INVITED ADDRESS

VIEWPOINT-DEPENDENT VERSUS INDEPENDENT 3-D OBJECT PERCEPTION: A DIRECT COMPARISON

Jan VANRIE and Johan WAGEMANS
University of Leuven

Although changing the viewpoint from which objects are perceived drastically alters the image they project on the retina, the visual system is generally able to recognize them. To explain this capacity, two basic accounts have been proposed: a viewpoint-independent and a viewpoint-dependent account. Both classes of theories have provided considerable evidence in their favor. In line with recent developments, however, we argue that a single mechanism does not suffice to explain the available experimental evidence. We report an experiment that shows that in a single experimental paradigm with highly similar stimuli, both viewpoint-dependent and viewpoint-independent results can be obtained. In addition, we review an fMRI-study in which we investigated the neurofunctional correlates underlying these behavioral patterns. Taken together, these findings corroborate the claim that there are multiple routes to achieve object recognition.

Introduction

The ability to recognize objects from different viewpoints is an impressive feature of our visual system: a bicycle seen from the front or from the side yields a completely different image, yet we easily recognize it as the same bicycle. Indeed, although a single object can yield almost an infinitive number of different retinal projections, we readily identify it as being the same object. To our subjective experience this process, which has been referred to as ‘visual object constancy’ (Lawson, 1999), proceeds in a seem-

---

This research was supported by research grants from the Research Program of the Fund for Scientific Research – Flanders (FWO G.0130.98) and the Research Council of the University of Leuven (IDO/98/002) to Johan Wagemans, and by a Belgian Psychological Society (BPS) award for the best undergraduate thesis in psychology to Jan Vanrie. We would like to thank the President and the members of the BPS committee, Axel Cleeremans who invited us to write the present paper and Wim Fias for helpful comments on a previous draft.

Correspondence concerning this article should be addressed to Jan Vanrie, University of Leuven, Department of Psychology, Tiensestraat 102, 3000 Leuven. Electronic mail may be sent to jan.vanrie@psy.kuleuven.ac.be.

1 Several factors can influence the final retinal projection of an object: lighting conditions, location in the visual field, distance (size of the object), ... In the present paper, however, we limit ourselves to changes in viewpoint.
ingly effortless manner. However, from a computational point of view, the issue is hardly trivial. For successful recognition to occur, a match has to be made between some sort of perceptual description of the retinal image and an object representation stored in memory. Changes in the viewpoint from which the object is perceived, however, can drastically alter this retinal image. The question then is how the visual system deals with these changes so that the object can still be recognized.

The above question has been investigated extensively by researchers in various domains, including cognitive psychology and psychophysics (e.g., Biederman & Bar, 1999; Bülthoff, Edelman, & Tarr, 1995; Hayward & Williams, 2000; Tarr, 1995; Willems & Wagemans, 2001), computer vision (e.g., Van Gool, Moons, Pauwels, & Oosterlinck, 1995; Van Gool, Moons, Pauwels, & Wagemans, 1994), single-cell recordings (e.g., Logothetis, & Sheinberg, 1996; Vogels, Biederman, Bar, & Lorincz, 2001), neuroimaging (e.g., Kosslyn et al., 1994), and neuropsychology (e.g., Davidoff & Warrington, 2000; Farah & Hammond, 1988; Humphreys & Riddoch, 1984). Similarly, several theoretical accounts have been put forward to explain how the visual system incorporates the factor 'viewpoint'. Although these accounts differ on numerous points, they can be roughly categorized in two major classes: viewpoint-independent and viewpoint-dependent accounts.

Viewpoint-independency. To deal with changes in viewpoint, the so-called 'viewpoint-independent' accounts (e.g., Biederman, 1987; Marr, 1982) propose that the visual system constructs a perceptual description of the retinal image, which is independent of the viewpoint. That is, from the information available in the retinal image (which is obviously viewpoint-dependent), the system derives a structural representation in an "object-centered" frame of reference. For example, the description of a bicycle in an object-centered frame of reference might include that "the handlebars are located above the front wheel". Now, suppose the bicycle is turned upside-down (e.g., to repair it). From the perspective of an observer, that is, described in the observer's frame of reference, the handlebars will now be below the wheel. However, for the object-centered description nothing changes: in the reference frame of the object itself, the handlebars will still be located above the wheel. Because the object representations stored in memory are assumed to be in the same object-centered reference frame, it is straightforward to find a correspondence for the viewpoint-independent perceptual description.

