



Technical Report No. 42

November 1996

What object attributes determine canonical views?

Volker Blanz, Michael J. Tarr, Heinrich H. Bülthoff, Thomas Vetter

Abstract

We investigated preferred or canonical views for familiar and novel three-dimensional objects using computer-graphics psychophysics. We assessed the canonical views for objects by allowing participants to actively rotate realistically shaded three-dimensional models in real-time. Objects were viewed on a Silicon Graphics Workstation and manipulated in virtual space using a three degree-of-freedom input device. In the first experiment, participants adjusted each object to the viewpoint from which they would take a photograph if they planned to use the object to illustrate a brochure. In the second experiment, participants mentally imaged each object based on the name and then adjusted the object to the viewpoint from which they imagined it. In both experiments, there was a large degree of consistency across participants in terms of the preferred view for a given object. Our results provide new insights on the geometrical, experiential, and functional attributes that determine canonical views under ecological conditions.

VB was supported by a graduate student stipend from the Max-Planck Society. MJT was supported by a grant from the Office of Naval Research, contract number N00014-93-1-0305, and a Max-Planck Society visitors' stipend.

This document is available as `/pub/mpi-memos/TR-042.ps` via anonymous ftp from `ftp.mpik-tueb.mpg.de` or from the World Wide Web, <http://www.mpik-tueb.mpg.de/projects/TechReport/list.html>.

Introduction

Seen from different directions, three-dimensional objects may look significantly different. Even though our visual perception is highly robust with respect to changes of viewpoint, we are more facile with certain views relative to others. This is demonstrated by the surprise that some people experience when they see a photograph of a familiar item taken from an unusual direction. Indeed, many an adolescent has been quite shocked when they first view their own profile in a mirror.

Previous research has shown that depending on the criteria applied, views of three-dimensional objects may be privileged or canonical in different senses:

- The viewpoint that is assigned the highest “goodness” rating by participants.
- The viewpoint that is first imagined in visual imagery.
- The viewpoint that is subjectively selected as the “best” photograph taken with a camera.
- The viewpoint found to have the lowest response time and error rate in recognition or naming experiments.
- The viewpoint inspected for the longest period of time in a free exploration task.

While there has been a great deal of interest in studying the recognition of novel objects through viewpoint generalization (see Bühlhoff, Edelman, & Tarr, 1995), as well as generalization across other types of variation, we still know little about the recognition of natural objects under ecological conditions. Here we attempt to address this issue by using common objects manipulated under somewhat more realistic conditions. In particular, we are interested in what canonical views tell us about the representation and recognition of objects and, more specifically, what attributes contribute to canonical views. We present two experiments, both of which assess observers’ canonical views using a three degree-of-freedom input device that allows real-time interaction with computer-graphics models of common objects. In Experiment 1 participants adjusted each object to the viewpoint from which they would take a photograph to illustrate a brochure and in Experiment 2 participants adjusted each object to the viewpoint from which they had mentally imaged each object. We were interested in: 1) the degree of consistency in the canonical view selected for a given object between different observers; 2) the degree of consistency between the “photograph” and “imagery”

tasks; and, 3) the attributes of the objects that contribute to canonical views in each experiment. We begin by reviewing what has been previously learned about canonical views.

Previous experiments on canonical views

The term “canonical views” was first used by Palmer, Rosch, and Chase (1981). In a series of experiments, they investigated the first four of the criteria listed above for a set of 15 familiar objects. Their first experiment used a goodness rating in which participants ranked 12 photographs of each object taken from different viewpoints. Objects were shown in their front, back, side, and top views, as well as intermediate views at 45° angles to these.

Their second experiment used a visual imagery task in which participants were given the name of each object and were then asked to form a mental image of the object. The specific viewpoint was measured by having participants describe their mental image in terms of the proportions of visible surfaces (i.e., front, back, side, top, bottom).

Their third experiment used an active selection task in which participants used a camera to take a photograph of their preferred viewpoint for each object. Independent raters then judged the proportions of surfaces visible in these pictures in order to allow a comparison with the results of the first and second experiments.

Palmer et al.’s (1981) analyses of these experiments indicated that participants consistently preferred the same viewpoints independent of the task. Specifically, participants preferred off-axis views – that is, viewpoints, such as a three-quarters view, that make a large number of surfaces visible. Their suggestion is that the same type of views should be preferred in all tasks. To confirm this, in a fourth experiment, they performed a naming task in which participants attempted to verbally name objects shown from the best, the worst, and two intermediate viewpoints as determined in the earlier experiments. Here they found that naming times decreased monotonically with increasing canonicity of the viewpoint. Thus, they obtained converging evidence for the canonicity of off-axis views.

In a more recent study of canonical views, Verfaillie and Boutsen (1995) used a paired comparison paradigm to obtain goodness ratings. Rather than photographs of real objects, they used a set of computer generated images created from three-dimensional models. They used up to 11 different viewpoints for each of 70 familiar objects. For

a given object they presented participants with all pairwise combinations of viewpoints and asked them to select the subjectively better viewpoint for the pair. Using these pairwise preferences they calculated viewpoint rankings for each object using Thurstonian scaling. An exploratory cluster analysis of the ranks grouped the set of objects into 8 clusters, and cluster membership could be explained in terms of coarse physical structure, for instance elongated front-back axes. Overall, they found that the three-quarter viewpoints were ranked the highest – a finding generally consistent with the results of Palmer et al. (1981).

One possible problem with both of these studies is that canonical views were assessed using a fixed set of viewpoints. In contrast, Perrett, Harries, and Looker (1992) used a real physical object that was free to rotate in all directions. The object was a wooden model of an arbitrary machined tool part. Participants viewed the object indirectly on a video screen that was updated continuously with their actions. Participants were instructed to rotate the object to a viewpoint that they considered to be “structurally most informative.” In this task most participants preferred off-axis views. However, when participants were first asked to close their eyes, then imagine the object, and then finally rotate the object to the viewpoint best matching their mental image, the majority of participants preferred plan views. In such views, the line of sight is orthogonal to a prominent face of the object, and most of the three-dimensional structure of the object is occluded. This latter result is inconsistent with the findings of Palmer et al. (1981), raising the question whether results for novel objects generalize to familiar objects. Perrett et al. (1992) speculated that this inconsistency might have been due to: 1) participants’ relatively brief experience with the novel object; combined with, 2) participants’ inability to imagine unfamiliar objects with a higher number of components under perspective distortion (although the test object was actually less complex relative to many familiar objects).

Edelman and Bülthoff (1992) also assessed the canonical views of novel three-dimensional objects (“paper-clips” and “amoeba”) – however, rather than rating or adjustment methods they measured canonicity using a recognition task. In a training phase, participants observed an object as it rotated through a wide range of viewpoints designed to provide good coverage of the viewing sphere. In a subsequent test phase, participants recognized the previously learned objects from static

images. Although all test images were seen in the previously presented motion sequence, response times and error rates were different for different views (and consistent across participants). In a second session, differences between test views diminished. Importantly, this pattern demonstrated that canonical views may be found in instances where all views were seen equally often – in such cases a preference for some views over others cannot be attributed to prior exposure. Although Edelman and Bülthoff (1992) concluded that these differences were due to intrinsic geometrical properties of the objects, a more recent recognition experiment by Cutzu and Edelman (1994) suggests that there may be other reasons for preferences for certain views. Specifically, they found that the preferred views of paperclips were not the same for all participants. Views that were recognized very accurately by some participants were difficult for others. They speculated that the reason for this was that the paperclips were quite complex and participants adopted different coding strategies for the same paperclip, for example, a Hebrew character or a star. Such differences in coding may have produced different regions where the iconic resemblance remained stable and, consequently, different maximally diagnostic views.

