Spatial Vision, Vol. 20, No. 1–2, pp. 79–106 (2007)

Also available online - www.brill.nl/sv

Low-level correlations between object properties and viewpoint can cause viewpoint-dependent object recognition

MAARTEN DEMEYER *, PETER ZAENEN and JOHAN WAGEMANS
University of Leuven, Belgium

Received 6 December 2005; accepted 30 May 2006

Abstract—Viewpoint-dependent recognition performance of 3-D objects has often been taken as an indication of a viewpoint-dependent object representation. This viewpoint dependence is most often found using metrically manipulated objects. We aim to investigate whether instead these results can be explained by viewpoint and object property (e.g. curvature) information not being processed independently at a lower level, prior to object recognition itself. Multidimensional signal detection theory offers a useful framework, allowing us to model this as a low-level correlation between the internal noise distributions of viewpoint and object property dimensions.

In Experiment 1, we measured these correlations using both Yes/No and adjustment tasks. We found a good correspondence across tasks, but large individual differences. In Experiment 2, we compared these results to the viewpoint dependence of object recognition through a Yes/No categorization task. We found that viewpoint-independent object recognition could not be fully reached using our stimuli, and that the pattern of viewpoint dependence was strongly correlated with the low-level correlations we measured earlier. In part, however, the viewpoint was abstracted despite these correlations.

We conclude that low-level correlations do exist prior to object recognition, and can offer an explanation for some viewpoint effects on the discrimination of metrically manipulated 3-D objects.

Keywords: 3-D object discrimination; object categorization; viewpoint dependence; signal detection theory.

INTRODUCTION

Viewpoint dependence and object representation

Our visual system copes with a 3-D world through 2-D input information. The loss of one dimension implies a reconstruction that is not deterministic: every 2-D projection may originate from an infinite number of 3-D objects. A longstanding

*To whom correspondence should be addressed. E-mail: maarten.demeyer@psy.kuleuven.be

problem in object recognition research is the nature of the object representation that allows us to perform this reconstruction in a sufficiently precise and swift manner. In particular, it has been much debated whether this representation is dependent or independent of viewpoint (e.g. Biederman, 1987; Biederman and Gerhardstein, 1993, 1995; Bülthoff and Edelman, 1992; Tarr, 1995; Tarr and Bülthoff, 1995). Typically this question was addressed by experiments that compare object discrimination performance within and across viewpoints (e.g. Biederman and Cooper, 1991; Hayward and Tarr, 1997; Hayward and Williams, 2000; Rock and DiVita, 1987). Viewpoint dependence in the performance data was then often attributed to viewpoint dependence in the object representation. However, the evidence is not conclusive. Abstract, metrically manipulated objects often yield viewpoint-dependent performance, whereas qualitatively different objects are often recognized equally well across viewpoints. Discrepancies to this pattern exist, however (e.g. Foster and Gilson, 2002; Tarr *et al.*, 1998; Vanrie *et al.*, 2001; Willems and Wagemans, 2001).

As a result, it has been concluded that the viewpoint dependence of object recognition performance is by itself not the informative question with regard to the nature of object representation (Stankiewicz, 2002; Wagemans *et al.*, 1996). Some efforts were made to dissociate the effects of representation from other effects that could potentially cause viewpoint dependence. For instance, Tjan and Legge (1998) noted that different tasks and stimuli pose different representational demands to object recognition. Thus, some tasks are inherently more viewpoint-dependent than others. They introduced the concept of Viewpoint Complexity (VX) of a task and stimulus set to explain previous findings, operationally defined as the number of randomly selected 2-D images needed by an ideal observer to solve the task. The authors argued that task constraints and stimulus representation properties need to be dissociated in order to study the latter.

In another paper, Tjan *et al.* (1995) measured the efficiency of human recognition of objects embedded in noise, compared to an ideal observer. They found that this efficiency was low, meaning that the observer himself and not only stimulus information is an important bottleneck for recognition efficiency. In particular, they speculated that the detection and discrimination efficiency for low-level 2-D features could significantly constrain human object recognition.

Foster and Gilson (2002) manipulated paperclip-like stimuli both metrically and qualitatively, and measured discrimination performances over a 360° range of views using signal detection theory methods. They found two additive effects on both types of manipulations: a 3-D based viewpoint-independent effect of object structure, and a 2-D based viewpoint-dependent effect independent of object structure. The authors suggested that viewpoint-dependent and viewpoint-independent effects need not be mutually exclusive, but may be the result of independent processes that can be combined to discriminate objects across viewpoints.

These studies inspired the present research. Tjan and Legge (1998) and Tjan *et al.* (1995) pointed to two important constraints on human object recognition. First, task

constraints (including stimulus information) can differ and can thus put different demands on the object recognition system, limiting the performance in different ways. Second, they focus on the bottleneck of inefficient 2-D feature discrimination. These are factors we will investigate as possible causes for viewpoint-dependent object recognition performance, but in a signal detection theory framework. Our aim is more specific than Tjan and Legge's and Tjan *et al*.'s, though: we are not concerned with the overall efficiency or representational demands of performing an object recognition task, but with one well-defined cause for systematic and predictable misjudgments of object identity across viewpoints. The methods we shall use, manipulating 3-D objects on different stimulus dimensions and measuring discrimination performances on these dimensions across views, are akin to the work of Foster and Gilson (2002). However, we will limit ourselves to a smaller range of viewpoints.

Low-level correlations and dimensional separability

We saw that factors outside object recognition itself could be at least partly responsible for viewpoint-dependent object recognition performance. Stimulus information is one such proven factor, but the early stages of the visual system are also deserving of attention. We believe that the recognition of metrically manipulated stimuli is especially prone to be influenced by these factors, and have envisioned a specific cause for the viewpoint dependence of the recognition of such objects. To explain this we will first present our general framework, based on multidimensional signal detection theory and strongly inspired by General Recognition Theory (Ashby and Townsend, 1986).

Figure 1 illustrates this framework. Suppose an observer has to monocularly discriminate two objects that differ only in curvature and viewpoint. These are the object dimensions that can be modeled here: the former we will call intrinsic, the latter extrinsic to the object. The input to our visual system is a 3-D object projected onto a 2-D surface (Fig. 1A), and is broken down into thousands of dimensions early on in the visual system (e.g. retinal receptors — Fig. 1B). The task for the subject

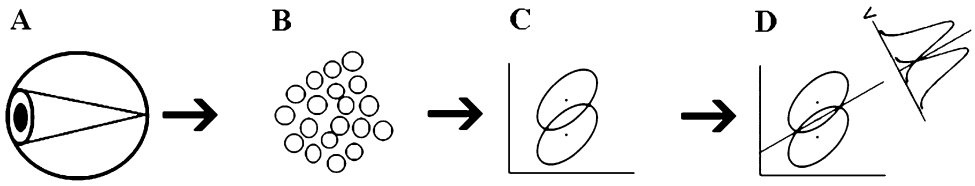

Figure 1. Our framework assumes four stages. (A) The projection of a 3-D object on the 2-D retina. (B) The deconstruction of this 2-D image into thousands of dimensions by, for instance, the retinal receptors. (C) The integration of these dimensions by early processing of the 2-D image. The objects and their noise distributions can be represented in a multidimensional space, containing as many dimensions as there are relevant object properties changing (in this case two). (D) The making of a decision by higher level processes about which object is perceived, using a decision rule.

could be to make a decision about either the 2-D image or the 3-D object, and will in either case require an integration of this multitude of dimensions.