According to these accounts, from (nearly) all views of a single object the same structural description will be constructed. A major prediction that follows from this assumption is then that the viewpoint from which the object is seen, does not alter the recognition process.
Viewpoint-dependency. The basic idea in the class of theories described as ‘viewpoint-dependent’, is that the recognition process employs features in the incoming image, which are represented in a viewpoint-specific frame of reference (Tarr, Bulthoff, Zabiniski, & Blanz, 1997). As a result the actual observer’s perspective will exert an influence on the recognition process. How this viewpoint-dependency is actually implemented varies quite strongly (see Willems & Wagemans, 2001), but all theories endorse the importance of a viewer-centered frame of reference.

A consequence of the use of a viewpoint-specific reference frame is that sometimes these frames may differ for the perceptual descriptions and the representations in memory, making a simple comparison impossible. Typically, to solve this discrepancy some sort of transformation is assumed to “align” the frames.

Predictions and experimental evidence. As mentioned, there has already been considerable research into the topic of object perception from different viewpoints. A frequently used paradigm in this research is the same-different paradigm in which observers judge whether or not two objects are identical, regardless of their orientation. By systematically manipulating the angular difference (AD) between the two objects, predictions of the two classes of theories can be contrasted. More specifically, viewpoint-independent accounts generally predict that varying the viewpoint will not affect recognition performance, while the viewpoint-dependent accounts do predict an effect.

Somewhat surprisingly, both classes of theories have provided substantial experimental evidence in their favor (for reviews, see Tarr, 1995; Vecera, 1998). In turn, this has lead to a debate on the validity of both the methods and the stimuli used in research on object perception. For example, Biederman and Gerhardstein (1993) have proposed a number of properties that objects need to have in order to qualify as appropriate stimuli. As a reply

---

2 In the most extreme case of viewpoint-dependency, i.e., that we store all views of an object in memory, no such transformation would be needed. However, this is clearly not a feasible strategy: an unlimited storage capacity would be needed to keep in memory all the views of every object we see. Moreover, such a mechanism makes the highly counterintuitive prediction that we would not be able to recognize an object when we look at it from a new perspective. Not surprisingly, this prediction has been refuted by experimental evidence that shows that objects can indeed be recognized from previously unseen viewpoints (Tarr, 1995).

3 Although in general the predictions of the two classes are straightforward (effect of AD or not), a note of caution is needed. The prediction of an absence of an effect of AD for the viewpoint-independent accounts presupposes that the time needed to extract a viewpoint-independent representation is indeed independent of the specific viewpoint in which the object is perceived. However, there is experimental evidence that the extraction of invariant features need not be invariant itself (Wagemans, Van Gool, & Lamote, 1996).
to an experiment in which wire-frames yielded effects of viewpoint (Edelman & Bülthoff, 1992), Biederman and Gerhardtstein (1993, experiment 5) inserted a 'geon' into the wire-frames. With these stimuli, participants' performance was clearly independent of the viewpoints in which the objects were shown. On the other hand, Tarr, Williams, Hayward and Gauthier (1998) claimed that even when recognizing a single geon, viewpoint-dependent effects can be observed.

In short, it seems that it is becoming more and more difficult to uphold the view of the visual system as operating in a rigid, fixed manner. For this reason, the idea of object perception as a continuum ranging from view-dependency to view-independency (e.g., Edelman, 1995; Hayward & Williams, 2000; Tjan & Legge, 1998), or the notion of multiple routes to object perception (Lawson, 1999), is receiving increased attention. The experimental work presented here was the first step in a project set up to disentangle two recognition processes, not only on the behavioral but also on the neurofunctional level. More specifically, one process that is associated with the viewpoint-dependent side, and another, which is associated with the viewpoint-independent side.