Finally, measurements of preferred views were obtained by Perrett and Harries (1988) using potatoes and real models of tetrahedra. To assess canonicity they allowed participants to freely explore each object as it rotated around the vertical axis and recorded the amount of time participants spent inspecting each view. Prior to exploration, participants were told that they should study each object for a subsequent recognition test. Similar experiments were run using clay heads (Harries, Perrett, & Lavender, 1991) and using the same machine tool part mentioned earlier (which was free to rotate in all directions; Perrett et al., 1992). For potatoes, they found that participants preferred views where the principal axis was aligned with or orthogonal to the line of sight. For the tetrahedra and the machined tool part, they found that participants preferred views aligned with the lines of sight orthogonal to major faces of each object. For the clay heads, inspection times were highest for the frontal view and a view close to the profile. Thus, for all classes of objects investigated, participants preferred views aligned with an object’s major axes relative to off-axis views. These results suggest that the consistency of canonicity across task (Palmer et al. 1981) cannot be extended to inspection tasks (a tech-

nique not used by Palmer et al.).

What is it that makes a view canonical?

Although there have been several studies demonstrating the existence of canonical views, there has been less attention paid to what properties of an object make a view canonical. To begin to understand canonicity, we should consider the following factors:

- *Goodness for recognition.* More specifically, in terms of:
 - Salience and significance of features for an individual observer. Depending on the strategies applied, we might expect individual variations between participants as well as training effects (e.g. Cutzu & Edelman, 1994).
 - Stability of the aspect with respect to small transformations (Koenderink & van Doorn, 1982). Generalization to novel views is impaired if small transformations lead to significant qualitative changes in the image or occlude relevant features. Accidental views, defined as unstable views for which even small rotations produce qualitative changes in the aspect, are therefore not expected to be easily recognized.
 - The number of occluded features. If too many or crucial features are occluded, recognition may become difficult due to insufficient information.
- *Familiarity.* Object representations and the strategies applied for recognition are influenced by the views that are encountered most frequently and by the views that are encountered during initial learning (Edelman & Bühlhoff, 1992; Tarr & Pinker, 1989; Tarr, 1995).
- *Functionality.* Within an action-perception framework, recognition should be tuned to the views that are most relevant for how we use an object (both in terms of grasping it and how it interacts with other objects). Many classes of objects, e.g., tools, are defined by their function rather than by their visual appearance and observers typically have a characteristic mode of interaction with each such object. The effects of functionality and familiarity are difficult to distinguish experimentally, since in our everyday life we typically encounter objects in views appropriate to their purpose.

- *Aesthetic criteria.* Geometric proportions can have considerable influence on which views are preferred. It is well known that rectangles with the ratio of the sides corresponding to the golden section (1:1.63) are more appealing than others (Sander, 1931). The same might be true for views of three-dimensional objects with contours forming golden sections. Difficult to assess scientifically, aesthetic criteria will not be considered here.



Figure 1: The “Spaceball” input device for rotation and translation of three-dimensional object models on a computer screen.

General Methods

Using an interactive measurement technique based on three-dimensional computer graphics, many of the technical limitations of previous experiments may be overcome:

Virtually infinite angular resolution of viewpoint. In the goodness-rating for familiar objects (Palmer et al., 1981 and Verfaille et al., 1995), participants had only a limited set of views from which to choose. Of these, only a small subset of views may serve as candidate canonical views to human observers, since the test set included views from behind or from below. As a result, prior studies could only measure canonicity with a very low angular resolution. Moreover, it is possible that a given image, although close to a potentially canonical view, was rejected due to idiosyncratic properties specific to that image, for example, the occlusion of a highly relevant feature. An experimental paradigm that allows participants to continuously rotate the stimulus objects may avoid this problem.



Figure 2: The set of three-dimensional object models used in Experiment 1a.



Figure 3: The set of object models used in Experiment 1b, which had the same design and procedure as Experiment 1a. We ran Experiment 1a and 1b on different groups of participants.

Direct measurement of viewing angles. In none of the experiments reported by Palmer et al. (1981) were the precise values of rotational angle available to the experimenter. Instead, they analyzed proportions of visible surfaces as an indirect measure of the view angle. In a computer graphics system, however, rotational angles of objects can be controlled and recorded with great precision.

Active exploration instead of passive viewing of a restricted set of views. In our experimental setup, participants were able to rotate objects in realtime on a computer screen interactively with a three degree-of-freedom input device (a “Spaceball” – see Figure 1). The mounting of the Spaceball’s sphere is equipped with sensors for forces and torques in all six degrees of freedom. In our experiments, the translational degrees of freedom were not used. When a torque around its center is applied to the sphere, the object shown on

the screen rotates at a proportional speed. Measuring only torques and forces, the Spaceball itself can not be rotated, so handling it requires some skill. After a preliminary training period of several minutes, however, most participants became proficient enough to orient an object in any desired viewpoint.

In both Experiment 1 and 2, participants explored each object extensively and oriented it according to a given set of instructions. Participants reported that this experimental paradigm is more natural and intuitive than rating tasks.

Stimuli

Figures 2 and 3 show the three-dimensional object models¹ used in two versions of Experiment 1.

¹Most of the models are in the public domain. They are available via Internet <ftp://avalon.viewpoint.com> or on a CDROM distributed by Syndesis Corp. Jefferson, WI.

Stimuli were shown on a Silicon Graphics workstation with 24bit color graphics using SGI Inventor software. The surfaces of all models were coloured and some textural properties were included (e.g., a metallic sheen). A parallel light source was located 45° above the virtual camera (no lateral displacement), and a Gouraud shading model was applied. Perspective projection was calculated as if the object were a toy located in the plane of the screen, seen from a distance of 65 cm. This kept perspective distortions to a minimum. Objects were viewed binocularly, but without stereo.

The set of stimuli can be divided into two main categories:

1. Novel objects with reduced effects of familiarity and functionality. If canonicity is merely due to contextual bias (familiarity and functionality), novel artificial objects should not exhibit any anisotropy of the distribution of views selected by participants. If we do find evidence for canonical views for novel objects, however, they most likely will inform us regarding the geometrical factors relevant for canonicity (see also, Bühlhoff & Edelman, 1992).
2. Realistic models of familiar objects. Formed under ecological conditions, canonical views of these objects reflect a complex interaction of factors. Participants' vast experience with these objects implies some familiarity with the details of their appearance and function as well as some knowledge regarding the relevance of different attributes for purposes of recognition (e.g., the spout of the teapot).

Experiment 1

Participants

We ran two versions of the experiment with slightly different stimulus objects. Experiments 1a and 1b used different groups of naive, paid participants aged between 18 and 35 years drawn from the Tübingen, Germany community. There were 36 participants in Experiment 1a and 28 participants in Experiment 1b.

Design and Procedure

Both Experiment 1a and 1b used the same design and procedure. Participants were given the following written instructions:

“Suppose you were making a brochure and you tried to give your customers the best possible impression of

the objects shown on the screen. Which views would you choose?”

The instructions are a mixture of the “goodness” and the “photography” tasks used by Palmer et al. (1981). Implicitly, all of the factors contributing to canonicity have to be taken into account by participants. The context of designing a brochure is intended to stress the “goodness” factor and to restrict the participants' potential for selecting unconventional views. In fact, highly non-canonical views are an important element of so-called “Creative Photography.” On the other hand, if we explicitly asked participants which views they considered the best for recognition, we might measure their conscious inferences rather than their spontaneous reactions/perceptions. Conclusions drawn from such data would be suspect. Even though our specification of the criteria, in terms of the specific view, was implicit and somewhat vague, participants had a clear concept of what they *thought* they were to do.