First, in the 2-D projection of the 3-D image, information is lost. The values of the two 3-D stimulus dimensions, viewpoint and curvature, can now be assessed only by means of their 2-D effects in the projection. However, often these 2-D effects are not projected independently. They could for instance, in the most extreme case, trade-off to negate each other, and yield exactly the same 2-D image for two different 3-D objects.

Second, the 2-D image properties are to be retrieved from the multidimensional input by the early stages of the visual system (Fig. 1C). This process is imperfect, however. At this stage, the detection or discrimination of the 2-D effects of the object properties might not only be noisy, but also interdependent. We model the internal representation of objects at this stage in a two-dimensional feature space. Note that the dimensions depicted herein are the two object dimensions (e.g. curvature and viewpoint), not the dimensions of 2-D image space (e.g. horizontal and vertical). The objects can then be represented in this space along with their Gaussian noise distributions, visualized as ellipses connecting points of an arbitrary equal density in the noise distribution. This noise distribution has a variance and a covariance, resulting from the joint effects of two stages — projection upon the retina and low-level processing. The size and shape of these noise distributions will, as in standard signal detection theory, define the distribution of the possible perceptual effects of a stimulus: the density of a distribution in a given point determines the chance that the originating object of the distribution will give rise to the perceptual effect corresponding to that point. Note, however, that this is not necessarily the final conscious percept by the subject. A perceptual effect is in our framework only relevant through its likelihood to belong to different distributions in this feature space. These likelihoods will be used as input information for decision processes later on (see below). The correlation between both dimensions that is contained in these distributions is what we suspect as being responsible for some viewpoint-dependent effects in object recognition, and what we will measure in our first study. This is the *interdependence* of both dimensions.

Third, the visual system has to make a decision about which object is being perceived when a perceptual effect arises (Fig. 1D). This is a task of higher-level areas of object recognition. A decision rule will be used to attribute perceptual effects in the two-dimensional feature space (consisting of the signal plus added noise) to either object. Following some assumptions we make, this decision rule will be linear (see below, in the Appendix, and in Ashby and Gott, 1988). If the decision is based on the likelihoods of the possible perceptual effects in this space to have originated from the noise distribution of either object, as we will assume, the decision can be made in a one-dimensional decision space. This decision space will be based on a decision variable we call v. The decision rule that would minimize the variance of the noise on this decision variable, and which is displayed in this example, is the rule whose boundary connects all points that have the same density

in both distributions. If the decision rule is to be viewpoint-independent, however, it must be parallel to the viewpoint dimension. It can be seen in this figure that the viewpoint-independent decision rule would not yield the most noise-free decision. We will discuss in the next section how viewpoint abstraction fits in with this framework.

Another important concept here is the *separability* of both dimensions — the subject's ability to judge them independently. Note that this is not to be confused with linear separability in classification spaces. We hypothesize that the low-level correlation can prevent the separation of both dimensions by the visual system (*integrality*), leading to a decision rule that is correlated with the low-level correlation. Again, we will detail this in the next section.

This framework relates directly to viewpoint dependence experiments. The example in Fig. 2 shows how a low-level correlation between object property and viewpoint, and a failure to separate these dimensions because of it, can lead to a pattern of results typically found using metrically manipulated stimuli. Suppose two objects differing only on the object property dimension (the two rows of distributions), e.g. curvature, are to be discriminated under different viewpoints (the columns). The middle distributions are the 'known' viewpoints, and thus the reference for the decision rule. If a subject can separate both dimensions, he will use a decision rule lying exactly in between and parallel to both rows to discriminate both objects, unaffected by viewpoint. If the low-level correlation prevents separability, however, the decision rule will be determined by this low-level correlation (see Appendix). As we see, this will lead to faulty categorizations that will lower the performance as the viewpoint difference gets larger. In reality we do not necessarily expect full integrality, but this example illustrates the influence that correlated noise distributions could have on object recognition performance.

To summarize, viewpoint dependence can already arise at relatively early stages in our visual system, where the 2-D effects of object property and viewpoint changes are analyzed, because either the effects themselves or their low-level discrimination are interdependent. In metrically manipulated stimuli, this is possible in a highly systematic manner.

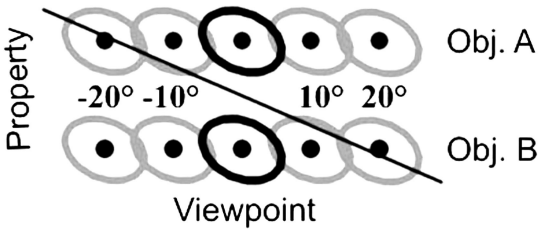

Figure 2. An example of how low-level correlations can lead to viewpoint-dependent object recognition performance. The task is to ignore viewpoint, but the decision rule is influenced by the low-level correlations and therefore slanted. This causes more faulty categorizations as the viewpoint under which a new stimulus is presented is further from the known viewpoint.

We noted that our framework is based on General Recognition Theory (Ashby and Townsend, 1986). But to be more exact, it is a special case of such a model. In GRT, each point in multidimensional feature space can have different variances and covariances, making an overall dimensional low-level correlation impossible to measure. However, if one assumes the covariance matrices of all points in two-dimensional feature space to be equal, the noise distributions in the entire two-dimensional feature space are determined by three variables only: the variances of the two individual dimensions, and their covariance. This is reminiscent of Tanner's (1956) model of dimensional interdependence, allowing for the measurement of a general covariance (and correlation) for the entire two-dimensional feature space. Note that while a general correlation is necessarily a general linear dependence between both dimensions, this is not an extra assumption: as Ashby and Townsend (1986) note, it follows from the earlier assumptions of constant covariance matrices and Gaussian noise distributions.

Within a small stimulus range and given abstract, metrically manipulated stimuli we feel the assumption of equal covariance matrices is justified. Goodness-of-fit tests of our data to our models will provide us with a test of the actual justifiability of this assumption.

Decision rules: a closer look

When two objects have to be discriminated, they could in theory yield perceptual effects all over the two-dimensional feature space, due to internal noise. This means that a discrimination task is actually a categorization task: which perceptual effect originated from which object? Thus, a decision rule must exist in this two-dimensional space, as we saw above. The optimal decision boundary connects all points in space where the likelihood of them belonging to the noise distribution of one object is equal to the likelihood of them belonging to the other object. Every deviation from this decision rule lowers the number of correct answers. Assuming Gaussian distributions of equal shapes but with different center points, this rule is known to be necessarily linear. This has been shown by, among others, Ashby and Gott (1988), who have also provided evidence that subjects can and do indeed use such an optimal rule. In a 2-D discrimination task (not abstracting viewpoint) it is influenced by the low-level correlation with viewpoint, and is therefore often not parallel to one of the dimensions.