Experiment

Our aim was not to map all processes referred to as "viewpoint-dependent and viewpoint-independent perception". The ultimate goal of the project was to dissociate a viewpoint-dependent process from a viewpoint-independent process 1) on the behavioral level, and 2) on the neurofunctional level. Importantly, in order to dissociate these processes we wanted to use a single experimental paradigm and a highly similar stimulus set. As an example of well-defined viewpoint-dependency, we opted for the process of mental rotation. For the viewpoint-independent process participants were able to make use of invariant features.

The process of mental rotation was first described by Shepard and Metzler (1971) who demonstrated that when subjects performed a same-different judgment with mirror objects, reaction times increase linearly with increasing angular differences between the objects. This classical finding has been replicated in many studies (e.g., Corballis, 1988; Kosslyn, 1994). Although mental rotation might not be a typical object recognition task (see Perrett, Oram, & Ashbridge, 1998; Willems & Wagemans, 2001), it is clearly an extreme example of viewpoint-dependency: in order to discriminate an object from a mirrored version, a viewer-centered frame of reference is needed. This is even acknowledged by viewpoint-independent accounts, because their object-centered descriptions do not generally code the differ-
ence between an object and its mirrored version (Biederman & Gerhardstein, 1993; Takano, 1989). A clear and perhaps equally extreme example of viewpoint-independency is the case in which an observer uses a viewpoint-invariant distinctive feature. The observer thus determines the identity of the object by extracting a so-called “invariant”, i.e., a feature that is not altered by changes in perspective (for a more elaborate discussion, see Van Gool et al., 1994). In this case, changing the orientation of the object will not result in changes in performance (cf. Eley, 1982).

To test this dissociation-hypothesis, participants performed a same-different judgment task with block-like objects, in which the angular difference between the comparison objects was manipulated. In the Rotation condition, the objects were either the same, yielding match trials, or they were mirrored versions, yielding non-match trials (because no matter how one rotates the objects, it is impossible to align all parts of an object and its mirrored counterpart). This condition was predicted to yield clear viewpoint-dependent results. In the Invariance condition, the objects differed in a viewpoint-invariant feature, namely the angles between the constituent parts of the object (see Method section for details). In this case, the difference between the comparison objects was an invariant feature: the angles between the object-components could be orthogonal (90°) or skewed. We predicted that participants would use this feature to solve the task, so performance would be independent of the angular difference between the objects.

In addition to being a first test of the dissociation-hypothesis, the present experiment served two somewhat related goals. First, it was aimed at validating the “viewpoint-independent” status of the Invariance condition. If performance in our Invariance condition is indeed independent of the magnitude of the angular difference, it should be possible to in- or decrease overall difficulty without introducing effects of AD. By changing the amount of skewing (see Method section for details), we were able to manipulate the salience of the invariant feature and thus the difficulty of the matching task in the Invariance condition. The second reason for using different amounts of skewing was to examine which amount of skewing would yield the most equivalent performance between the Rotation and the Invariance condition. Changing the amount of skewing will probably have an impact on the level of performance. If we achieve a comparable level of performance between the two conditions, this will make the interpretation of the fMRI results more

4 We want to emphasize that it is the ensemble of the three angles that constitutes the invariant feature. Willems and Wagemans (2000) demonstrated that the observation of a single angle from an arbitrary viewpoint yields considerable perceptual ambiguity. By simultaneously skewing all three angles, however, the stimulus does allow for a veridical interpretation over all viewpoints.
straightforward. However, because we can arbitrarily change the skewing size, the most appropriate amount is an empirical question.

Methods

Subjects. Eight subjects (six male, two female) took part in the experiment. All of them had normal or corrected-to-normal vision and all but one were naive with respect to the purpose of the experiment.

Stimuli. Stimuli were 2-D perspective drawings of a set of 36 objects (see Figure 1). The objects consisted of four components: a main limb and three smaller limbs of different sizes. The smaller limbs were attached face-to-face onto the main limb. We systematically placed the side-components in such a way that there would always be two side-components in the same plane but on opposite sides of the main limb. This procedure resulted in a standard set of 9 front/back and left/right asymmetrical, solid objects.