Prior to the experiment, participants were given practice familiarizing themselves with the Spaceball input device. Following this practice, they pressed a key and the first object appeared on the screen. The order of objects and the initial viewpoint of each object was randomized. There was no time limit, allowing participants to explore each object thoroughly. Having found the viewpoint they preferred the most, they pressed a key, and after this viewpoint was recorded, the next object appeared. The entire session took about 30 minutes.

As mentioned, the only difference between Experiments 1a and 1b were the objects used as stimuli. The fourteen objects used in Experiment 1a are shown in Figure 2, and the seventeen models used in Experiment 1b are shown in Figure 3.

Results

The results of Experiments 1a and 1b will be reported together (there were no significant differences between the two). The results may be summarized as follows:

- The term “canonical view” is meaningful. For example, Figure 4 shows a distinct clustering of the angular distributions of viewpoints selected by participants. Most participants looked at the cow in an oblique direction from the left or from the right and from a position only slightly above the ground.
- Most participants preferred off-axis views to straight front- or side-views. This result is

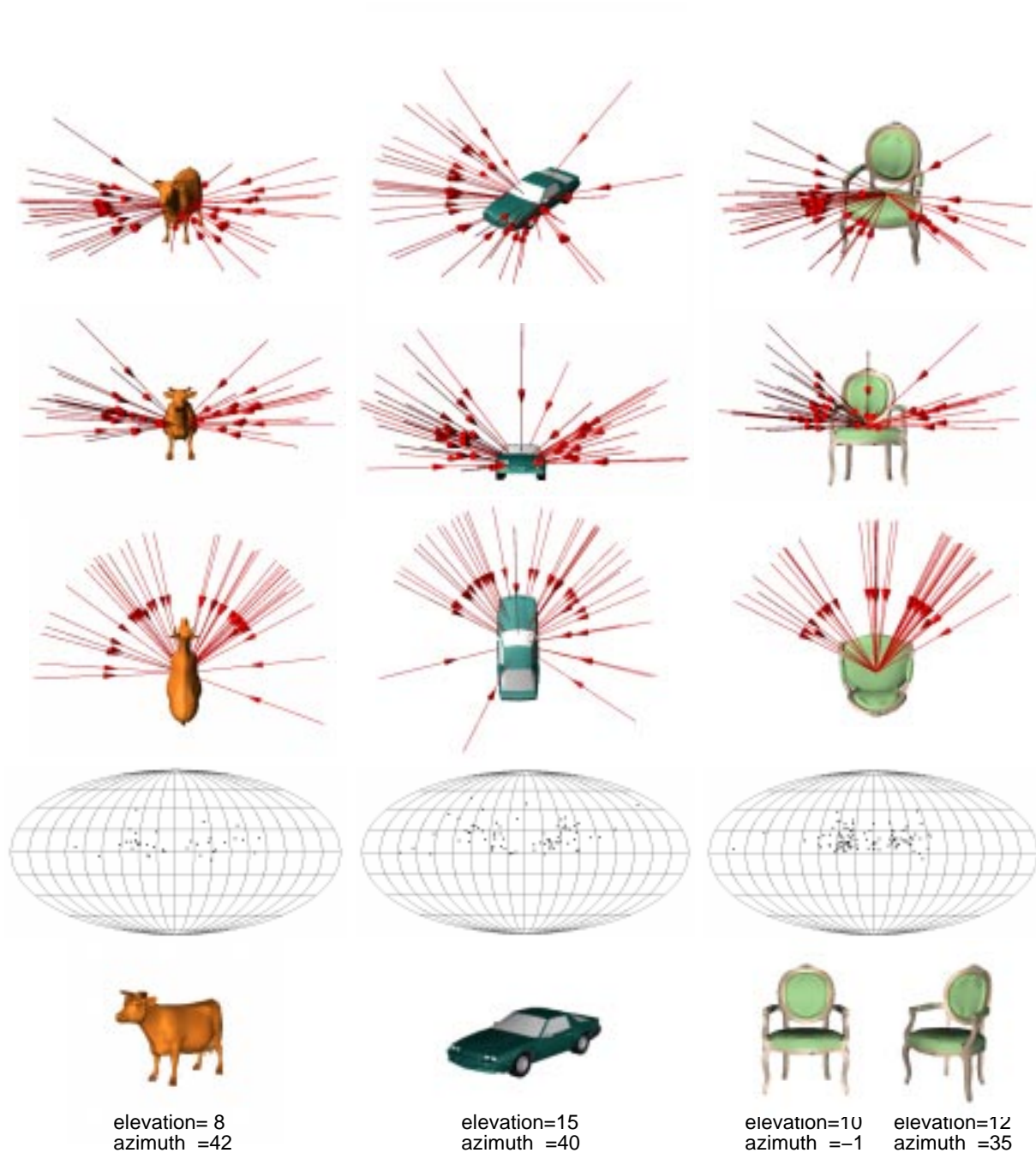


Figure 4: The arrows pointing to the center of rotation of each object indicate viewing directions selected by individual participants in Experiment 1. Most responses can be assigned to a small number of distinct clusters. The diagrams in the fourth row are Mollweide projections of the entire viewing sphere, with each dot standing for one selected view. Grids are centered with respect to the objects' frontal axis, and spaced in steps of 20° . Car and chair data were pooled over several three-dimensional models. All three objects show two symmetrical lateral clusters of responses, and the distribution for chair models exhibits an additional cluster in the median plane. In the bottom row, the mean views of these clusters, computed using circular statistics, are shown (Batschelet, 1981). In circular statistics, the mean of a set of directions can be defined as the mean of a set of unit vectors pointing to the given directions. For lateral clusters, a mirror reflection of views into one hemisphere was applied.

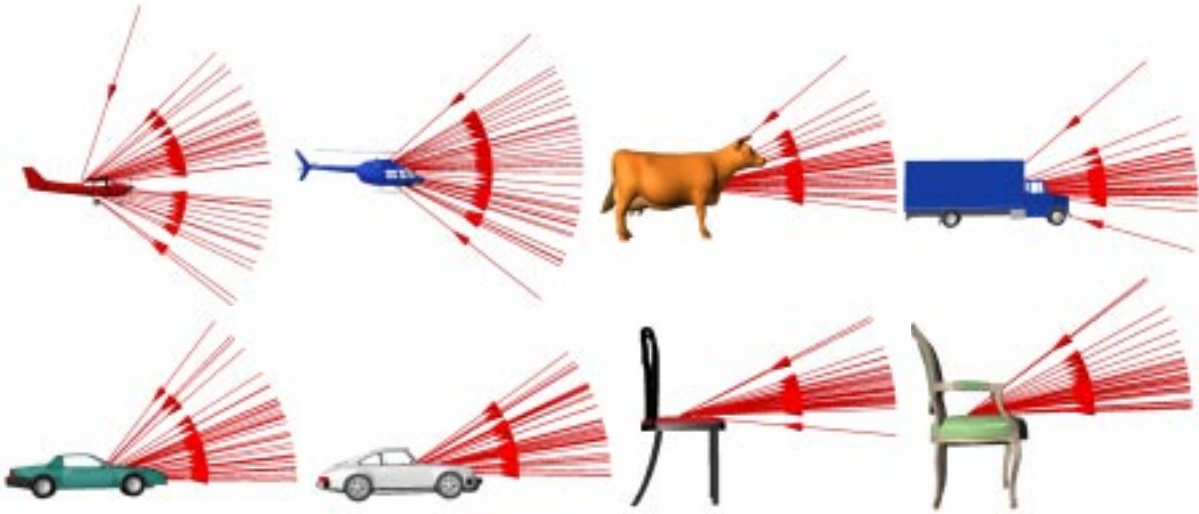


Figure 5: In order to illustrate the distribution of viewing directions in height, all arrows were rotated into the objects' symmetry planes. While participants selected views on aircraft both from above and from below, the other objects show only views from the upper half of the viewing sphere. Note the gap at horizontal directions in the distributions for the helicopter and the airplane.