However, a true object recognition task is not a 2-D discrimination task; it is a 3-D discrimination task in which the viewpoint dimension has to be ignored and the property dimensions have to be discriminated. The goal of object recognition is to recognize the object, not the 2-D image. The task is then to use a decision rule parallel to the viewpoint dimension and to not take into account the low-level correlation with viewpoint. However, this rule will only be achieved if separability exists.

To put it more formally, the visual system will use a linear rule that connects all points in two-dimensional feature space that have the same likelihood to belong

to both distributions in it (a likelihood ratio of 1). A linear rule can be rewritten as a weighted linear combination of the dimensions; if separability exists, it is possible to downweight the viewpoint dimension to 0. This reduces all effective variances of that dimension to 0 and eliminates the covariance. Thus, reweighting the dimensions will change the likelihoods in feature space. These likelihoods, and as a result the decision rule, will then be based on the property dimension only. This yields a viewpoint-independent decision boundary when connecting all points in which the likelihood ratio is 1. Compared to Fig. 1D and its description above, the viewpoint-independent rule will now no longer have a more noisy decision variable than the previously optimal rule. If no separability is possible, the viewpoint dimension can not be singled out for downweighting to 0, and the low-level correlation between both dimensions will influence the likelihoods in property–viewpoint space. If we still use a likelihood ratio of 1 to define the decision boundary, this will result in a decision rule that is correlated with the low-level correlation. We want to test our hypothesis that low-level correlations can cause viewpoint-dependent object recognition by comparing low-level correlations to the slopes of decision rules.

Based on the above rationale, we will continue as follows. First, we aim to measure low-level correlations and check the goodness-of-fit of our models to the data. Then we will investigate what relation they bear to viewpoint dependence at the 3-D object recognition stage, as opposed to the earlier stages at which the low-level correlations are situated. We would like to refer to Ashby and Gott (1988) and Maddox (2001) for similar work on categorization that, to our knowledge, has never been applied to 3-D object recognition across viewpoints.

STIMULUS CONSTRUCTION AND CALIBRATION

Since our research is partly concerned with the effects of stimulus information, constructing the stimulus carefully is of great importance. Our stimulus manipulations need to be metrical and dimensional in nature for two reasons: first, to be able to relate the results to relevant previous research, and second, for reasons of mathematical elegance and ease of application of multidimensional signal detection theory. We will manipulate several dimensions of the stimulus, to cover several possible degrees of expected viewpoint dependence.

Our stimuli consist of two conjoined tubes that can be manipulated metrically along four dimensions. Three of these are intrinsic to the object: Bend (curvature of both tubes), Radius (width of both tubes), and Angle (difference in rotation between both bent tubes). These are comparable to the metric manipulations used by Foster and Gilson (2002) and Stankiewicz (2002).

One dimension is extrinsic to the object, namely View (viewpoint under which the stimulus was rendered). Based on our subjective experience while creating this stimulus, we expected a strong low-level correlation with View for Angle, a moderate one for Bend, and no or little correlation for Radius.

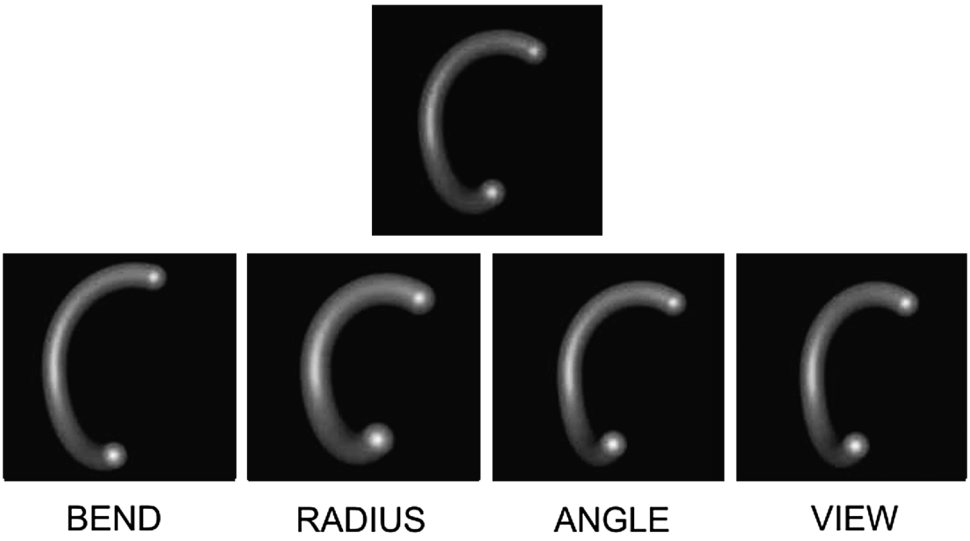

Figure 3. Our four stimulus dimension manipulations, compared to the standard stimulus above.

Next, we performed a standard calibration on the stimulus dimensions: we derive the internal noise variance of each dimension around the standard values by constructing a psychometrical function, using methods described by Wichmann and Hill (2001). This way, we have a more psychologically relevant unit of stimulus manipulation than the arbitrary physical stimulus parameters, and we can perform manipulations on the different dimensions that are on the same psychological scale. This is necessary, because it is a prerequisite to expect a full correspondence between low-level correlation and object categorization slope if full dimensional integrality exists (see Appendix). The standard deviation of the internal noise distribution on each dimension was estimated to be the average physical stimulus difference that yielded, according to the psychometrical function, a d' of 1 as discrimination performance. The subjects for this experiment were the same as in our other studies. Since these measurements were done for calibration purposes mostly, defining the unit of measurement for the other experiments, we do not report them here. Figure 3 illustrates the four stimulus dimensions we used. The upper stimulus is the 'standard' stimulus, having all dimensions set to their standard values. In the bottom part of Fig. 3, the four dimensions are manipulated over five standard deviations of internal noise each (at short stimulus durations).

EXPERIMENT 1

We will first measure the low-level correlations between View and every intrinsic object property dimension. We do this by constructing a two-dimensional stimulus space for each of these property–viewpoint combinations, and sampling a rectangular and equidistant grid of nine stimuli from it (see Fig. 4A). We will then measure

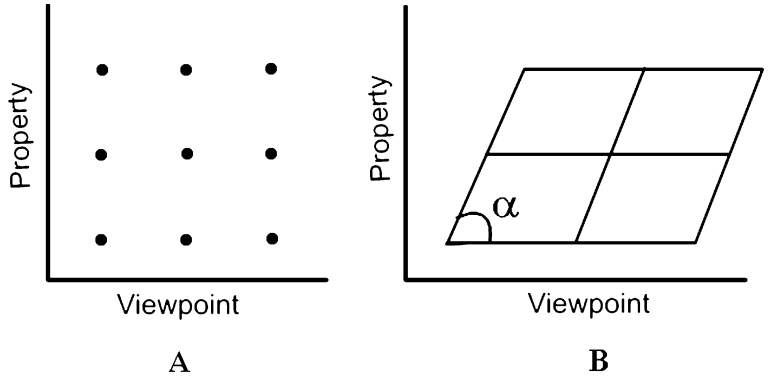

Figure 4. (A) An example of a stimulus grid in stimulus space. We sampled nine points for each property–viewpoint combination, and measured all 36 discrimination performances between them. (B) An example of a model applied to the behavioral data. Because we assume equal covariance matrices, its geometrical shape is that of a parallelogram. The angle α is directly related to the low-level correlation we want to measure.

all 36 possible discrimination performances between these stimuli, instructing the subjects to judge the 2-D image instead of abstracting the viewpoint. The subject was presented with only one stimulus per trial, and had to recognize it as one of two stimuli defined before each block. This discrimination of two stimuli is formally equivalent to the classical Yes/No task in signal detection theory. In such a task there is but one stimulus interval, and the subject makes a binary decision about it. Answering Yes or No (e.g. whether a stimulus is present) is then comparable to answering the identity of the one stimulus presented, out of a stimulus set of two. We will therefore refer to our task as a Yes/No task. We will describe it in greater detail in the Methods section.