In addition to the standard set, a specific set of objects was constructed for each separate condition. For the Rotation condition, we relocated the two co-planar side-components of each object to the opposite side of its main limb (e.g., Figure 1, lower left). This yielded a set of 9 mirrored versions of the standard objects. For the Invariance task, the angle of attachment of the side-components to the main limb was altered. Whereas the face-to-face connection in the standard objects resulted in orthogonal angles, the attachment of a smaller limb was now no longer orthogonal. With the main limb in the vertical orientation, the side-components were tilted up- and downward resulting in a “skewed” object. Two versions of skewed objects were constructed, which only differed in the amount of skewing: 5° in one version and 10° (e.g., Figure 1, upper right) in a second version. The direction of tilt (up- or downward) was random but had to be opposite for the two limbs in the same plane. The relatively small but qualitative difference was thus the angle formed by a small limb and the main limb: this was 90° (orthogonal) in one set of objects and either 85°/95° or 80°/100° (skewed) in the other set of objects.

Of each of the 36 objects (i.e., 9 standard, 9 “mirrored”, 9 “5°-skewed” and 9 “10°-skewed” objects), twelve different images were produced. The objects, with the main limb oriented vertically, were first rotated 45° in the plane in a counterclockwise direction and then slanted 45° backwards. For every object, the view obtained in this way was taken as the standard view (i.e., the “0° view”). Rotating the object around the length-axis of the main limb in steps of 30° generated twelve different views of each object, which represented the actual stimuli. These were copper-colored against a black
background and subtended approximately 3.30° x 3.10° of visual angle. Two views were presented simultaneously, next to one another, separated by 7.40° of visual angle on a 17 inch SVGA CRT with a spatial resolution of 1024 x 768 and a temporal resolution of 75 Hz. In the Rotation condition, these views were either from the same object (match trials, Figure 1 upper left) or from an object and its mirrored version (non-match trials, Figure 1 lower left). In the Invariance condition, the views were either from the same object (match trials, Figure 1 upper right for two views of the same skewed object) or they were from an object with different angles between the parts (non-match trials, Figure 1 lower left, the left object is skewed, the right object orthogonal).

Design. All subjects performed two conditions: the Rotation and the Invariance condition. Two groups of four subjects were formed. The 5°-group received the 5°-version in the Invariance condition, the 10°-group received the 10°-version. In each condition the images of 18 objects were used: the standard set of 9 objects plus the 9 condition-specific objects. Of all 18 objects, one of the 12 viewpoints was selected that would appear in all trials in which the object in question was used. The selected view was then combined with each of the 12 views of the same object, resulting in 12 “match” trials with angular differences (ADs) between the two presented views of 0°, 30°, 60°, 90°, 120°, 150° and 180°. For the “non-match” trials, the selected view of each of the 18 objects was combined with each of the 12 views of the corresponding object in the other set. So, a non-match trial in the Rotation condition consisted of a view of an object and a view of its mirrored counterpart (e.g., Figure 1, lower left). In the Invariance condition, the two objects were either both orthogonal or both skewed (e.g., Figure 1, upper right) in case of a match trial, whereas for a non-match trial they differed in the angle of attachment of their side-components, that is, one was orthogonal and the other skewed (e.g., Figure 1, lower right). So, the objects in a non-match trial could only differ in one feature: the location of the side-components in the Rotation condition or the angle of attachment to the main limb in the Invariance condition.

In half of the 12 trials, the selected viewpoint was presented as the left object. The angular difference between the two viewpoints varied from 0° to 180° in steps of 30°. In sum, for each condition 18 (objects) x 2 (match) x 12 (views) = 432 trials were presented. The order of trials was randomized independently for each subject.

Procedure. All subjects performed two separate sessions, on two separate occasions, in which one condition was presented. Half of the subjects in each group started with the Rotation condition, the other half with the
Figure 1. Example of a match trial and a non-match trial in the Rotation condition (left column) and an example of a match trial (both objects are skewed) and a non-match trial (left object is skewed, right object is orthogonal) in the Invariance condition (right column) with the 10°-version.
Figure 2. Mean response time (A) and proportion correct responses (B) on match trials per angular difference (AD) for both the Rotation and the Invariance condition for the two Groups (5° and 10°).
Invariance condition. In both conditions, a simultaneous, same-different judgment paradigm was used.