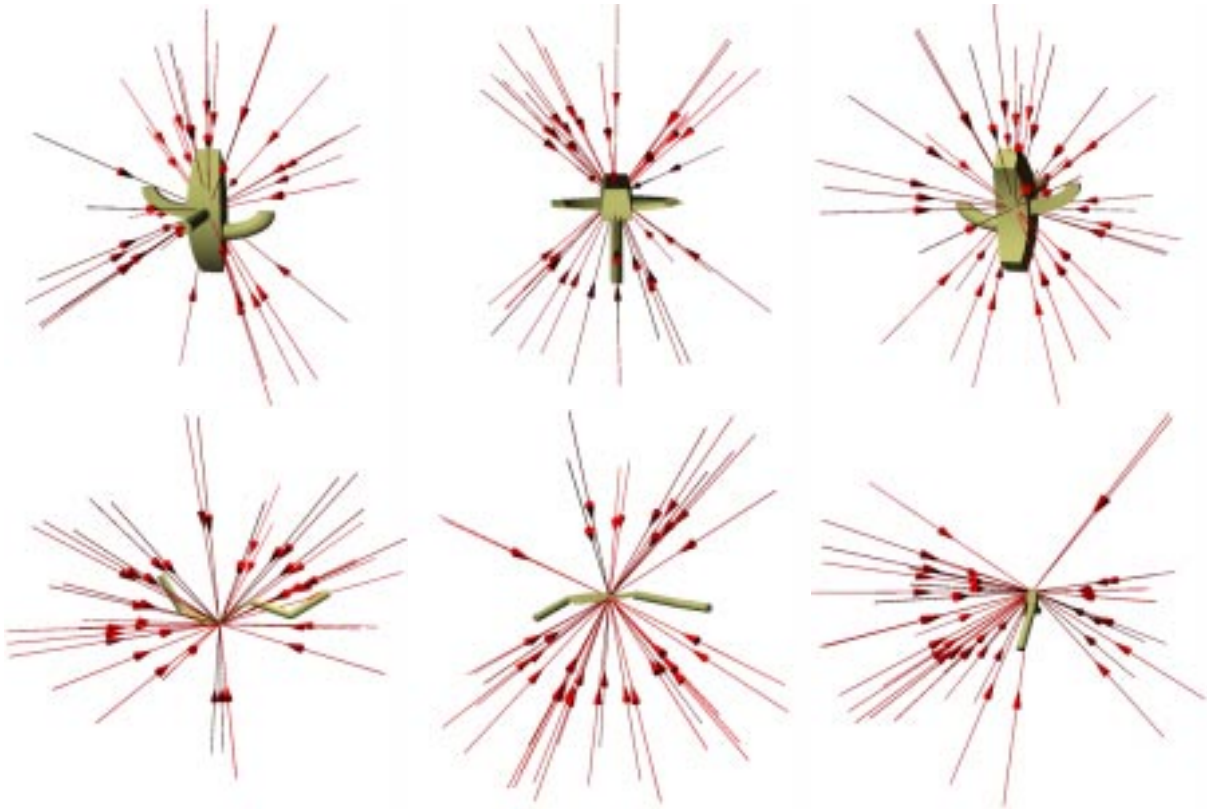


Figure 6: Even though novel objects have no overall preferred view, distributions of responses are still not isotropic. Clearly, participants avoided occlusions and accidental views. This can be seen in the top view of the 'w'-object (top row, center) and in views along the plane defined by the 'w'-shaped object (bottom row, center and right).

consistent with the idea that oblique views make more surfaces simultaneously visible and therefore tend to yield images with a high degree of informativeness (a result consistent with the findings of Palmer et al., 1981).

- Preferred viewing directions tend to be relatively low to the ground. For the chair models (Figure 4), the views reflect what we see standing approximately seven meters away from a chair, or sitting at a distance of four to five meters. Even though the average view that we calculated (Figure 4) does not seem unusually low, its height may in part be a result of the experimental setup: The distance in perspective projection corresponded to 4.3m for a real-sized chair, and this might have biased participants. A second possible consequence of the specific setup is that participants look on the screen almost horizontally, and they perceive the object in the plane of the screen or behind it. If they adjust the object on the screen to a view that seems upright in this reference frame, they inevitably get viewing directions that are horizontal in the object’s reference frame. However, our approach supposes that participants completely disregard their environment and only reproduce the retinal image of their preferred view (see Hinton & Parsons, 1988). In fact, this abstraction from the actual reference frame is often achieved: For example we find it natural to look at an image that shows a landscape but is placed flat on the table.
- According to context, some objects seem to have preferred views from below. The views selected for aircraft spanned both below and above the midline.² Views for cars, however, were almost exclusively above the midline. This is not surprising after all, since bottom views of cars are neither informative nor familiar to most of us.
- Accidental views are avoided. This is most striking for the airplane model in Figure 5. In the figure, all arrows are rotated into the symmetry plane of the object in order to visualize the angular distribution of viewpoints in height. The airplane data show a distinct gap at horizontal viewing directions. Horizontal

²For another airplane, a jet, only top views were observed. This might be due to the particular shape of the model: it had no landing gear, and the flat lower side contained no interesting structures at all. In this case, views from above were much more diagnostic.



Figure 7: Responses for the teapot model are restricted to a relatively narrow interval that achieves an upright orientation of the model on the screen without producing any occlusions.

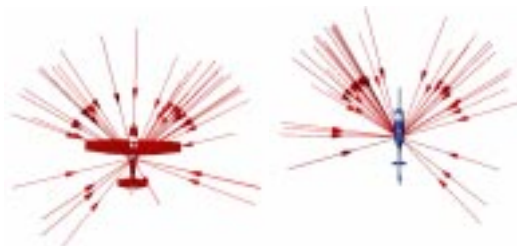
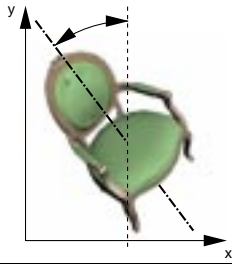


Figure 8: Most participants selected 3/4-views for the aircraft.

Table 1: The number of times participants selected left, central and right views of objects in Experiment 1. Note that there is no evidence for a preference for either side.

	left-handed participants			right-handed participants		
	left	mid	right	left	mid	right
teapot	7	0	7	13	0	8
geon-object	6	4	4	8	1	12
cow	7	0	7	9	0	12
helicopter	7	0	7	14	0	7
airplane	7	1	6	8	1	12
car (Camaro)	4	1	9	9	1	11
car (Porsche)	6	1	7	15	1	5
truck	7	0	7	11	0	10
chair (victorian)	4	3	7	5	4	12
chair (modern)	4	0	10	8	2	11
chair (office)	5	1	8	10	3	8
total	64	11	79	110	13	108



2D-orientation of objects' vertical axis	Chairs	Cars
less than 2°	55.6%	44.4%
less than 5°	85.2%	68.1%
less than 10°	92.6%	84.7%
less than 20°	99.0%	94.4%

Table 2: Percentage of trials where the objects' vertical axes were oriented upright on the screen with different precision. The data are based on $3 \cdot 36 = 108$ trials for chairs (3 models, 36 participants) and $2 \cdot 36 = 72$ trials for cars.