In the analysis we estimate the dimensional angle α that fits the observed data optimally (see Fig. 4B). This angle α is directly related to the low-level correlation between viewpoint dimension X and intrinsic property dimension Y:

$$r(X, Y) = -\cos(\alpha). \tag{1}$$

We will also validate this measurement using an adjustment task. Theoretically, this should yield the same low-level correlations as the Yes/No task. In an adjustment task, a subject is presented with a test stimulus, like in a Yes/No task, but the answer is given by metrically manipulating a probe object to match the test stimulus. The low-level correlation can be computed using these data (see below), and compared to the Yes/No results.

It has been noted that low-level correlations could also be caused by the first stage in our framework, the projection of a 3-D object on a 2-D plane. We are interested in the relative importance of this stage compared to the 2-D processing stage, and to this end we will do a pixel analysis on the 2-D images. This will allow us to dissociate stimulus information from low-level processes. The analysis is similar to the analysis we do on the Yes/No data, but using pixel luminance differences

instead of discrimination performance. To exclude the possibility that any deviation from the behavioral results could be caused by the selective use of certain spatial frequencies by the human visual system, we applied different spatial frequency filters to the images, both low-pass and band-pass.

Methods

Subjects. Two of the authors (P. Z., M. D.) and one additional subject (B. W.) participated. All had normal or corrected-to-normal eyesight and were not naïve with respect to the aim of the study. We do not consider this a problem given the nature of the task. The same three subjects participated in all experiments described herein.

Apparatus. The experiment was run on a PowerMac G4 computer at a processor clockspeed of 733 MHz. The stimuli were presented on a Sony GDM-F520 19″ CRT monitor, driven by a GeForce2MX graphical card. Luminance measurements were made but no gamma correction was applied during stimulus presentation, because we found it disturbed the impression of a realistic 3-D object, and we could not see any apparent problems stemming from not using gamma correction in our experiments. However, the input images for the pixel analysis were corrected for the actual screen luminances. The screen had a spatial resolution of 1024 by 768 pixels and a temporal resolution of 85 Hz, and was viewed monocularly in a dark room at a distance of 60 cm. It was covered with black paper, except for a circular central aperture with a diameter of 20 cm, in which the stimulus was displayed.

The subject's head was fixed with a chinrest, and the screen was viewed through a 30 cm wide and 60 cm long black tube. This eliminated all environmental illumination.

Stimuli. The stimuli were 161 by 161 pixels grayscale images generated in 3D Studio (Autodesk, 1993), and presented through Matlab 5.0 using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The size of the actual object displayed could vary somewhat across conditions, but the standard object was 5 cm long and 3 cm wide (4.8 by 2.9 visual degrees). The object was gray on a black background. Its 3-D shape was defined by Lambertian shading, and rendered by a virtual camera at a distance of 120 cm, always aimed at the center of the object. The object was illuminated by one virtual light source 20 cm behind, 10 cm above and 5 cm left of the camera. The light intensity dimmed with distance, from full strength at 80 cm until darkness at 200 cm, following an inverse square function.

Yes/No experiment. In the Yes/No experiment and the pixel analysis, a grid of nine stimuli — a three by three orthogonal and equidistant grid — was used for each property–viewpoint combination. The exact stimulus values used were

based on the psychometrical curves, so as to avoid ceiling and floor effects in performance data. The left-bottom point in Fig. 4A was the standard stimulus. All 36 discrimination performances between the nine grid points were assessed to estimate the dimensional low-level correlation we wanted to measure.

Adjustment experiment. For the adjustment experiment, three stimulus spaces were created, one for each combination of View and intrinsic property (Bend–View, Radius–View, and Angle–View). The standard stimulus was used as the central point. To make the adjustment of the probe stimulus, the subject could then navigate through this space, yielding an impression of fluent stimulus manipulation of both dimensions. Test stimuli were drawn randomly from the middle 50% range of the stimulus space to avoid the subject reaching the end of the stimulus space while adjusting the probe object.

Pixel analysis. The same stimuli as in the Yes/No experiment were used.

Procedure

Yes/No experiment. Each combination of View and intrinsic property was programmed as a separate experiment. These experiments were taken in a different order by the three subjects. During each block of trials, the subject had to discriminate two stimuli. Before each block, the subject could study both stimuli and learn to associate the '1' and '2' keys of the NumPad part of the keyboard to each of them. Ten test trials with feedback and unlimited exposure duration were done to ensure the stimulus difference and the key association had been learned. The two stimuli were then presented 25 times each in a random order. Each trial started with a 235 ms central fixation cross, followed by 294 ms of stimulus exposure at a random location on the screen, but with a maximal deviation of 12 pixels from the center of the screen in each direction. A mask consisting of a scramble of 40 by 40 pixel squares randomly drawn from the stimulus set then followed for 235 ms, at the location of the stimulus. Immediately after the subject's answer, the next trial started. No feedback was given. Following each block of 50 trials, the subject could choose to stop the experiment and continue another time. Each experiment consisted of 72 blocks, resulting in 100 trials per data point.

Adjustment experiment. Each combination of View and intrinsic property was programmed as a separate experiment. These experiments were taken in a different order by the three subjects. Every adjustment experiment encompassed 24 blocks of 26 trials, and three conditions of which the order was randomized. A block only contained trials of one condition. In two of these conditions only one dimension was manipulated, and in one both had to be adjusted. In the unidimensional conditions, the irrelevant dimension was not adjustable and was fixed to its standard value. These unidimensional conditions are not reported in this paper, because they are

not relevant in this context (they were used as a validation of the psychometrical curves). In the two-dimensional conditions, both dimensions could be adjusted simultaneously. The condition was announced by a text message before each block.

Every trial consisted of two phases. In the first phase a fixation cross, the test stimulus and a mask were shown, exactly like in the Yes/No experiments. This was immediately followed by the second phase, where the subjects could adjust a probe stimulus with the keyboard's arrow keys. The task was to make the probe stimulus resemble the test stimulus as closely as possible. The probe stimulus, which was visible throughout the adjustment, responded immediately to the observer's changes. Also note that the test stimulus was no longer visible in the second phase. Both relevant dimensions (one intrinsic property and the viewpoint) could be adjusted simultaneously with the up–down and left–right keys respectively; the other dimensions were locked at their standard values. This phase was not time-limited, though the subjects were instructed to respond swiftly. When they had completed the adjustment, the subjects could end the trial by pressing the Enter key. The next trial started immediately after this key press.

