286	.568112

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. $(N = 28, df = 3) = 1{,}476563$ $p < {,}68769$

Coeff. of Concordance = ,01758 Aver. rank r = -,0188

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	2,357143	66,00000	,214286	,417855
ATTSAME	2,464286	69,00000	,285714	,460044
IGNROT	2,625000	73,50000	,392857	,628890
IGNSAME	2,553571	71,50000	,321429	,475595

Experiment 9

Prime and Probe Trial

Descriptive Statistics: Latencies in ms

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	40	625.6420	7479.14	86.4820	13.67401
ATTSAME	40	599.4135	6195.15	78.7092	12.44502
IGNROT	40	768.3987	10415.72	102.0574	16.13670
IGNSAME	40	722.3522	9974.58	99.8728	15.79128
UNPRIMED	40	787.1753	18679.37	136.6725	21.60982

Descriptive Statistics Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	40	.083333	.014245	.119352	.018871
ATTSAME	40	.087500	.018501	.136017	.021506
IGNROT	40	.062500	.009526	.097603	.015432
IGNSAME	40	.091667	.008476	.092064	.014557
UNPRIMED	40	.070833	.008387	.091579	.014480

Summary of all Effects Errors in Percent; Design: 1-Group. 2-Attention. 3-View

	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	.017361	38	.018896	.918762	.343859
2	1	.002778	38	.010965	.253333	.617643
3	1	.011111	38	.010234	1.085714	.304007
12	1	.011111	38	.010965	1.013333	.320473
13	1	.002778	38	.010234	.271429	.605398
23	1	.006250	38	.011001	.568106	.455658
123	1	.006250	38	.011001	.568106	.455658

Probe Trial Errors

Descriptive Statistics Probe Errors in Percent

	Valid N	Mean	Variance	Std.Dev.	Standard Error
ATTROT	40	.087500	.071207	.266847	.042192
ATTSAME	40	.062500	.049412	.222289	.035147
IGNROT	40	.062500	.049412	.222289	.035147
IGNSAME	40	.091667	.123860	.351938	.055646
UNPRIMED	40	.087500	.123914	.352014	.055658

Friedman ANOVA and Kendall Coeff. of Concordance

ANOVA Chi Sqr. ($N = 40$, $df = 3$) = 1.500000 $p < .68227$

Coeff. of Concordance = .01250 Aver. rank r = -.0128

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	2.450000	98.0000	.150000	.533494
ATTSAME	2.500000	100.0000	.125000	.334932
IGNROT	2.450000	98.0000	.100000	.303822
IGNSAME	2.600000	104.0000	.200000	.464096

Friedman ANOVA and Kendall Coeff. of Concordance

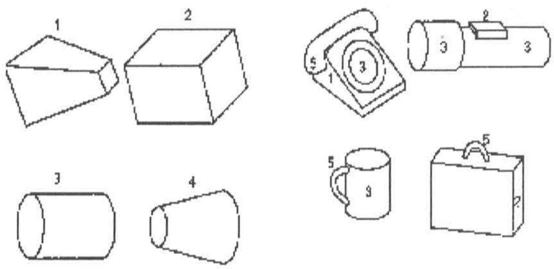
ANOVA Chi Sqr. ($N = 40$, $df = 4$) = 1.407407 $p < .84290$

Coeff. of Concordance = .00880 Aver. rank r = -.0166

	Average Rank	Sum of Ranks	Mean	Std.Dev.
ATTROT	2.937500	117.5000	.150000	.533494
ATTSAME	2.987500	119.5000	.125000	.334932
IGNROT	2.925000	117.0000	.100000	.303822
IGNSAME	3.112500	124.5000	.200000	.464096
UNPRIMED	3.037500	121.5000	.175000	.446496

Appendix 2: Rating Study - Introduction and Instructions

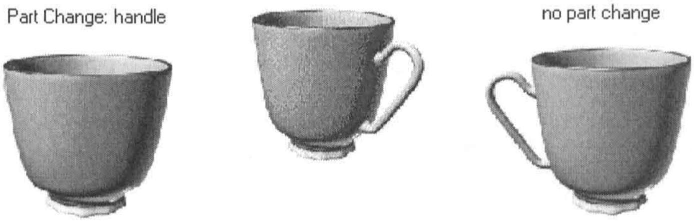
In this Questionnaire you are asked to rate how objects change when they are rotated in depth!
Before we start, please look at the figure below. It shows that you can think of objects as consisting of certain basic shapes or parts.



If we rotate an object, ist possible that some parts disappear or new parts appear that were not visible before.

Part Change: handle

no part change



Instructions:

In each display you will be shown 2 views of one object, which differ in their orientation in depth.

Please compare the two views and look for major changes in visible parts (like legs, handles, ears, a.s.o.) between them.

If you think one view contains parts that are not visible in the other view (e.g., because they are occluded) please press P!

If you think both views contain the same visible parts of the object please press "S"!
Press SPACE to start with a few practice trials!