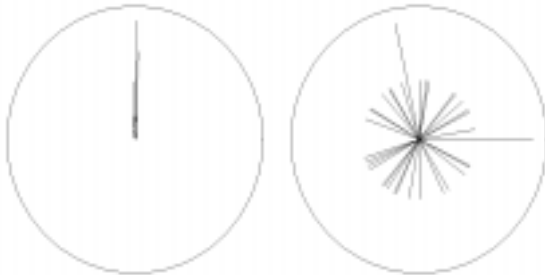


Figure 9: Polar histograms of the distribution of orientations of the objects' vertical axis. Left: chairs (108 trials); Right: enlarged histogram for the novel geon-object (36 trials). Chairs were oriented in an upright orientation very precisely, whereas there was no preference for the novel object.

views would be accidental with respect to the wings because at these particular directions the wings are projected to lines. This constellation of features is visually unstable, atypical, and provides little information about the shape of the wing.

- Occlusions are avoided. In Figure 7, it is clear that the teapot was oriented such that neither the spout nor the handle was occluded. Similarly, for the novel geon-object (Figure 6, top line), participants attempted to select a view that made all geometric components visible.

Novel objects have no canonical views. As illustrated in Figure 6, novel objects do not seem

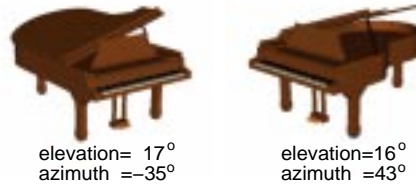


Figure 10: Unlike most other objects used in this study, the grand piano is not symmetrical, and indeed the distribution of selected views shows two lateral clusters at different azimuths. Their means are displayed here.

to have clearly preferred views. This finding supports the notion that context and experience have major impact on the formation of canonical views. The fact that different participants do not share the same preferred views for these objects is consistent with the findings of Cutzu and Edelman (1994) who obtained a similar result in studying human recognition performance on paperclip objects.

Some views of novel objects are still avoided. For the novel geon-object, the preferred view distributions exhibit gaps at viewing directions that would lead to accidental views or to occlusions of components. The “w”-shaped object in Figure 6 rarely has a preferred view at directions along its object plane since these views would be accidental. Interestingly, the viewing direction orthogonal to the object plane is not really predominant either. Some views avoid maximally spreading or greatly compressing the distance between components, and in conjunction with marked shading, such oblique views give a better impression of the object's geometry than straight top views.

Responses are highly symmetrical. For symmetrical objects, views from the left were selected about as often as views from the right (Table 1).³ Still, participants did not consistently select the same side for all of the objects, nor did they simply select sides at random. In debriefing, they stated that for individual items they clearly preferred one side to the other.

Functionality seems to play only a minor role. The number of preferred views for the teapot (Figure 7) were about equivalent between the left and right sides. Even though teapots are

³Views were divided into three categories: left, right, and centered (mid). Centered views were defined as preferred views that had an angle of less than 10° away from the symmetry plane.

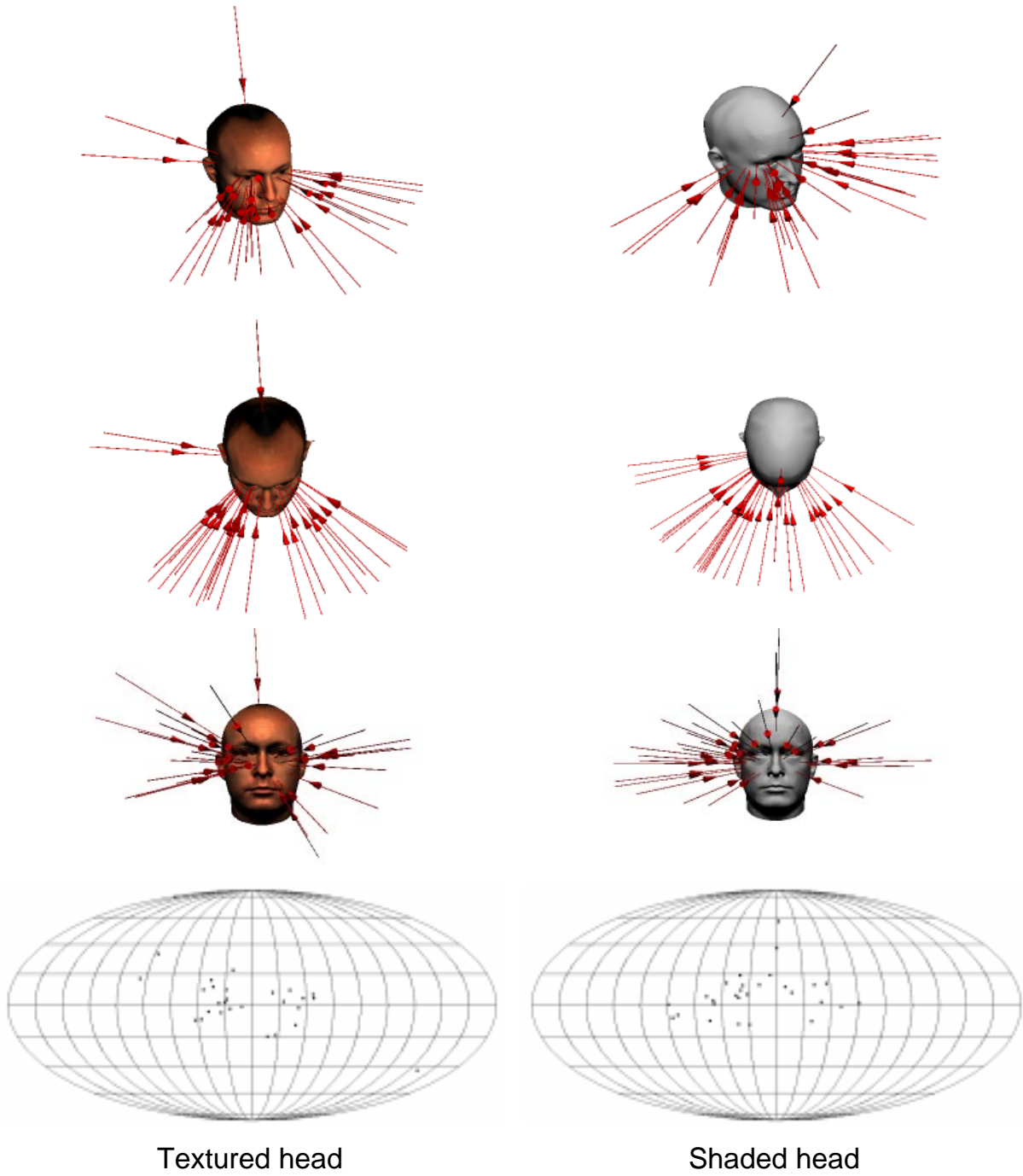


Figure 11: Selected views for a textured head model (left) and the same head model without texture (right). For the textured head, participants showed a clear preference of directions about 30° from the median axis. However, for the shaded head the distribution is more uniformly spread from full-face to profile-views.

symmetrical objects, their left and right sides are far from equivalent to observers: We would expect right-handed participants to turn the handle to the right, and left-handed participants to turn the handle to the left. Surprisingly, such an effect of functionality does not seem to occur for left-handed participants, and for right-handed participants the results reveal only a very small asymmetry (Table 1).