Pixel analysis. The pixel analysis used the sum of the absolute values of all pixel luminance differences between two images, corrected to the actual screen luminances, as a measurement of the physical difference between the 2-D images. The analysis was done on the same stimuli as the Yes/No experiment, and the data were processed in the same way. Spatial filtering, when applied, was done after correcting for actual screen luminances. For low-pass filters, we used a range of 0.1 to 15 c/deg as cutoffs, using 502 equidistant intermediate values (on a logarithmical scale). Our band-pass filters were rectangular with a width of 1 octave, and had centers ranging from 0.1 to 10.6 c/deg, using 466 equidistant intermediate values (again, on a logarithmical scale). The 15 c/deg upper limit is the Nyquist frequency. 10.6 c/deg is the highest center of a one octave wide rectangular band-pass filter whose highest frequency does not exceed this Nyquist frequency.

Analysis

Yes/No experiment. The low-level correlation was estimated through a maximum likelihood procedure, by fitting a parallelogram model on the data (see Fig. 4B). The geometrical distances within such a parallelogram represent discrimination performances in d': the longer the side, the easier the discrimination. The restriction of the parallelogram shape stems from the assumption that covariance matrices across two-dimensional space are equal. Thus, all opposite sides are equal in length and all midpoints are exactly in the middle, since the stimulus grid was equidistant as well. Also, the length of a side is equal to the sum of its two segments.

We constructed a large number of models for each viewpoint–property space and each subject. First, we made twenty equidistant manipulations to each dimensional variance within a probable range. This probable range was arbitrarily defined

by the second largest and second smallest measurement of the variance of a given dimension. These measurements were based on all opposite sides of the parallelogram, and all segments of these opposite sides. Then we combined each of these 20×20 models with each of 201 possible correlations between -1 and 1 (up to a precision of 0.01) for a total of 80400 models.

Now we compute the total likelihood of each model. The likelihood of every observed d' given the theoretical d' of a model was computed. We used the formula of Gourevitch and Galanter (1967) as an estimate of the variance of each theoretical d', assuming no bias and a Gaussian sampling distribution of d'. If H is the hit rate, n_H the number of hits, Φ_H the density at $z(H)$, and F the false alarm rate,

$$\sigma_{d'}^2 = H\frac{1-H}{n_H\Phi_H^2} + F\frac{1-F}{n_F\Phi_F^2}. \tag{2}$$

The sum of all 36 loglikelihoods for a given model was then taken as the total loglikelihood of the model. The theoretical correlation of the underlying model with the highest total loglikelihood was taken as the correlation of the data.

To estimate the standard error of our correlation measurements, we used a bootstrap procedure (Efron, 1979). All the original data-points (the hit rates and false alarm rates making up the d' values) were resampled 2000 times, with replacement. We then applied the above procedure on each set of 36 bootstrap samples. We used the resulting distribution of correlations to estimate the standard error.

We will also test the goodness-of-fit of each best-fitting model to the behavioral data. To this end, we perform an Anderson–Darling test (as recommended by Stephens, 1986). In such a test, cumulative distributions are tested for equality. The Anderson–Darling test is more sensitive than the classical Kolmogorov–Smirnov test, especially near the tails. The cumulative distribution we will test here is that of the z-scores of the data points: the discrepancy in d' between data and model prediction, corrected for the expected standard error of the model prediction. This distribution will be compared with the theoretical cumulative z-distribution. If our multidimensional signal detection models fit the data, a cumulative z-distribution should not be rejected as a statistical model for the data.

Adjustment experiment. In the adjustment data, total answer variance may arise due to internal noise, but also due to external variance, since different random stimuli are shown on each trial. Thus, the answer value x_a on dimension X equals the stimulus value x_s plus the internal noise value x_i:

$$x_\mathrm{a} = x_\mathrm{s} + x_\mathrm{i}. \tag{3}$$

Of interest to us is the correlation between the internal noise values of both dimensions X and Y. However, the data only provide us with the covariance

between the total answer values:

$$\begin{aligned} \mathrm{COV}(x_\mathrm{a}, y_\mathrm{a}) &= \mathrm{COV}(x_\mathrm{s} + x_\mathrm{i}, y_\mathrm{s} + y_\mathrm{i}) && (4)\\ &= \mathrm{COV}(x_\mathrm{s}, y_\mathrm{s}) + \mathrm{COV}(x_\mathrm{s}, y_\mathrm{i}) + \mathrm{COV}(x_\mathrm{i}, y_\mathrm{s}) + \mathrm{COV}(x_\mathrm{i}, y_\mathrm{i}). && (5) \end{aligned}$$

$\mathrm{COV}(x_\mathrm{s}, y_\mathrm{s})$ equals 0 because our stimulus generation is random, and $\mathrm{COV}(x_\mathrm{s}, y_\mathrm{i})$ and $\mathrm{COV}(x_\mathrm{i}, y_\mathrm{s})$ equal 0 because anything else is a violation of our assumption of equal Gaussian noise distributions. Thus, we assume that:

$$\mathrm{COV}(x_\mathrm{a}, y_\mathrm{a}) = \mathrm{COV}(x_\mathrm{i}, y_\mathrm{i}). \quad (6)$$

Given $x_\mathrm{a} = x_\mathrm{s} + x_\mathrm{i}$, $\sigma^2_{x_\mathrm{i}} = \sigma^2_{x_\mathrm{a}} - \sigma^2_{x_\mathrm{s}}$, and:

$$r(x_\mathrm{i}, y_\mathrm{i}) = \frac{\mathrm{COV}(x_\mathrm{i}, y_\mathrm{i})}{\sigma_{x_\mathrm{i}}\sigma_{y_\mathrm{i}}} \quad \Longleftrightarrow \quad r(x_\mathrm{i}, y_\mathrm{i}) = \frac{\mathrm{COV}(x_\mathrm{a}, y_\mathrm{a})}{\sqrt{\sigma^2_{x_\mathrm{a}} - \sigma^2_{x_\mathrm{s}}}\sqrt{\sigma^2_{y_\mathrm{a}} - \sigma^2_{y_\mathrm{s}}}}. \quad (7)$$

This allows us to compute the internal noise correlation based on observable data. Again we estimate the variance of the measurement using a bootstrap procedure. This time the raw adjustment answers are resampled.

Pixel analysis. The same analysis used on the Yes/No data was applied, using the sum of absolute pixel luminance differences instead of d' as a discrimination measurement.

Results and discussion

We assessed low-level correlations between object properties and viewpoint using both Yes/No and adjustment tasks. The results can be seen in Table 1. We found variable low-level correlations across subjects, but they were nonetheless generally consistent across tasks. When plotting the results of both behavioral tasks against each other (Fig. 5) it can be seen that they are highly linearly related ($r = 0.96$), with a slope close to 1 (0.92) and an intercept close to 0 (-0.11). In other words, they are very similar. Unfortunately, for practical reasons, two of the datasets could not be completed: Radius PZ and Angle BW, both Yes/No data. Therefore, we could not include them in the comparative linear plot.

