9
Object Recognition: Attention and Dual Routes

Volker Thoma[1] and Jules Davidoff[2]

1 History of Hybrid Models

1.1 Introduction

The human capacity for visual object recognition is characterized by a number of properties that are jointly very challenging to explain. Recognition performance is highly sensitive to variations in viewpoint such as rotations in the picture plane (e.g., Murray 1995, 1998; Jolicoeur 1985) and to some rotations in depth (e.g., Hayward 1998; Lawson and Humphreys 1996, 1998) but invariant with the location of the image in the visual field (Biederman and Cooper 1991; Stankiewicz and Hummel 2002), the size of the image (Biederman and Cooper 1992; Stankiewicz and Hummel 2002), left-right (i.e., mirror) reflection (Biederman and Cooper 1991; Davidoff and Warrington 2001), and some rotations in depth (Biederman and Gerhardstein 1993). Second, object recognition is remarkably robust to variations in shape (Davidoff and Warrington 1999; Hummel 2001). For example, people spontaneously name the picture of a Collie or a Pomeranian both as simply a "dog" – a phenomenon termed "basic level" categorisation (Rosch et al. 1976).

Theorists traditionally struggle to account for these properties. In so called view-based theories (e.g., Olshausen et al. 1993; Poggio and Edelman 1990) representations mediating object recognition are usually based on metric templates derived from learned views. Although more recent accounts allow for combinations of template fragments (e.g., Edelman and Intrator 2003), the object features in view-based representations are fixed to certain locations in the image. Therefore, these accounts can readily explain effects of view-dependency in object recognition. In contrast, so-called structural description theories assume that the visual system extracts a more abstract representation from the 2D image on the

[1]School of Psychology, University of East London, Romford Road, London E15 4LZ, UK
[2]Psychology Department, Goldsmiths University of London, Lewisham Way, London SE14 6NW, UK

retina by encoding an object's constituent parts and their spatial relations (e.g., Biederman 1987; Hummel and Biederman 1992). Such a description is unaffected by many view-changes (such as changes in size, left-right reflection) and it also applies to many different exemplars of an object, permitting generalisation over metric variations of shapes (see Hummel 2001).

1.2 View Specific vs *Abstract Representations*

Not surprisingly, theorists have for some time sought to explain object recognition phenomena by integrating two qualitatively different types of representations. We will call these accounts hybrid models. For example, Posner and his colleagues (Posner 1969; Posner and Keele 1967) found an advantage for the sequential matching of identical letters in comparison with the matching of letters with the same name but differing case. However, this advantage was found only with short interstimulus intervals. These results were confirmed by other researchers with more realistic stimuli (Bartram 1976; Ellis et al. 1989; Lawson and Humphreys 1996) and were taken as evidence for the existence of a rapid, stimulus-specific representation and a more durable, abstract representation that generalises over variations in shape.

There is also neuropsychological evidence in support of representations that are either view-specific or more abstract. Warrington and her associates (Warrington and James 1988; Warrington and Taylor 1978) asked brain-damaged patients to recognize objects from canonical or non-canonical views. Observers with damage to the right posterior areas of the brain were particularly poor at non-canonical object recognition; therefore, Warrington and Taylor (1978) proposed that visual object recognition involves in two main stages. In the first stage, perceptual object constancy is achieved, relying 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 (Warrington and James 1988). There are more recent accounts based on such hemispheric differences in which an abstract-category recognition system is assumed to be dominant in the left brain hemisphere whereas a specific-exemplar subsystem is thought to be working more effectively in the right hemisphere (Marsolek 1999).

Somewhat different representations working in two parallel pathways were proposed by Humphreys and Riddoch (1984). Their patients were shown 3 photographs of objects. The task was to match two different views of a target object by discriminating the object from a visually similar distracter object. Four of their patients with right-hemisphere damage only showed impairment in this task when the principal axis of the target object was foreshortened in one of the photographs. In contrast, a 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 Humphreys and Riddoch (1984), this double dissociation indicates that two functionally independent routes are responsible for achieving object constancy. 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.

One particular shortcoming of these early hybrid 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. One type of attempt to clarify those conditions is to invoke process differences such as mental rotation (Jolicoeur 1990; Corballis 1988) or holistic *vs* analytic processing (Farah 1990, 1991). These will not be dealt with here but for a critical review, see Humphreys and Rumiati (1998) and Lawson (1999).

1.3 Representation Use according to Task-Demands

Tarr and Bulthoff (1995) 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. Shape discriminations usually fall within the extremes of the continuum and recognition is mediated by viewpoint-dependent and viewpoint-independent mechanisms according to the nature of the task, the similarity and familiarity of the stimuli, and other context conditions. Although this account seems intuitive, its predictions are rather general and the experimental evidence is somewhat unclear (Murray 1998; Hayward and Williams 2000).