Preferred views for familiar objects are closely aligned with the gravitational upright. The above analyses only considered viewing direction, but not image plane rotations. This additional degree of freedom can be described in terms of the angle between the projection of the object’s vertical axis and the y-axis in the image (see Table 2). Of course, familiarity and functionality are likely to preclude depicting chairs or cars upside down, and indeed most participants very precisely adjusted the models to an upright orientation. Novel objects, however, were portrayed in arbitrary orientations (see Figure 9).

Results specific to the head model

Because face perception is sometimes thought of as a “special” recognition subsystem (e.g., Farah et al., 1995), we will consider results for the human head model separately. The stimulus, a three-dimensional scan of a human head (cf. Troje & Bühlhoff, 1996), was presented either coloured with a realistic texture or with an untextured shaded gray surface (Lambertian shading). The two versions of the head appeared in random order as two members of the object set of Experiment 1b.

The results are illustrated in Figure 11. Preferred viewpoints for the textured head are clustered at about 30° from the median axis. The mean azimuth is 32° with only very few responses above 45° , so although the average preferred view falls between a full-face and profile, there is a tendency towards the full-face view. In contrast, the distribution of viewpoints for the shaded, untextured head is spread over a much wider range of viewpoints from a full-face view to almost a complete profile view.

At least for textured heads, the viewpoints selected by participants seem appropriate: Recognition experiments with face models under varying pose (Troje & Bühlhoff, 1996) showed that optimal recognition performance was achieved if textured heads were learned from viewpoints of about 30° from the median plane, and at about 60° for the untextured faces. Interestingly, according Troje

and Bühlhoff the viewing direction of the test views (as opposed to the learned views) had little effect on recognition performance. Finally, the preferred views found in Experiment 1 do not correspond to those observed in inspection time measurements. For example, Harries et al. (1991) found peak inspection times at full-face, 0° , and near-profile, 72° , views with model clay heads placed on a turntable.

Experiment 2

A second task used by Palmer et al. (1981) involved assessing the viewpoints selected by observers when they were asked to spontaneously form mental images of objects. This task is considered diagnostic of the views that are most heavily weighted in long-term visual representations of objects in that it is assumed that such views will be most accessible in memory.⁴ To address the issue of the views that are most readily imaged, we slightly modified the experimental setup used in Experiment 1.

Participants

Forty-five naive participants aged between 18 and 35 years participated in the experiment for pay. Again the participants were drawn from the Tübingen, Germany community.⁵

Design and Procedure

Participants were first familiarized with the Spaceball device during a training period in which they played a simple three-dimensional orienting game. Following training, the name of an object was displayed on the computer screen, and participants were asked to imagine this particular object. As soon as they felt they had a stable mental image, they pressed a key, and a three-dimensional model of the object appeared on the screen. Participants had to orient the object according to their mental image, and press a key when they felt that the viewpoint of the model corresponded with the viewpoint in their image. The viewpoint was recorded, and the name of the next object for mental imaging was displayed. Again, there was

⁴At least for a view-based representation. See, for example, Bühlhoff et al., 1995.

⁵Ten of the participants performed both Experiment 1 and Experiment 2. In this case, Experiment 2 was run first.



Figure 12: The set of three-dimensional object models used in Experiment 2.

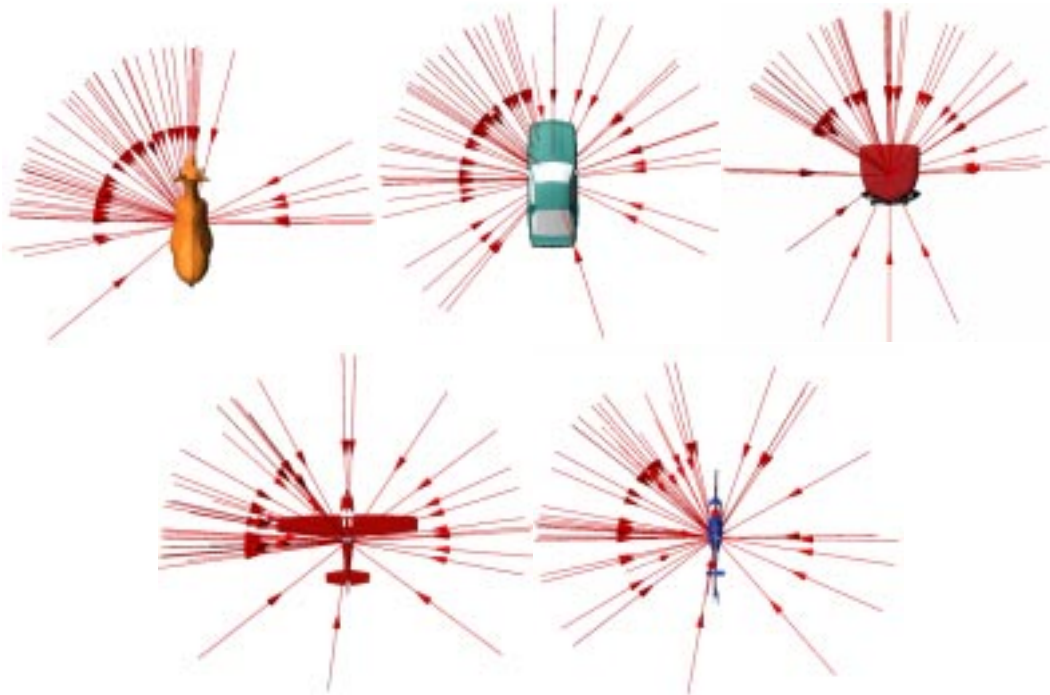


Figure 13: In Experiment 2, selected views are spread over a wider range than responses in Experiment 1. Straight front or side views are much more frequent, sometimes even predominant. Moreover, the number of left and right views is very unequal.

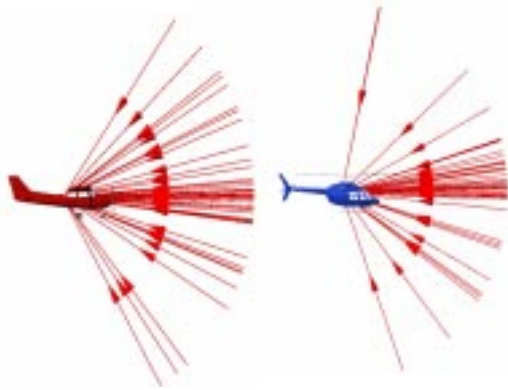


Figure 14: The distribution in height of the selected views in Experiment 2 is almost complementary to the one of Experiment 1. In the imagery task participants prefer horizontal viewing directions, whereas these views are avoided in the photography task. As in Figure 5, all arrows were rotated to the objects' symmetry planes.

no time limit. Figure 12 shows the eleven objects used in Experiment 2.⁶

Results

The pattern of preferred views was fundamentally different in the imagery task. As illustrated in Figures 13 and 15, participants often selected different views in the imagery task as compared to the photography task (Experiment 1). This result is inconsistent with the work of Palmer et al. (1981) who suggested that there was a consistency in canonical views across tasks. Supporting our finding, Perrett et al. (1992) obtained a similar result with a single, novel object. However, they speculated that for familiar objects, responses might be consistent in both tasks. Here we find that this is not the case.