The results of our Anderson–Darling goodness-of-fit tests for the Yes/No models, also listed in Table 1, led to the rejection of two models. Bend MD ($A^2 = 3.06$, $p = 0.03$) and Radius MD ($A^2 = 2.93$, $p = 0.03$) both deviated more from the theoretical z-distribution than was to be expected from sampling errors. Smaller p-values indicate worse fits. This does not mean that the results are meaningless, but they should be taken as unreliable. It might for instance explain why Radius MD is the only dataset where the Yes/No correlation is quite different from the adjustment one. As a summary graph, Fig. 6 shows the fit of the cumulative z-values of all data to the expected cumulative z-distribution ($A^2 = 2.00$, $p = 0.09$). The biggest lack of fit is seen near the lower end of the curve, but this is caused mainly by the

Table 1.
Overview of correlation data, goodness of fit and standard errors

Property	Pixel correlation	Subject	Yes/No correlation	Goodness of fit	Adjustment correlation
Bend	$r = 0.16$	BW	$r = 0.40$ $SE = 0.11$	$A^2 = 0.81$ $p = 0.47$	$r = 0.47$ $SE = 0.14$
		PZ	$r = 0.08$ $SE = 0.12$	$A^2 = 0.47$ $p = 0.78$	$r = 0.15$ $SE = 0.24$
		MD	$r = -0.38$ $SE = 0.22$	$A^2 = 3.06$ $p = 0.03$	$r = -0.41$ $SE = 0.13$
Radius	$r = 0.05$	BW	$r = -0.31$ $SE = 0.12$	$A^2 = 0.31$ $p = 0.93$	$r = -0.34$ $SE = 0.11$
		PZ	$r =$ N/A $SE =$ N/A	$A^2 =$ N/A $p =$ N/A	$r = 0.35$ $SE = 0.11$
		MD	$r = -0.33$ $SE = 0.16$	$A^2 = 2.93$ $p = 0.03$	$r = 0.07$ $SE = 0.12$
Angle	$r = -0.57$	BW	$r =$ N/A $SE =$ N/A	$A^2 =$ N/A $p =$ N/A	$r = -0.84$ $SE = 0.03$
		PZ	$r = -0.97$ $SE = 0.02$	$A^2 = 0.51$ $p = 0.74$	$r = -0.92$ $SE = 0.09$
		MD	$r = -0.99$ $SE = 0.02$	$A^2 = 0.36$ $p = 0.89$	$r = -0.90$ $SE = 0.03$

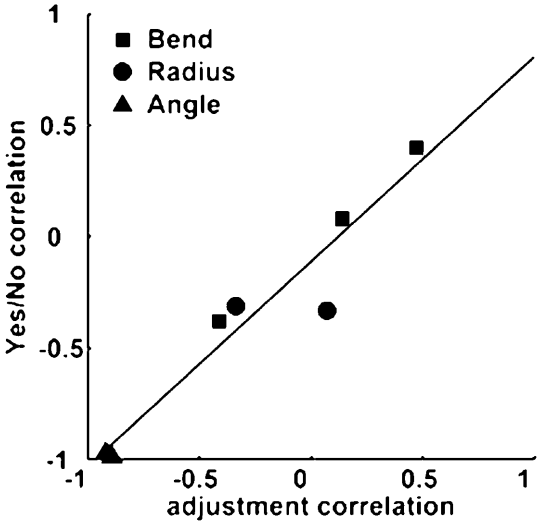

Figure 5. Scatterplot and linear fit to compare the low-level correlations measured by Yes/No and adjustment tasks.

considerable lack of fit in that region in the Bend MD data set. Without it, the fit is much better ($A^2 = 1.15$, $p = 0.29$).

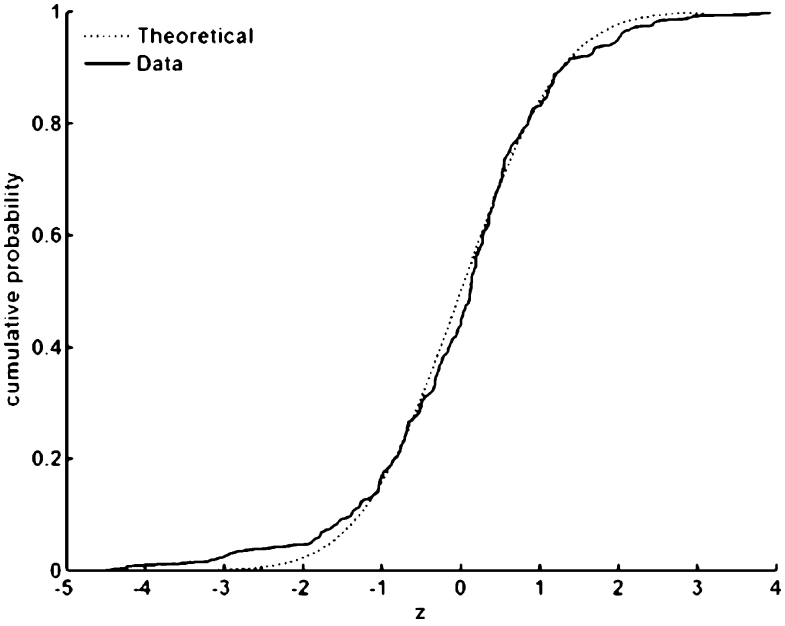

Figure 6. Graphical illustration of the goodness-of-fit of our low-level correlation models. In this overview graph, the cumulative distribution of the observed z-values of all models is compared with the theoretical cumulative z-distribution. It can be seen that the fit is good, with the greatest lack of fit around the lower end of the curve.

The pixel analysis (see Table 1) results follow the pattern we had envisioned when devising the stimulus: a strong correlation for Angle, less so for Bend and almost absent for Radius. Because the individual differences observed in the Bend and Radius conditions allow for little direct comparison with the pixel analysis results, we will consider only the Angle condition for the spatial frequency filtering analysis. Without spatial frequency filtering, this condition resulted in a pixel correlation that was markedly more moderate than measured behaviorally ($r = -0.57$). Figure 7 shows the results of our low-pass and band-pass filtering on the input images for Angle. It can be seen that retaining lower frequencies generally yields stronger negative correlations, though never as strong as those measured behaviorally. The difference is still very significant ($p < 0.001$) for the Yes/No and the spatially filtered pixel correlation that differed the least.

Thus, we found a good correspondence in estimated low-level correlations between Yes/No and adjustment tasks. What makes this correspondence even more remarkable is that the consistency across tasks exists despite large individual differences. As long as these individual differences are constant over time, as they seem to be, they do not pose a problem for our research. It also shows that the first stage of our framework (the input image) does not entirely determine the measured low-level correlation. It is well possible that differential attention for different parts or aspects of the stimuli is responsible for the individual differences found.