2 A Hybrid Model of Object Recognition and Attention

2.1 The Hummel Model

Most of the previous hybrid accounts incorporate representations that have properties similar to structural descriptions (e.g., Hummel and Biederman 1992) or view-like representations (e.g., Olshausen et al. 1993). However, which type of representation is employed may depend on attention (Hummel and Biederman 1992). The next section will describe a hybrid account of object recognition that specifies how visual attention affects the representation of object shape.

The fact that both structural descriptions and view-based representations of shape can account for some, but not all of the properties of object recognition led Hummel (Hummel and Stankiewicz 1996; Hummel 2001) to propose that objects are recognized based on a hybrid representation of shape, consisting of

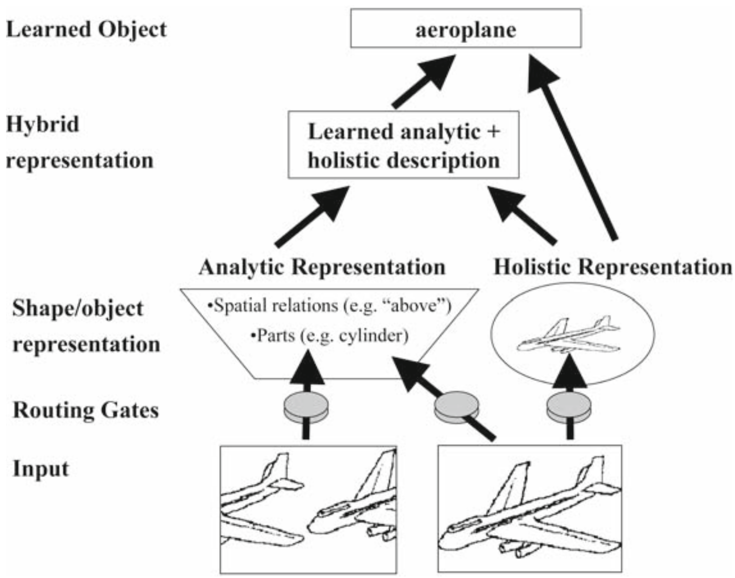

FIG. 1. A simple sketch of the architecture of JIM.3 (adapted from Hummel 2001). Units in the input layers of the model are activated by the contours from an object's line drawing. Routing gates propagate the output to units with two representational components: The independent units represent the shape attributes of an object's geons, and the units in the holistic map represent shape attributes of surfaces. The activation patterns of both components are learned individually, then summed in a higher layer over time. Units in the uppermost layer code object identity

a holistic (i.e., "view"-like) representation as well as an analytic representation (i.e., a structural description) of shape (Fig. 1).

Given a 2D image (such as a line-drawing) of an object, the Hummel model (JIM.3) generates both an analytic and a view-based representation. The analytic representation codes an object's shape in terms of the object's parts and their categorical interrelations. This representation has the properties of structural description (Biederman 1987) and is largely robust to many variations in viewpoint (such as translation, changes in scale, left-right reflection and some rotations in depth) but it is sensitive to rotations in the picture plane (see Hummel and Biederman 1992). The analytic representation allows generalization to novel views and to novel exemplars of known categories. However, it requires processing time and visual attention to be able to represent parts and spatial relations independently of each other (Hummel and Biederman 1992; Hummel 2001).

The holistic representation, in contrast, does not specify parts of an object or their categorical spatial relations. Instead, object parts are represented in terms of their topological positions in a 2-D coordinate system (see Hummel 2001). Since the holistic representation does not require attention for binding parts to their spatial relations, it can be generated rapidly and automatically. The representation formed on the holistic map is sensitive to left-right reflections as well as to rotations in the picture plane and in depth because the units representing object surfaces are spatially separated. However, the holistic representation is invariant with translation and scale.

2.2 Previous Tests of the Hummel Model

Stankiewicz et al. (1998) tested the predictions of the hybrid analytic/holistic model regarding changes in viewpoint using an object naming task with paired prime/probe trials. A prime trial consisted of a fixation cross followed by a box to the left or right of fixation, which served as an attentional cue (see Fig. 2 for