Many more straight frontal- and side-views were preferred. In contrast to the photography task, the number of straight frontal- and side-views was much greater in the imagery task, and in the majority of cases these views were selected more often than oblique views.

Accidental views are no longer avoided. Even though straight frontal- or side-views are very often accidental, they were frequently selected in the imagery task. The most striking ex-

⁶Obviously, in this paradigm no nonsense objects can be used, nor can we distinguish between different shapes for exemplars of the same class. Only items with well-known names or clear descriptions on a subordinate-level, such as different types of cars, could be used.

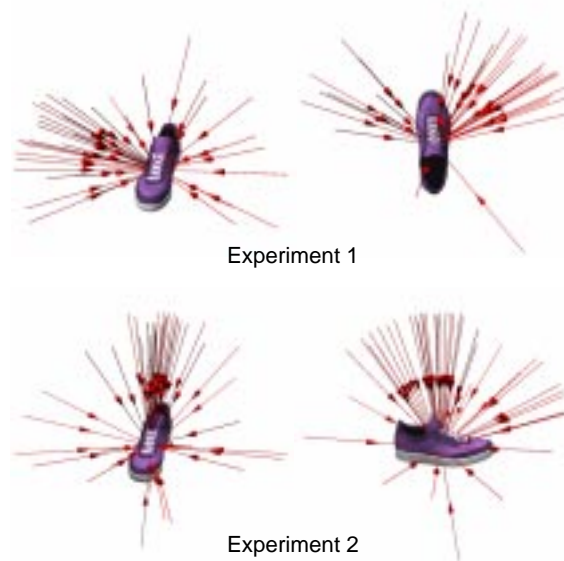


Figure 15: The shoe provides a good example of the inconsistency of responses in the photography and imagery tasks. While in Experiment 1 most participants had preferred off-axis views on the outer side of the shoe, in Experiment 2 straight top views were predominant.

ample of this is the airplane (Figure 14. Compare this to Figure 5) from Experiment 1 where we find almost complementary results. Specifically, in Experiment 1, there was a gap at horizontal viewing directions, whereas it is precisely these views that are preferred in Experiment 2.

Substantially different responses are also observed for the shoe. In Experiment 1, most participants positioned the shoe as if they were looking at someone else's foot. That is, they selected its outer side, the area that is often used for emblems on shoes. Moreover, hardly anybody picked the view that one would see when looking down on our own foot. In contrast, in Experiment 2, we find that participants strongly preferred straight top-views. In this case the experimental setup may have biased participants to imagine top views. In order to make sure that the image on the screen corresponded with the mental image, the additional information "right shoe" was given when participants were asked to form their mental image. Since the most typical way to imagine a right or left shoe might be to think in terms of one's own egocentric frame of reference, this procedure could result in a bias for top views.

Asymmetrical distribution of viewing directions. Unlike the photography task in Ex-

Table 3: In Experiment 2, right-handed participants selected many more left views than right views.

	left-handed participants			right-handed participants		
	left	mid	right	left	mid	right
teapot	6	1	4	26	2	5
cow	7	1	3	24	5	4
helicopter	7	1	3	25	2	6
airplane	6	1	4	25	2	6
car	5	1	5	26	1	6
truck	6	1	4	20	8	5
chair	3	2	6	20	8	5
sidechair	3	6	2	18	10	5
total	43	14	31	184	38	42

periment 1, the imagery task leads to a pronounced asymmetry in preferred viewpoints (Table 3). Also, more individual participants were consistent in their selection of a given side for all objects. One reason may be that responses were correlated with the right- and left-handedness of participants. While right-handed participants showed a strong preference for the left sides of objects, left-handed participants showed little preference and were almost symmetrical.

An example of this asymmetry may be found for the teapot – a case where functional considerations are likely to produce differences between left- and right-handed participants. As in Experiment 1, we would expect that right-handed participants would prefer a view of the teapot with the handle pointing to the right and left-handed participants would prefer a view one with the handle pointing to the left. While most right-handed participants indeed formed a mental image in the functionally-appropriate view, left-handed participants often imagined a functionally-inadequate view.⁷ Debriefed after the experiment, participants said they realized that their preferred view was not functional, but offered no explanation for their seemingly unusual preference.

Discussion

We investigated the nature of canonical views (e.g., Palmer et al., 1981) using an experimental setup where participants were able to rotate three-dimensional objects on a computer screen in real-time. This technique allows more precise measurements of canonical views than ranking and naming experiments using limited sets of fixed views.

⁷We made sure that they really would grasp a teapot with their left hands.

Differences between tasks

In both a photography task and a mental imagery task, we found relatively distinct clusters of preferred viewing directions, so the term “canonical views” is legitimate in the context of our experiments. However, unlike some earlier studies, we found that canonical views were not identical for different tasks. While participants clearly preferred off-axis views in a photography task, many participants selected straight frontal- or side-views in a mental imagery task. These results contradict those of the Palmer et al. (1981), but are consistent with the results of Perrett et al. (1992). In particular, our results extend those of Perrett et al. (1992) to highly familiar objects, and we provide some evidence that this difference in responses is not due to a lack of experience with off-axis views of particular objects.

The differences in preferred views between Experiments 1 and 2 might be due to a tradeoff between diagnosticity and simplicity of views. In the photography task, a good view should contain as much information about the object as possible, and it may well be highly complex, as long as the important features of the object are salient enough. Mental images, however, are subject to internal storage and processing economy, which might restrict image complexity in terms of both information content and effects of perspective projection. In this respect, plan-views tend to be less complex than off-axis views. Preserving the objects’ cross-sectional shape (Perrett et al., 1992), plan-views can still be highly diagnostic.

Another reason for the different results in Experiments 1 and 2 may be found in the level at which objects were accessed. In Experiment 2, the specific descriptions displayed on the screen suggested a basic-level discrimination, and indeed a chair can be discriminated from a car very well if both are shown in a side view. Within the particular context given in Experiment 1’s instructions, however, a subordinate-level discrimination was suggested, since brochures usually emphasize specific features of the advertised items with respect to the competitors’ products of the same type.

Sources of canonicity

Stability, familiarity, and functionality. In the photography task, participants clearly avoided accidental views and occlusions. Much of their responses can be explained in terms of maximal stability and information content of views. However, there are also strong effects of context. For objects

with a clear front and back side, the front side was clearly preferred. Views from below were avoided for cars and many other objects, but not for aircraft. This may be due to familiarity as well as functionality. Our data with a teapot model suggest that functionality seems to play a minor role. We would have expected right-handed participants to rotate the handle to the right, and left-handed persons rotate it to the left, so that the teapot could be grasped with the dominant hand. We found only small correlations between the orientation of the teapot and participants' handedness.

Observer biases. While the distribution of viewpoints selected in the photography task was symmetrical with respect to the objects' median axis for symmetrical objects, there was a strong asymmetry in responses in the imagery task. A closer analysis showed that while left-handed persons selected left views about as often as right views, right handed persons preferred left views. We have no explanation for this effect.

Object geometry. An important issue of the canonical view phenomenon is the role of geometrical features of the object on the one hand and contextual factors such as familiarity and functionality on the other hand. Using novel objects, we investigated participants' preferences in the absence of context. Unlike realistic objects, novel objects turned out to have no overall preferred views. Still, there was some structure in the distribution of selected views, and this had to be completely determined by the objects' geometry. Indeed, responses could be explained in terms of accidental views and occlusion.