Subjects were seated in front of the CRT at about 80 cm, in a dimly lit room. Prior to each session subjects were informed about the nature of the stimuli, the way in which they could differ and the task they had to perform. After some practice trials, the actual experiment began. Two views were presented next to one another (see Figure 1), subjects judged whether the objects were the same (M-key) or different (Z-key), regardless of possible rotations in depth. Stimuli were visible until subjects responded, after which a 500 ms black screen was shown, followed by the display of a new stimulus pair. The presentation of the stimuli was controlled by a program developed in Superlab Pro (Cedrus, Corp 1997) that also recorded the response and response time. Subjects were instructed to respond as quickly and accurately as possible.

Results

Prior to the analysis 1.4 % of the data were removed as outliers, i.e., trials with a response time larger than the mean response time plus three standard deviations (calculated per subject for each condition separately).

Response times. The analysis was performed only on trials in which the subject responded correctly. For each subject, response time means per angular difference for each of the two conditions were calculated, differentiating between matches and non-matches. The data were analyzed separately per Group (with 2 levels: one group of subjects who received the 5°-objects in the Invariance condition and a second group who received the 10°-objects) through an analysis of variance (ANOVA) having as within-subjects variables: Process (Rotation or Invariance), AD (0°, 30°, 60°, 90°, 120°, 150° and 180°) and Match (same or different).

Although it did not reach significance in either group, the effect of Process was clearly different for the two groups: Rotation was performed faster in the 5°-group (2856 ms vs. 4067 ms), but slower in the 10°-group (3184 ms vs. 2339 ms). The interaction between Process and AD was significant for the 10°-group ($F(6,18) = 4.17, p < .01$), but not for the 5°-group ($F(6,18) < 1$). Most importantly, the Process x AD x Match interaction was significant in both groups ($F(6,18) = 3.56, p < .02$ and $F(6,18) = 9.93, p < .001$ for the 5° and 10° group, respectively). Mean response times for the match trials can be seen in Figure 2A. Post-hoc comparisons (Schéffé) indicated that for match trials in the two Invariance conditions no differences were found between the different angular differences.
Because the Rotation condition was identical in the two groups, we performed an ANOVA on the results of the Rotation conditions with Group (5° and 10°) as between-subjects variable and AD and Match as within-subjects variables, to check for significant differences between the 5°- and the 10°-group. The analysis indicated performance in the two groups did not differ (F(1,6) < 1). The failure to find a significant three-way interaction Group x Match x AD, F(6,36) < 1, also indicated that the Match x AD interaction was not different.

**Accuracy.** Per subject, the mean proportion of correct responses per angular difference for both matches and non-matches was calculated for both conditions. The analysis was as for the response times.

Compared to the Invariance condition, accuracy in the Rotation condition was not significantly higher for the 10°-group (.96 vs. .87), but clearly so for the 5°-group (.92 vs. .68, F(1,3) = 30.51, p = .02). The Process x AD interaction failed to reach significance in both groups. The three-way interaction, finally, was only significant for the 5°-group (F(6,18) = 7.93, p = .001). This is, however, most likely the result of the huge difference in performance for the 0°-AD between matches (.96) and non-matches (.40) in the Invariance condition. Figure 2B shows the mean accuracy for the match trials.

Again the performance of the two groups on the Rotation condition was not significantly different: no main effect of Group, F(1,6) = 3.21, p = .12, nor a significant Group x Match x AD interaction was found, F(6,36) < 1).

**Discussion**

Two groups performed a Rotation condition and an Invariance condition. Performance was predicted to be dependent on changes in viewpoint in the former condition, but independent in the latter. Moreover, one group performed the Invariance condition with 5°-objects, the other group with 10°-objects. If the experimental manipulation (skewing) used in the Invariance condition is indeed an "invariant feature" for the visual system, changing the amount of skewing would not lead to a differential effect of viewpoint in the two groups. The results seem to indicate that this is indeed the case. Although changing the amount of skewing influenced the mean response times and accuracy, it did not fundamentally change the effects of viewpoint.