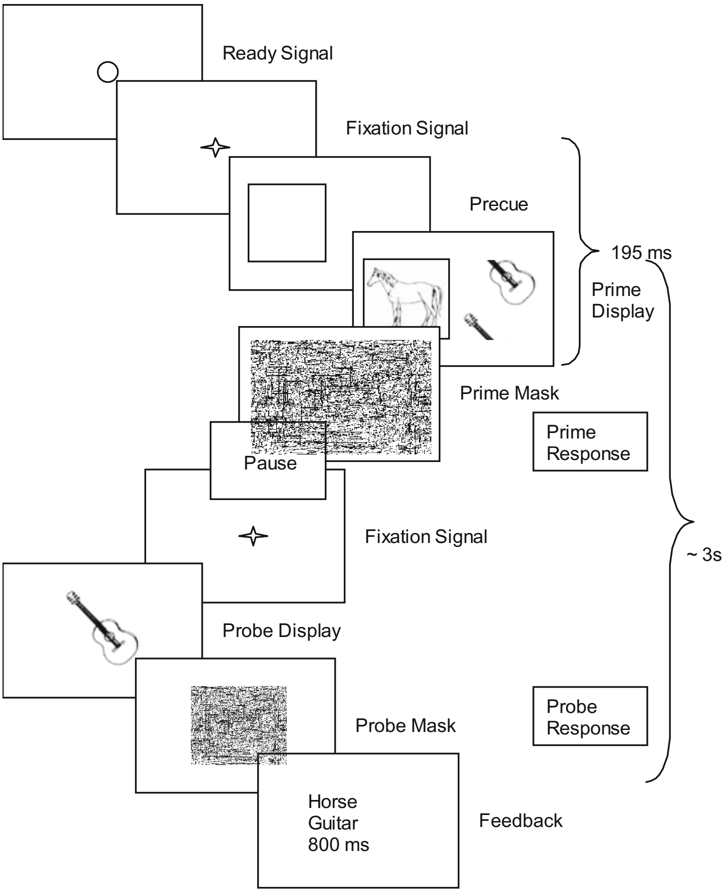

FIG. 2. Sequence of displays in a typical short-term priming paradigm (here an example from Experiment 1)

a similar paradigm used in Thoma et al. 2004). This was followed by two line drawings of common objects, one of which appeared inside the cueing box, and the other appeared on the other side of fixation. The participant's task was to immediately name only the cued image (the attended prime) and not respond to the other image (the ignored prime). The entire prime trial (from cueing box to mask) lasted only 195 ms, which is too brief to permit a saccade away from fixation. Each prime display was masked and after 2 seconds followed by a probe display containing a single image of an object at fixation. Again, the task was to name the object which was either the same object as the attended prime, the same object as the ignored prime, or an object the participant had not previously seen in the experiment (an unprimed probe, which served as a baseline to measure priming). Images of repeated objects (i.e., other than unprimed probes) were either identical to the corresponding primes, or were left-right reflections of them. Priming was measured as the difference in latencies between repeated (previously attended or ignored) and unrepeated (unprimed) probe images. The results showed that attended prime images reliably primed both themselves and their left-right reflections. However, ignored prime images only primed themselves in the same view. Moreover, the effects of attention (attended vs. ignored) and reflection (identical images vs. left-right reflections) were strictly additive: The priming advantage for same view prime-probe trials was equivalent in both attended and unattended conditions (about 50 ms). The fact that attention and reflection had additive effects on priming provides strong support for the independence of the holistic and structured representations of shape in the hybrid model. A holistic representation contributes to priming in a strictly view-dependent way and is independent of attention, whereas an analytic representation contributes to priming regardless of the view but depends on attention. Stankiewicz and Hummel (2002) tested the hybrid model's predictions concerning changes in position and scale using a similar paradigm as in Stankiewicz et al. (1998). As predicted, priming for attended and ignored objects was not affected by view changes such as translation and scaling (i.e., changes in position and size).

2.3 Testing Configural Distortions in the Hybrid Model

Here, we report 8 further experiments that examine aspects of the Hummel model using a priming paradigm similar to that employed by Stankiewicz et al. (1998). The findings of Stankiewicz and colleagues are clearly consistent with the hybrid model, but they cannot provide a direct test for the model's primary theoretical assertion – that the object shape is represented in a hybrid analytic and holistic fashion. To test the assumption of truly analytic representations underlying object recognition, we employed images that would not resemble any holistic representations. Whereas analytic representations of shape should be necessarily robust to configural distortions – such as scrambling of component parts – a holistic representation should be very sensitive to such image variations. Consider the manipulation of splitting an image down the middle and moving the left half of the image to the right-hand side (Fig. 1). A holistic representation of the intact aeroplane (e.g., stored as a view as in a typical image-based model; e.g., Poggio and Edelman 1990) would be matched, in its entirety, against an object's image to determine the degree of fit between the image and the holistic representation (i.e., view) in memory.

According to this holistic measure of similarity, the intact and split images of the aeroplane are very much different. However, a structural representation could compensate for this manipulation as long as the shapes of the object's parts are recoverable from the information presented in each half of the image (Biederman 1987; Hummel and Biederman 1992). In the split image, the front of the aeroplane is not connected to the back, yet the two halves retain enough structural information to allow the identification of the object.

Experiments 1–3 are from Thoma et al. (2004) and were designed to directly test the central theoretical assertion of the hybrid model that the representation of an attended image is analytic and holistic whereas the representation of an ignored image is only holistic. Experiment 1 investigated the role of attention in priming for split and intact object images. Participants named objects in pairs of prime-probe trials (as in Stankiewicz et al. 1998). Half of the prime images were presented intact, and half were split either horizontally or vertically, as illustrated in Figure 2. The factors of attention (attended *vs* ignored image) and image type (intact vs. split) were crossed orthogonally. The probe image was always intact and corresponded either to the attended prime, the ignored prime, or it was an image the observer had not previously seen in the experiment (which served as a baseline).

As predicted by the hybrid model, split images primed their intact counterparts only when the split images were attended, but both attended and ignored intact images primed their intact counterparts (see Fig. 3a). There was a reliable priming advantage for intact primes over split primes. Thus, the effects of attention (attended *vs* ignored) and configuration (intact vs. split) were strictly additive as in Stankiewicz et al. (1998).