Appendix 3: List of Stimuli

Experiment 1

rabbit	peacock	pocketbook	tennisracket
horse	swan	broom	table
skunk	hammer	spoon	pot
brush	telephone	mouse	shirt
bird	windmill	hanger	pen
sailboat	trumpet	helicopter	chicken
key	gun	violin	duck
train	gorilla	dog	spider
ladder	elephant	frying pan	kangaroo
pig	frog	cat	snake
truck	bed	desk	bear
rocking chair	plane	dresser	shoe
giraffe	saw	sheep	
wrench	car	iron	
fish	lion	house	
guitar	cow	axe	
snail	pipe	fork	
motorcycle	screwdriver	chair	
kite	umbrella	gator	
plug	ant	wagon	
racoon	bike	scissor	
camel	squirrel	cup	
basket	kettle	piano	
tiger	toothbrush	french horn	

Experiment 2

cow	broom	key	wagon
cow2	broom2	key2	wagon2
plug	cup	sailboat	pig
plug2	cup2	sailboat2	pig2
windmill	frog	swan	fork
windmill2	frog2	swan2	fork2
table	bed	kettle	umbrella
table2	bed2	kettle2	umbrella2
basket	brush	car	saw
basket2	brush2	car2	saw2
pot	cat	chicken	rabbit
pot2	cat2	chicken2	rabbit2
house	telephone	elephant	kangaroo
house2	telephone2	elephant2	kangaroo2
duck	piano	mouse	pen
duck2	piano2	mouse2	pen2
trumpet	shoe	truck	snake
trumpet2	shoe2	truck2	snake2
iron	bear	pan	kite
iron2	bear2	pan2	kite2
fish	horse		
fish2	horse2		

Experiment 3

Target		Filler	
frog	hammer	bike	accordeon
giraffe	cup	bird	ashtray
tennisracket	swan	camel	nut
motorcycle	axe	car	brush
harp	lion	glove	owl
guitar	butterfly	duck	corn
fish	rocking chair	fence	plug
house	dog	alligator	onion
trumpet	sheep	grapes	fhorn
elephant	mouse	gorilla	mitten
windmill	umbrella	hanger	lobster
cat	snake	cow	goat
watch	snail	kangaroo	screwdriver
horse	aeroplane	ostrich	kite
spider	eagle	carrot	football
wrench	fork	peacock	pen
piano	ladder	shirt	flower
rooster	gun	pipe	artichoke
		tiger	saw
		pig	apple
		rabbit	toothbrush
		scissor	banana
		spoon	violin
		squirrel	leaf

Experiment 4

No-Base	Base	Animals
banana	baby carriage	ant
brush	bed	bird
carrot	bike	elephant
comb	boot	camel
football	cake	cat
fork	candle	chicken
glove	chair	cow
guitar	couch	dog
gun	cup	donkey
hammer	desk	duck
key	harp	frog
knife	helicopter	gorilla
leaf	house	grasshopper
lock	iron	horse
pen	ironing board	monkey
ring	kettle	mouse
saw	lamp	owl
scissors	pitcher	peacock
screwdriver	sailboat	pig
spoon	table	rabbit
tennisracket	telephone	snail
trumpet	ashtray	squirrel
watch	truck	tiger
whistle	watering can	turtle

Experiment 5

Target		Filler	
saw	carrot	banana	rolling pin
leaf	lock	screwdriver	nailfile
guitar	broom	envelope	cigarette
umbrella	trumpet	anchor	mitten
book	plug	comb	chisel
glove	fork	knife	pliers
scissors	french horn	toothbrush	watermellon
pineapple	hammer	spoon	ball
key	violin	whistle	belt
butterfly	brush	baseball bat	potato
lightbulb	watch	barrel	ruler
lobster	corn	tennisracket	spool
		wrench	star
		football	sun
		pen	tomato
		pepper	nut

Experiment 6

Target		Filler	
alligator	bird	apple	accordeon
couch	giraffe	axe	anchor
desk	house	banana	ant
kettle	plane	barrel	ashtray
piano	snail	basket	bear
squirrel	truck	baseball bat	beetle
		broom	belt
boot	bike	brush	cake
chicken	chair	bottle	cherry
helicopter	frog	cup	doll
candle	pants	fish	fence
jacket	duck	flag	flower
wineglass	telephone	glasses	frying pan
		guitar	grapes
bed	baby carriage	gun	hanger
cat	car	hammer	iron
elephant	dog	key	kite
mouse	motorcycle	knife	pineapple
sailboat	rabbit	pipe	pliers
table	suitcase	scissors	sheep
		screwdriver	sled
		shoe	swing
		snake	toaster
		trumpet	vase

Experiment 7

axe	glasses	pistol
banana	guitar	plane
bike	hammer	shark
boot	harp	ship
bus	helicopter	shoe
camel	hippo	shovel
cannon	horse	snail
car	iron	sofa
carriage	ironing board	spoon
chair	kettle	stapler
coffee machine	key	suitcase
cow	knife	toilet
crocodile	lamp	toothbrush
desk	microscope	torch
dog	motorbike	turtle
dolphin	phone	vacuum cleaner
drill	piano	watch
eagle	pig	wrench
fork	pipe	

Experiment 8

See Experiment 7

Experiment 9

Target		**Filler**	
ship	pipe	fire extinguisher	corkscrew
motorbike	stapler	harp	hammer
watch	bird	ironing board	shovel
dolphin	iron	camera	lock
cow	glasses	phone	tank
chair	spanner	baseball bat	pincers
truck	helicopter	cup	pen
gun	camel	banana	chicken
hoover	duck	piano	canopener
axe	plane	guitar	binoculars
lamp	bus	boot	plunger
turtle	toilet	horse	carriage
car	scissors	skateboard	dog
bike	crocodile	suitcase	pig
bed	snail	desk	toothbrush
		shoe	cannon
		fork	microscope
		sofa	spoon