A second goal of the present experiment was to find the most suitable amount of skewing. Ideally, the two conditions would demonstrate an equivalent level of performance in terms of accuracy and mean response times. It appeared that the mental rotation task was performed slower than the Invariance task with the 10°-version (+845 ms), but faster than with the 5°-
condition (-1211 ms). The proportion correct responses was lower for both Invariance conditions compared to the Rotation condition. For the 10°-version, this difference (7%) was non-significant. For the 5°-version, however, accuracy was much lower, namely 24%.

Given a simple linear relation between size of skewing and the behavioral measure, an angle of intermediate size (7.9°) would be the most appropriate in terms of speed of performance. However, in terms of accuracy, an even larger angle (12.1°) would give the best fit. As it turns out, the equilibrium in this trade-off between speed of performance and accuracy seems to be located exactly on the 10° skewing angle.

In sum, the conclusion from the present experiment is threefold: i) the two conditions seem to demonstrate a differential effect of AD, ii) changing the overall difficulty of the Invariance condition does not influence effects of AD, and iii) the most appropriate amount of skewing for the Invariance condition is 10°.

On the basis of these conclusions and after a follow-up study in which the behavioral properties of the two conditions have been investigated in more detail (Vanrie, Willems, & Wagemans, 2001), we set up an fMRI-study in order to investigate the neural correlates of these processes. So, in what follows we will summarize the main findings of this fMRI-study in which participants performed the two tasks described here (in a slightly modified version). This experiment constituted the second major step in the project set up to dissociate the two processes. The original experiment and results have been described elsewhere (Vanrie, Béatse, Wagemans, Sunaert, & Van Hecke, 2001).

fMRI – Experiment

The aim of the neuroimaging experiment was to investigate whether different routes to object perception were also dissociable on the neurofunctional level. Six participants performed 4 time series in which the Rotation condition and the Invariance condition (with the 10°-objects) were presented in alternating blocks and 4 time series in which one of the experimental conditions was alternated with a control condition. To investigate the underlying cortical activation patterns, we measured cerebral blood flow in a 1.5 Tesla MR scanner. The results can be summarized in three major points.

First, the neurofunctional data showed that in comparison with the control condition all the areas that were activated in the Invariance condition were also activated in the Rotation condition. Most important in these common activations was the ventral, occipito-temporal stream with a bilateral focus in the fusiform gyrus. Second, compared to the control condition, we
also observed a shift in the main point of activation. More specifically, in the Invariance condition occipito-temporal areas were more strongly activated than parietal areas, while the reverse pattern was found for the Rotation condition. Third, a direct comparison of the two experimental conditions revealed that a number of cortical areas were either uniquely or more strongly activated by the Rotation condition. The differentially activated regions included amongst others the lateral and inferior part of the intraparietal sulcus (IPS) in both hemispheres, the superior parietal lobes, and an extensive cluster in BA6 containing the frontal eye fields and the superior frontal sulcus.

A detailed description and interpretation of the result is beyond the scope of the present study (see Vanrie, Béatrice et al., 2001). However, the differential activation patterns for the two conditions are clearly compatible with the dissociation-hypothesis and the experimental results here presented. Moreover, the neurofunctional results can easily be linked to the functional dissociation between the dorsal and the ventral pathway described by Milner and Goodale (1995). In their account, the ventral stream (going from the primary visual cortex V1 to the inferior temporal cortex) processes information about the features of an object in a viewpoint-independent reference frame in order to recognize or categorize an object. On the other hand, the dorsal stream (going from V1 to the parietal cortex) comes into play when the goal of the visual analysis is to program and control actions towards these objects. In this case, the information is thought to be encoded in viewer-centered coordinates. In general, the results of the fMRI experiment thus seem to corroborate this coupling between viewpoint-independency and the ventral stream (cf. the Invariance condition), as well as that between viewpoint-independency and the dorsal stream (cf. the Rotation condition).

Conclusion

Notions such as ‘multiple routes to object perception’ are receiving increasing support in recent years. The present experiment, taken together with the differences in cortical activation patterns observed in the MRI-study, are fully compatible with this notion. Our visual system seems to operate in variable ways, depending on the input (what is perceived) and the goal (task). For this reason, we believe that it would be more rewarding for future research to focus not so much on whether recognition is viewpoint-dependent or independent, but rather on the circumstances in which it is one or the other.
References


Received July, 2001
Revised and accepted September, 2001