Experiment 2 was designed to estimate what fraction of the priming observed in Experiment 1 was due to visual (as opposed to concept and/or name) priming. Images in the identical-image conditions of Experiment 1 (attended-intact, ignored-intact) were replaced with images of objects having the same basic-level name (e.g., "piano") as the corresponding probe object, but with a different shape (e.g., "grand piano" instead of "upright piano"). The results of Experiment 2 showed than an intact probe image was primed more (about 80 ms) by an attended split image of the same exemplar (e.g., a grand piano) than by an attended intact different exemplar of the same basic-level category (e.g., upright piano). Since in both cases participants responded with the same name in prime and probe trials, this difference indicates a strong visual component to the priming in the attended and ignored conditions. There was no priming for unattended primes (split or different exemplar), suggesting that all the priming observed in the unattended condition of Experiment 1 was specifically visual.

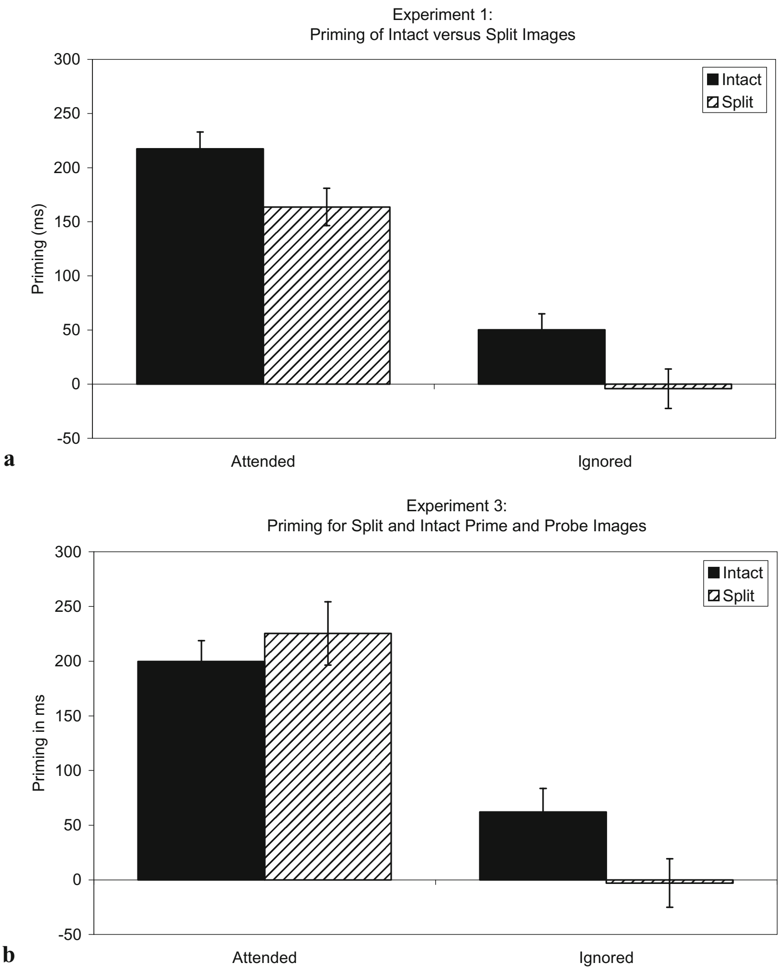

FIG. 3a,b. Priming (baseline RT minus RT in each experimental condition) means (ms) and standard errors in **a** Experiment 1 for intact probe images (Thoma et al. 2004) as a function of whether the prime image was attended or ignored and intact or split (n = 42). **b** Priming means in ms and standard errors for Experiment 3 (Thoma et al. 2004) as a function of whether the prime object was attended or ignored and whether both prime and probe were split or intact

The priming patterns in Experiments 1 and 2 are predicted by the theory that the visual system generates holistic representations of ignored images and analytic representations of attended images (Hummel 2001; Hummel and Stankiewicz 1996). However, an alternative interpretation is that all the observed priming resides in early visual representations (i.e., rather than in representations responsible for object recognition, as assumed by the hybrid model). Identical images may simply prime one another more than non-identical images, and attended images prime one another more than unattended images. If this "early priming" account is correct, then the advantage for identical images over non-identical images and the advantage for attended images over unattended images could yield the effects found in Experiment 1. This interpretation is challenged by the results of Stankiewicz and Hummel (2002), who showed that priming for ignored images is invariant with translation and scale. Nevertheless, a third experiment was designed to establish whether the results of Experiments 1 and 2 reflect a reliance on holistic processing for ignored images, as predicted by the hybrid model. Applied to the current paradigm, the logic is as follows: If the results of the first two experiments reflect the role of holistic representations in the recognition of ignored images, and if these holistic representations are encoded in LTM in an intact (rather than split) format, then ignoring a split image on one occasion should not prime recognition of the very same image on a subsequent occasion. However, if the results are due to priming early visual features (in both the attended and ignored cases), then ignoring a split image on one trial should prime recognition of that (split) image on the subsequent trial. By contrast, both models would predict that attending to a split image on one trial should permit the encoding and, therefore, priming of that image.