Viewing position. Measuring viewing angles quantitatively, we found that in most cases elevation of preferred viewpoints is significantly below 45° . Likewise, the azimuth angles between preferred viewpoints and the objects' median axes are often significantly less than 45° . Such data might be useful for the design of further experiments on canonicity and object recognition.

A computational analysis

A separation of geometrical and contextual factors on realistic, highly familiar objects can be achieved by comparing our psychophysical results with a ranking that is based on a purely objective criterion of "goodness." A number of suitable definitions of "goodness" are conceivable. In an analysis that has not been described in the previous sections (Blanz, Schölkopf, Bühlhoff, Burges, Vapnik, & Vetter, 1996), we assessed the preferred views using the view generalization of a view-based ob-

ject recognition algorithm. On a given pair of images, after centering and downsampling them, our algorithm applies oriented filters to smear the images in four different directions separately. Then, the four pairs of equally filtered images are compared. If at least one of the comparisons yields a below-threshold image distance, the initial images are classified as showing the same object. The algorithm is quite robust with respect to in-depth rotations, and its view generalization from given views is mainly governed by view stability.

The algorithm was run in a subordinate-level classification task on chairs, airplanes, and a set of novel geon-objects (similar to the one in Figure 6, top row, with different volumetric primitives attached in slightly different positions.) We then computed a ranking of views according to generalization performance upon adjacent viewing directions. Measurements were based on sets of computer-generated views of three-dimensional models from directions that were spread over the entire viewing sphere. Comparing these computational results with our psychophysical data, we found that for chairs the views selected by participants in Experiment 1 were also ranked high in generalization performance. Coincidence was even higher if the set of ranked views was restricted to the upper frontal quarter of the viewing sphere at a finer spacing. Favoring off-axis views, the computational data fit better to the results of Experiment 1 than Experiment 2. Purely geometric criteria as measured by our ranking could not account for the preferred views of aircraft and geon-objects, however. Even though all of the ranked views on the airplanes depict stable, upright states, some of them look more natural than others. For the geon-object, it should be mentioned that, unlike our participants, the algorithm was quite tolerant with respect to occlusions of components. Processing mainly low spatial frequency structures, it more heavily weighted the locations rather than the identities of the components composing the object.

Conclusions

We began this paper by suggesting that canonical views may inform us regarding the nature of object representations and recognition. Indeed, we believe that our results do have relevance for current theories of object recognition.

Let us consider two prototypical classes of theories, "geon-structural-descriptions" (Biederman & Gerhardstein, 1993) and "view-based" theories (Bühlhoff, Edelman, & Tarr, 1995; Poggio & Edel-

man, 1990). Both are view sensitive in that they differentially encode different viewpoints within the representation of the object. However, in the case of geon-structural-descriptions, each distinct view is instantiated with one trial of exposure according to changes in visible part configurations (Biederman & Gerhardstein, 1993) – that is, a qualitative geometric encoding similar to an aspect graph (Koenderink, 1987). In the case of view-based models, each distinct view is instantiated in large part according to the degree of familiarity with clusters of viewpoint-specific features – that is, a quantitative experiential encoding. For example, Tarr (1995) created “canonical” views for a small set of arbitrarily selected viewpoints. Given extensive practice recognizing novel three-dimensional objects at these viewpoints, participants became progressively faster and more accurate for these familiar views as compared to unfamiliar views. Similar results have been obtained by Bülthoff and Edelman (1992) and Humphrey and Khan (1992).

With regard to our present results, the heavy weighting of familiarity in determining canonicality suggests that observers are relying more on their experience with particular views than on object geometry or functional considerations. Such a pattern is unlikely to arise if views are determined purely by encounters with previously-unseen geometric configurations (Biederman & Gerhardstein, 1993). In contrast, this is precisely the pattern predicted by theories of recognition in which views are determined by statistical learning methods over repeated exposures (Poggio & Edelman, 1990). Thus, our results on canonical views generally support the view-based approach. Interestingly, there is recent neurophysiological evidence from single-cell studies that is consistent with this argument. Logothetis and Pauls (1995) trained monkeys to recognize “paper-clip” objects from many viewpoints. Much as in psychophysical studies with humans (Bülthoff & Edelman, 1992; Tarr, 1995), the monkeys’ performance became near-equivalent at all familiar viewpoints. However, single-cell recordings in visual cortex (STS) revealed the presence of arrays of view-tuned neurons for specific paperclips – that is, neurons coding for preferred views that, as a set, spanned the view space of familiar viewpoints.

In conclusion, we wish to emphasize that although canonical views appear to be largely familiarity based, there are also clear effects of geometrical properties. Specifically, participants rarely selected views that were accidental or produced a

large number of occlusions. However, there is little evidence that functional considerations *per se* play a large role in determining which views are canonical. Thus, we suggest that canonical views, much as is argued for view-based representations in general, are the product of a complex interaction between experience, task, and geometry – all three factors contributing to the organization of the representation.

References

- [1] Batschelet, E. (1981). *Circular Statistics in Biology*, London: Academic Press.
- [2] Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19(6), 1162-1182.
- [3] V. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, and T. Vetter. Comparison of view-based object recognition algorithms using realistic 3d models. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, *Artificial Neural Networks — ICANN'96*, pages 251 – 256, Berlin, 1996. Springer Lecture Notes in Computer Science, Vol. 1112.
- [4] Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci USA*, 89, 60-64.
- [5] Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, 5(3), 247-260.
- [6] Cutzu, F., & Edelman, S. (1994). Canonical Views in Object Representation and Recognition. *Vision Research*, 34(22), 3037-3056.
- [7] Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32(12), 2385-2400.
- [8] Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, 35(14), 2089-2093.
- [9] Harries, M. H., Perrett, D. I., & Lavender, A. (1991). Preferential inspection of views of 3-D model heads. *Perception*, 20, 669-680.

- [10] Hinton, G. E., & Parsons, L. M. (1988). Scene-based and viewer-centered representations for comparing shapes. *Cognition*, *30*, 1-35.
- [11] Koenderink, J. J. (1987). An internal representation for solid shape based on the topological properties of the apparent contour, In: *Image Understanding: 1985-86*, W. Richards & S. Ullman (Eds.), Norwood, NJ: Ablex Publishing Corporation, 257-285.
- [12] Koenderink, J. J., & van Doorn, A. J. (1982). The shape of smooth objects and the way contours end. *Perception*, *11*, 129-137.
- [13] Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representation in the primate. *Cerebral Cortex*, *3*, 270-288.
- [14] Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects, In: *Attention and Performance IX*, J. Long & A. Baddeley (Eds.), Hillsdale, NJ: Lawrence Erlbaum, 135-151.
- [15] Perrett, D. I., Harries, M. H., & Looker, S. (1992). Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part. *Perception*, *21*, 497-515.
- [16] Perrett, D. I., & Harries, M. H. (1988). Characteristic views and the visual inspection of simple faceted and smooth objects: "Tetrahedra and potatoes". *Perception*, *17*, 703-720.
- [17] Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*, 263-266.
- [18] Sander, F., (1931). Gestaltpsychologie und Kunsttheorie. Ein Beitrag zur Psychologie der Architektur. *Neue Psychologische Studien*, *8*, 311-333.
- [19] Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, *2*(1), 55-82.
- [20] Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*(28), 233-282.
- [21] Troje, N., & Bühlhoff, H. H. (1996). Face recognition under varying pose: The role of texture and shape. *Vision Research*, *36*(12), 1761-1771.
- [22] Verfaillie, K., & Boutsen, L. (1995). A corpus of 714 full-color images of depth-rotated objects. *Perception & Psychophysics*, *57*(7), 925-961.