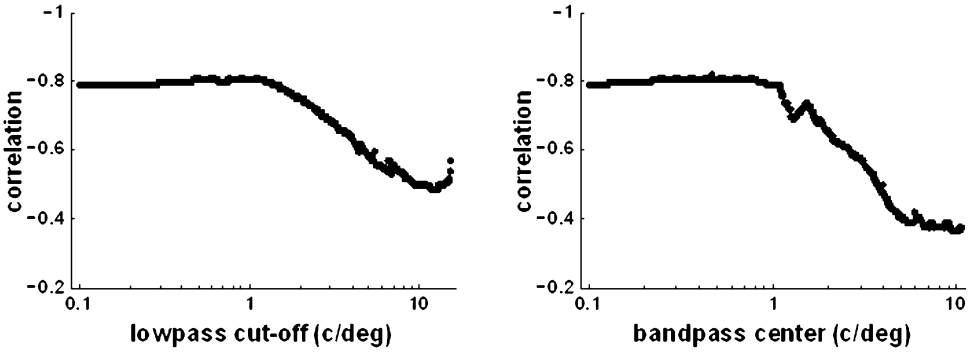

Figure 7. Pixel based low-level correlations for Angle in our stimuli after spatial frequency filtering. The left panel shows the results for low-pass filtering, the right panel for band-pass filtering.

The pixel analysis showed that the 2-D projection of a 3-D object does contain some strong correlations in itself, following the pattern we had expected: a strong correlation with View for Angle, less so for Bend and almost absent for Radius. But unfortunately, the behavioral data allow for little comparison with the pixel analysis due to the observed individual differences. The case of Angle is an exception, however, because all correlations between Angle and View are strongly negative. Here, it can be seen that the behavioral correlations are stronger than the pixel correlations of the 2-D projection in all three subjects: the input image is not enough, by itself, to explain the low-level correlations measured. The pixel correlations can be considerably higher when applying spatial frequency filtering though, especially when the lower frequencies are retained. This was to be expected, since blurring the images obscures the differences by which Angle and View manipulations can be disentangled. Still, we see two reasons to dismiss the alternative explanation that spatial frequency filtering causes the behavioral Angle correlations to be stronger than the pixel-based correlations. First, even at the lowest frequencies the correlation was never strong enough to match the behavioral Yes/No correlation observed. It did not differ significantly from some of the adjustment correlations measured, but we focus on the Yes/No task because that is the task used by the pixel analysis. Second, and more importantly, these very low frequencies are not what humans typically use to identify objects. For complex stimuli like faces, spatial frequencies around 2–5 c/deg are more relevant (Gold *et al.*, 1999). It can be seen in Fig. 7 that the pixel analysis results for this spatial frequency range are not very different from what was found without spatial frequency filtering. Thus, we have found no support for the alternative hypothesis that spatial frequency filtering can explain the differences between behavioral and pixel analysis results for the Angle manipulations.

In sum, we have measured consistent low-level correlations within subjects, and will now proceed to check to what extent they can cause viewpoint-dependent object recognition.

EXPERIMENT 2

We have concluded that viewpoint and property dimensions can be correlated prior to object recognition itself. But are they separable, i.e. can they be judged independently even if their noise distributions are correlated? In other words, can the metrical status of an object property be assessed irrespective of viewpoint?

We will measure the decision rule while instructing subjects to ignore viewpoint, and then compare the slope of the linear decision rule to the low-level correlation that was measured earlier without abstracting viewpoint. Finding a decision rule implies doing a categorization task, since a decision rule subdivides the space in two categories. We will define these categories by means of two prototype objects in the same viewpoint, differing on one intrinsic stimulus dimension only. Then we will present the subjects with stimuli drawn in a systematic manner from the entire stimulus space (see below), and ask them to judge which prototype object each stimulus resembles most, abstracting viewpoint. Thus, we will use a categorization task of 3-D objects, as opposed to the discrimination task of 2-D images that we used to measure low-level correlations. Given this and the assumptions we made, the slope of the categorization rule will be equal to the low-level correlation if full dimensional integrality exists (see Appendix).

Assuming Gaussian internal noise distributions, the categorization rule is, as noted, linear. A linear categorization rule implies a linear weighted combination of dimensions. We will estimate these internal weights by adding external Gaussian noise to the prototype objects, a procedure similar to the work of, for instance, Burgess and Barlow (1983) and perturbation analysis (e.g. Landy *et al.,* 1995), but to our knowledge never applied in research on viewpoint dependence. The slope of the decision rule can be computed based on these weights.

We do this as follows. The decision rule is a weighted linear combination of dimensions X and Y. For all perceptual effects (x, y), the decision will be based on a decision variable v defined by:

$$v = \omega_x x + \omega_y y \tag{8}$$

$$\Longleftrightarrow \quad -\omega_y y = \omega_x x - v \tag{9}$$

$$\Longleftrightarrow \quad y = -\frac{\omega_x}{\omega_y} x + \frac{v}{\omega_y}. \tag{10}$$

The decision variable v is monotonically related to the likelihood ratio. It is the 1-D decision space that is mentioned in Fig. 1D as the final step of our framework. Our aim is to retrieve the weights ω_x and ω_y attributed to dimensions X and Y, so we can compute the slope of the decision rule and compare it to the low-level correlation. We know that:

$$\sigma_v^2 = \sigma^2_{\omega_x x + \omega_y y} = \omega_x^2 \sigma_x^2 + \omega_y^2 \sigma_y^2 + 2\omega_x \omega_y \,\mathrm{COV}(x, y), \tag{11}$$

where σ_x^2 and σ_y^2 are the internal variances for all points (x, y) on dimensions X and Y, and $\mathrm{COV}(x, y)$ is the internal covariance between them.

Now we add external (stimulus) variance on one dimension to the equation:

$$\sigma_v^2 = \omega_x^2(\sigma_x^2 + \sigma_{nx}^2)+\omega_y^2\sigma_y^2 + 2\omega_x\omega_y \,\mathrm{COV}(x, y), \tag{12}$$

where σ_{nx}^2 is the amount of the external variance added to dimension X.

If we manipulate but σ_{nx}^2 and keep the rest constant this results in:

$$\sigma_v^2 = \omega_x^2\sigma_x^2 + \omega_x^2\sigma_{nx}^2 + \omega_y^2\sigma_y^2 + 2\omega_x\omega_y \,\mathrm{COV}(x, y) \tag{13}$$

$$\Longleftrightarrow \quad \sigma_v^2 = \omega_x^2\sigma_{nx}^2 + c, \tag{14}$$

where $c = \omega_x^2\sigma_x^2 + \omega_y^2\sigma_y^2 + 2\omega_x\omega_y \,\mathrm{COV}(x, y)$ and constant.

Thus, we will find a linear dependence between total answer variance and external noise. The total answer variance is derived from the d' measure of discriminability of the two prototype objects with added external variance. The square root of the slope of this function will give us an unsigned estimate of the weight of the dimension that had external noise added to it. The situation is similar for dimension Y. The sign of the slope we derive by manipulating external covariance:

$$\sigma_v^2 = \sigma_{\omega_x x}^2 + \sigma_{\omega_y y}^2 + 2\omega_x\omega_y[\mathrm{COV}(x, y) + \mathrm{COV}(nx, ny)], \tag{15}$$

where $\mathrm{COV}(nx, ny)$ is the external covariance added between dimensions X and Y.