The results of Experiment 3 showed that a split image primed itself when attended but not when ignored, whereas an intact image primed itself under both conditions (see Fig. 3b). Critically, in the ignored conditions priming was found only for intact images but not for repeated split images. This 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 levels of priming under both ignored conditions.

The results reported by Thoma et al. (2004) strongly support the central theoretical tenet of the hybrid model of object recognition (Hummel 2001; Hummel and Stankiewicz 1996), that object recognition is based on a hybrid analytic + holistic representation of object shape. Attended intact, attended split and ignored intact images primed subsequent recognition of corresponding intact images, whereas ignored split images did not prime their intact counterparts. This pattern of effects is predicted by the hybrid account because attended images are represented both analytically and holistically, whereas ignored images are represented only holistically.

2.4 Plane Rotations

Object recognition is well-known to be sensitive to orientation in the picture plane (for a review, see Lawson 1999). The principal aim of Experiments 4 and 5 (Thoma, Davidoff, and Hummel 2007) was to test the hybrid model with picture plane rotations. A distinction between base (objects with a preferred upright position) and no-base objects (objects without a definite base) was made, which has previously been found to have importance for both behavioural (Vannucci and Viggiano 2000) and neuropsychological (Davidoff and Warrington 1999) investigations of object orientation. In a simple object naming study Thoma et al. (2007) confirmed the finding that objects with a definite base (e.g., a house) incurred increasing recognition performance costs when rotated, whereas no-base objects (e.g., hammer) were equally recognisable in all picture plane orientations. Subsequently, in Experiment 4, the pattern of priming effects observed for plane-rotated no-base objects clearly replicated the findings of Stankiewicz et al. (1998) with mirror-images and those of Thoma et al. (2004) with split images. Thus, the general notion of a hybrid model consisting of a holistic and analytic representation is supported by the fact that attended objects primed themselves in both the same view and the rotated view, whereas ignored objects only primed themselves in the same view (Hummel 2001).

In an attempt to test whether low-level early priming could have yielded these results, a replication of Experiment 3 was attempted using base-objects (e.g., a house). The relevant prime objects (attended or ignored) and the corresponding probe images were shown in the same orientation – both appeared in either an upright (familiar) or rotated (unfamiliar) view. The particular interest was in the ignored trials. Once more, Experiment 5 found a significant amount of priming in one condition (upright prime and identical probe image) and no priming in the other. Importantly, the lack of priming here was for ignored identical views that were unfamiliar (rotated view of base objects). Thus, the lack of priming in the ignored conditions for rotated no-base objects seen in Experiment 4 cannot be attributed to 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 in Experiment 5 (and previously with split objects, see Fig. 3b) is perhaps the most direct evidence that images of ignored objects achieve priming from access to stored familiar views. The data also fit previous findings that attention is necessary to establish view-independent representations (Murray 1995).

2.5 Depth Rotation

Experiments 6–8 (Thoma and Davidoff 2006) are concerned with depth rotations in the Hummel model. Just as with plane-rotations there are many documented effects of rotations in depth on recognition performance. Many researchers have shown view-dependent effects after depth rotations of familiar objects (e.g., Hayward 1998, Lawson and Humphreys 1996, 1998). However, Biederman and his colleagues (Biederman and Gerhardstein 1993) have obtained view-invariant effects after some rotations in depth that did not alter the visible part-structure of an object. The hybrid theory of object recognition may offer an explanation for mixed findings on depth-rotation effects.

Certain rotations in depth produce a mirror transformation of the image if the object is bilaterally symmetric. In Experiment 6, the findings of Stankiewicz et al. (1998) with mirror images were replicated with a new set of photorealistically rendered objects. Once more, the effects of attention and viewpoint were additive. Attended objects primed both themselves and their reflected versions, whereas ignored objects only primed themselves but not their mirror versions. Thus, the hybrid model may account for effects of depth rotations in which the part structure is not changed between views.

In contrast to mirror reflections, rotations in depth between study and test can affect the analytic representation because visible parts may be occluded or new parts may be revealed (Biederman and Gerhardstein 1993). Depth-rotations that differ from those akin to mirror-reflections should therefore provide an opportunity to further test the theory that two representations work in parallel because depth rotation may affect both representational components (analytic and holistic) instead of just one (holistic). The aim was to test whether depth-rotation involving part changes affects priming for attended objects (analytic plus holistic representation) more than for ignored objects (holistic representation only).

The logic underlying Experiment 7 comprises 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 the object should additionally reduce the amount of priming for attended images (because only then will the analytic representation be affected), but not for ignored images. In summary, if a part-based representation is involved for attended images but not for ignored ones, object rotations involving part changes should affect priming for attended images (holistic and analytic change) more than for ignored images (holistic change only).

In Experiment 7, objects were rotated in depth to produce an altered part-structure between views. To achieve a qualitative change in view orientation, objects were rotated in depth and depicted in two views. One was a complete side view (Fig. 4c) that would be primed by a more conventional view or vice versa (Fig. 4b). As a consequence, some parts of the object seen in one view (Fig. 4c) are not visible in the second view (e.g., the tail in Fig. 4b) and vice versa (e.g., legs in Fig. 4c). The effect of part-change was verified in a pilot study.

a,b **c**

Fig. 4a–c. Three views of an example object as used in Thoma and Davidoff (2006). 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 are hidden in view **c** whereas a new part (the tail) appears

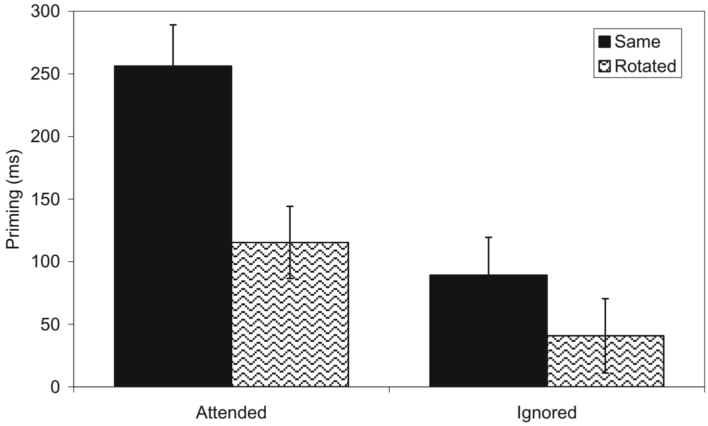

FIG. 5. Priming means (ms) and standard errors in Experiment 7 (Thoma and Davidoff 2006, Experiment 2) as a function of whether the object was attended or ignored in the prime display and whether the probe objects were presented in the same orientation or rotated in depth

The results of Experiment 7 replicated the previous findings of priming for attended images in the same view and in a changed (here: depth-rotated) orientation while ignored objects only primed themselves in the same view (see Fig. 5). Unlike previous tests of the hybrid model, the data show a unique interaction between attention and view-change: The difference between identical and depth-rotated views was significantly greater for attended than for ignored images.[1] This novel priming pattern is in line with the prediction of the hybrid model that depth-rotations may cause qualitative changes in analytic representations that depend on attention.

The data are not in line with view-based accounts. If attention plays a role in matching input with representations based on metric properties, one would expect enhanced priming effects for rotated objects in attended conditions relative to ignored conditions because attention would serve to aid the matching process (e.g., Olshausen et al. 1993). This was not the case here – the priming difference between rotated and identical view was greater for attended than for ignored objects.

As predicted from the hybrid model of Hummel (2001), viewpoint and attention produced additive effects of priming between qualitatively similar views, just as observed in Experiment 6. In Experiment 8, there was a greater degree of angular separation (90°) between the prime and probe view than in Experiment 7 (60°), yet the former view pairs (Fig. 4a,b) were rated by observers as more similar (in terms of visible parts) than the view pairs of Experiment 7 (Fig. 4b,c). Thus, the differences between the attended conditions of Experiments 7 and 8 confirm previous findings (Hayward 1998; Lawson 1999) that the amount of angular rotation (60° *vs* 90°) is not a reliable predictor of recognition performance as would be expected if object shape was represented only metrically.

The results for attended images also confirm that object recognition depends on whether the same parts are visible across views (Biederman and Gerhardstein 1993; Srinivas 1995). The hybrid's model general notion that object recognition across rotations in depth involves both an analytic and a holistic representation is also corroborated by Foster and Gilson (2002) who used novel 3-D objects that were to be discriminated in matching tasks either by a metric or a non-accidental (i.e., structural) property.

3 Multiple Representations in the Brain

The recent confirmation of the Hummel model from behavioural evidence finds support from neuroscience. Janssen et al. (2000) showed that neurons in the superior temporal sulcus were selective for three-dimensional shape whereas neurons in the lateral TE were generally unselective for 3D shape, though equally selective for 2D shape. Functional imaging studies (e.g., Vuilleumier et al. 2002) also support the notion that two types of object representations can be distinguished according to view-invariance in priming tasks. Vuilleumier et al. (2002) showed that repetition of images of common objects decreased activity (i.e., showed priming) in the left fusiform area independent of viewpoint (and size), whereas a viewpoint-dependent decrease in activation was found in the right fusiform area. Interestingly, the latter area was sensitive to changes in orientation but not in size – properties of the holistic component directly predicted by the hybrid model (Hummel 2001) and confirmed in behavioral studies (Stankiewicz and Hummel 2002).

As we have outlined in a previous section, numerous studies of patients with (limited) object agnosia indicate qualitatively different representations (Warrington and James 1988; Humphreys and Riddoch 1984). More recent evidence seems to corroborate the idea of multiple representations in the brain. Davidoff and Warrington (1999, 2001) studied patients who were extremely impaired at recognising object parts. Nevertheless, they were normal in naming intact objects though only when seen in familiar views. In terms of the hybrid model, the patients' holistic components seemed intact, allowing object recognition from familiar views, whereas analytic components were impaired preventing recognition of object parts or from unfamiliar views.

There is also neuropsychological evidence that attention may play a role in shape representation. Patients demonstrating unilateral neglect usually fail to respond to stimuli presented on the side contralateral to their lesion. Despite showing poor response to contralesional stimuli, there is evidence that these patients can nevertheless process semantic and shape properties in that field (Marshall and Halligan 1988; McGlinchey-Berroth et al. 1993). Recently, Forti and Humphreys (in press) have shown that the processing of shape information in the neglected hemifield depends on viewpoint as proposed by Stankiewicz et al. (1998) and seems qualitatively different from non-neglected stimuli. Similar findings come from studies on extinction, in which patients are able to detect ipsilesional stimuli presented alone but not when they are presented simultaneously with a stimulus on the contralesional side. Importantly, a recovery from extinction can be observed for global form information (Humphreys et al. 2000).

4 Conclusions

As we have seen, studies from different areas of cognitive science indicate the coexistence of multiple or hybrid representations of shape, resembling a distinction between holistic and analytic processing (Hummel 2001). This chapter has focused on the processing of shape in Hummel's model of object recognition because it is currently the most detailed model describing the role of attention in hybrid representation. Studies using traditional (rotation, reflection, scaling, translation, exemplar change) and novel (splitting) manipulations of object shape clearly confirmed the model's predictions regarding analytic/holistic representations. However, there are still many aspects of object recognition that are yet to be integrated into the model.

The hybrid model is largely based on a structural descriptive approach to object recognition (Hummel and Biederman 1992), which has been criticized in the past (e.g., Tarr and Bulthoff 1995; Edelman and Intrator 2003). For example, it is unclear how the model (and its predecessors) extracts axes of geons from 2D images. Another critique concerns the representation of irregular objects without obvious parts (such as a bush). One solution could be that in these cases, recognition relies more on the holistic component (Hummel 2003). A further way in which aspects of the Hummel model may be employed is to consider the role of time. For example, Zago et al. (2005) showed that visual priming for objects was maximal for an exposure time of 250 ms, then decreases. Therefore, they argued that certain aspects of an initial broad representation may be fine-tuned, becoming more stimulus specific.

In summary, it seems that attention is not necessary for object recognition but that the representations underlying object recognition differ according to whether an object is attended or not. An analytic representation is formed for attended objects and will be relatively robust to changes in view or configuration, except for part-changes. A holistic representation of an object is formed with and without attention allowing rapid recognition, but such a representation is very sensitive to any changes in view of global shape.

Note

1. The level of priming in Experiment 7 under all priming conditions was slightly higher than that in other experiments, and there was a slight trend toward positive priming for the ignored rotated prime. This may be due to the fact that the probe views were slightly less canonical (foreshortened) which produced longer identification times for baseline conditions (~50 ms compared to Experiment 6) and allowed more room for priming.

References

Bartram DJ (1976) Levels of coding in picture-picture comparison tasks. Mem Cognit 4:593–602

Biederman I (1987) Recognition-by-components: a theory of human image understanding. Psychol Rev 94:115–147

Biederman I, Cooper EE (1991) Evidence for complete translational and reflectional invariance in visual object priming. Perception 20:585–593

Biederman I, Cooper EE (1992) Size invariance in visual object priming. J Exp Psychol Hum Percept Perform 18:121–133

Biederman I, Gerhardstein PC (1993) Recognizing depth-rotated objects – evidence and conditions for 3-dimensional viewpoint invariance. J Exp Psychol Hum Percept Perform 19:1162–1182

Corballis MC (1988) Recognition of disoriented shapes. Psychol Rev 95:115–123

Davidoff J, Warrington EK (1999) The bare bones of object recognition: implications from a case of object recognition impairment. Neuropsychologia 37:279–292

Davidoff J, Warrington EK (2001) A particular difficulty in discriminating between mirror images. Neuropsychologia 39:1022–1036

Edelman S, Intrator N (2003) Towards structural systematicity in distributed statically bound visual representations. Cogn Sci 27:73–110

Ellis R, Allport DA, Humphreys GW, Collis J (1989) Varieties of object constancy. Q J Exp Psychol A 41A:775–796

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

Farah MJ (1991) Patterns of co-occurrence among the associative agnosias: implications for visual object representation. Cognit Neuropsychol 8:1–19

Forti S, Humphreys GW (in press) Representation of unseen objects in visual neglect: effects of view and object identity. Cognit Neuropsychol (in press)

Foster DH, Gilson SJ (2002) Recognizing novel three-dimensional objects by summing signals from parts and views. P Roy Soc Lond B Bio 269:1939–1947

Hayward WG (1998) Effects of outline shape in object recognition. J Exp Psychol Hum Percept Perform 24:427–440

Hayward WG, Williams P (2000) Viewpoint dependence and object discriminability. Psychol Sci 11:7–12

Hummel JE (2001) Complementary solutions to the binding problem in vision: implications for shape perception and object recognition. Vis Cogn 8:489–517

Hummel JE (2003) The complementary properties of holistic and analytic representations of object shape. In: Rhodes G, Peterson M (Eds) Analytic and holistic processes in the perception of faces objects and scenes. Greenwood, Westport CT, pp 212–234

Hummel JE, Biederman I (1992) Dynamic binding in a neural network for shape-recognition. Psychol Rev 99:480–517

Hummel JE, Stankiewicz BJ (1996) An architecture for rapid hierarchical structural description. In: Inui T, McClelland J (Eds) Attention and performance XVI: information integration in perception and communication. MIT Press, Cambridge MA, pp 93–121

Humphreys GW, Riddoch MJ (1984) Routes to object constancy – implications from neurological impairments of object constancy. Q J Exp Psychol A 36:385–415

Humphreys GW, Rumiati RI (1998) Agnosia without prosopagnosia or alexia: evidence for stored visual memories specific to objects. Cognit Neuropsychol 15:243–277

Humphreys GW, Cinel C, Wolfe JM, Olson A, Klempen N (2000) Fractionating the binding process: neuropsychological evidence distinguishing binding of form from binding of surface features. Vision Res 40:1569–1596

Janssen P, Vogels R, Orban GA (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. Mem Cognit 13:289–303

Jolicoeur P (1990) Identification of disoriented objects: a dual systems theory. Mind Lang 5:387–410

Lawson R (1999) Achieving visual object constancy across plane rotation and depth rotation. Acta Psychol 102:221–245

Lawson R, Humphreys GW (1996) View specificity in object processing: evidence from picture matching. J Exp Psychol Hum Percept Perform 22:395–416

Lawson R, Humphreys GW (1998) View-specific effects of depth rotation and foreshortening on the initial recognition and priming of familiar objects. Percept Psychophys 60:1052–1066

Marshall JC, Halligan PW (1988) Blindsight and insight in visuo-spatial neglect. Nature 336:766–767

Marsolek CJ (1999) Dissociable neural subsystems underlie abstract and specific object recognition. Psychol Sci 10:111–118

McGlinchey-Berroth R, Milberg W, Verfaellie M, Alexander M, Kilduff PT (1993) Semantic processing in the neglected visual field: evidence from a lexical decision task. Cognit Neuropsychol 10:79–108

Murray JE (1995) The role of attention in the shift from orientation-dependent to orientation-invariant identification of disoriented objects. Mem Cognit 23:49–58

Murray JE (1998) Is entry-level recognition viewpoint invariant or viewpoint dependent? Psychon B Rev 5:300–304

Olshausen B, Anderson C, Van Essen D (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J Neurosci 13:4700–4719

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

Posner MI (1969) Abstraction and the process of recognition. In: Spence JT, Bower G (Eds) The psychology of learning and motivation. Academic Press, New York, pp 43–100

Posner MI, Keele SW (1967) Decay of visual information from a single letter. Science 158:137–139

Rosch E, Mervis CB, Gray WD, Boyes-Braem P (1976) Basic objects in natural categories. Cognit Psychol 8:382–439

Srinivas K (1995) Representation of rotated objects in explicit and implicit memory. J Exp Psychol-Hum L 21:1019–1036

Stankiewicz BJ, Hummel JE (2002) Automatic priming for translation- and scale-invariant representations of object shape. Vis Cogn 9:719–739

Stankiewicz BJ, Hummel JE, Cooper EE (1998) The role of attention in priming for left-right reflections of object images: evidence for a dual representation of object shape. J Exp Psychol Hum Percept Perform 24:732–744

Tarr MJ, Bulthoff HH (1995) Is human object recognition better described by geon structural descriptions or by multiple views – comment on Biederman and Gerhardstein (1993). J Exp Psychol Hum Percept Perform 21:1494–1505

Thoma V, Davidoff J (2006) Priming for depth-rotated objects depends on attention and part changes. Exp Psychol 53:31–47

Thoma V, Hummel JE, Davidoff J (2004) Evidence for holistic representation of ignored images and analytic representation of attended images. J Exp Psychol Hum Percept Perform 30:257–267

Thoma V, Davidoff J, Hummel JE (2007) Priming of plane-rotated objects depends on attention and view familiarity. Vis Cogn 15:179–210

Vannucci M, Viggiano MP (2000) Category effects on the processing of plane-rotated objects. Perception 29:287–302

Vuilleumier P, Henson RN, Driver J, Dolan RJ (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat Neurosci 5:491–499

Warrington EK, James M (1988) Visual apperceptive agnosia – a clinico-anatomical study of 3 cases. Cortex 24:13–32

Warrington EK, Taylor AM (1978) Two categorical stages of object recognition. Perception 7:584–694

Zago L, Fenske MJ, Aminoff E, Bar M (2005) The rise and fall of priming: how visual exposure shapes cortical representations of objects. Cereb Cortex 15:1655–1665