Similar to the above equations we find:

$$\sigma_v^2 = \sigma_{\omega_x x}^2 + \sigma_{\omega_y y}^2 + 2\omega_x\omega_y \,\mathrm{COV}(x, y) + 2\omega_x\omega_y \,\mathrm{COV}(nx, ny) \tag{16}$$

$$\Longleftrightarrow \quad \sigma_v^2 = 2\omega_x\omega_y \,\mathrm{COV}(nx, ny) + c. \tag{17}$$

Thus, comparing equations (10) and (17), the sign of the slope of the decision rule is opposite to the sign of the linear function that connects σ_v^2 and $\mathrm{COV}(nx, ny)$.

On the face of it, subjects will be presented with a simple categorization task: attribute an object in a two-dimensional space to either of two prototype objects, ignoring viewpoint. But underlying this simple task, each stimulus presented is actually one of the two prototype objects with added external (co)variance and will be scored as such. This does indeed imply that a subject's categorization answer that is physically correct could be scored as incorrect if the originating prototype object (before adding external variance) is physically closer to the stimulus than the subject's answer. The external variance is added through random sampling from a Gaussian distribution, using the (co)variance required by the condition. As shown, this will allow us to compute the linear categorization rule.

Methods

Subjects. The same three subjects described in the first experiment participated.

Apparatus. The same apparatus described in the first experiment was used.

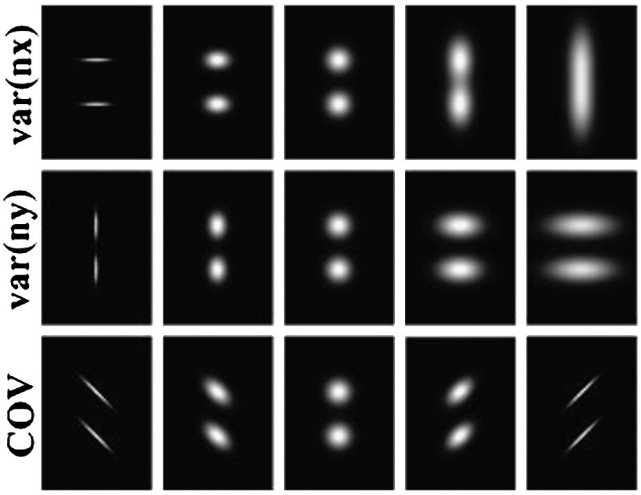

Figure 8. Illustration of our 15 conditions of added external (co)variance. The two distributions are centered around our two prototype objects, differing on one object property dimension but not in viewpoint. var(nx) are the possible external noise distributions added in the conditions where the variance on intrinsic property dimensions was manipulated, var(ny) those on the viewpoint dimension, and COV illustrates the external noise distributions added in the conditions where the external covariance was manipulated.

Stimuli. The stimuli were generated in the same way as in the other experiments. Each property–viewpoint combination was an experiment on its own, and defined a two-dimensional 39 by 39 stimulus space. Both dimensions were on the same scale: they both subtended a range of 13 internal noise standard deviations. The center of this stimulus space was the standard stimulus. Note that while this is a considerable range of stimuli, most stimuli presented were close to the standard stimulus. This is because we sample our stimuli from Gaussian external noise distributions, whose standard deviations were never bigger than three times the internal standard deviation, and whose centers were at the above mentioned prototype stimuli. These two prototype objects were defined as being in the standard viewpoint, but one standard deviation of internal noise removed in each direction from the standard value of the intrinsic property concerned. External (co)variance was added to each stimulus by sampling randomly from a Gaussian external noise distribution defined by the condition. Figure 8 illustrates the 15 external noise conditions we used (further detailed in the Procedure section). By drawing different categorization slopes on these figures, it can be seen intuitively how these conditions yield different results for different decision rules. They reveal to us the underlying slope, as the equations above show.

Procedure

The task was similar to the earlier Yes/No experiment. Each property–viewpoint combination was a separate experiment of 60 blocks of 100 trials, resulting in

400 trials per condition. There were five possible manipulations of each dimension's external variance (0.01, 0.5, 1, 3 and 9 times the measured internal variance) and five manipulations of external covariance ($r = -0.99, -0.5, 0, 0.5$ and 0.99). Thus, we had three independent variables with five levels each. This results in a total of 15 conditions. Only one independent variable was manipulated at a time for a given trial, keeping the other two independent variables at their standard value. The standard value was always the middle level of an independent variable: one time the internal variance for the external variance manipulation, and a covariance of zero for the external covariance manipulation. The actual stimulus presentation and response was exactly the same as in the previous Yes/No experiment, but this time the task was one of categorization instead of discrimination. Prior to each block, the subject could study the prototype objects, and then had to judge which prototype object the stimuli resembled most, abstracting the viewpoint. The prototype objects were the same in all blocks in a given experiment. Therefore, the exercise trials at the beginning of each block were omitted. All 15 conditions were randomized on a trial-to-trial basis across all blocks and thus hidden for the subjects.

Analysis

The analysis was done as described above, by finding the linear functions that describe the relationships between total answer variance and external (co)variance best, and using their slopes to estimate the slope of the decision rule. The linear fitting on the data was not done with a simple linear regression, but the standard error of the datapoints was taken into account by choosing the linear fit with the smallest χ^2 statistic. We will also test the goodness-of-fit of our data to these predicted linear relationships using this χ^2 goodness-of-fit statistic. In such a test, the sum of the squared z-scores is assumed to be χ^2 distributed with n-p degrees of freedom, where n is the number of datapoints and p the number of underlying parameters. We calculated this statistic for each model of 15 datapoints (three conditions with five datapoints each).

Ultimately, the result of interest is the slope of a linear fit of categorization slopes versus low-level correlations. We want to compare this result to two possible outcomes of theoretical interest: a slope of 1 (full determination of categorization by the low-level correlation) and 0 (full abstraction of viewpoint). To allow for a statistical test, we again used a bootstrap procedure (Efron, 1979). All the original datapoints were resampled 2000 times, with replacement. We then applied all the above procedures on each of these simulations, and calculated the resulting slope of a linear fit of categorization slopes versus low-level correlations. We used the resulting distribution of bootstraps to calculate the standard error of our observations and to perform a significance test.

Because any difference between the 2-D discrimination and the 3-D object categorization task could be due to task or stimulus differences, we also used the sum of absolute pixel luminance differences as input data to this experiment. If by not abstracting the viewpoint one does indeed use a categorization slope equal to the

low-level correlation — as theoretically predicted (see Appendix) — then the result of this pixel analysis should show a perfect correspondence with the pixel analysis on the 2-D task. We will not use low-pass or band-pass filters here, because we are only interested in excluding task or stimulus differences as an explanation for our results, rather than comparing behavioral and pixel analysis results.

Results and discussion

Figure 9 shows an example of the raw Bend data for subject MD. It can be clearly seen that total answer variance does indeed rise with a different slope on both dimensions. We use these slopes to compute the slope of the decision rule, as is explained above.

In Fig. 10 and Table 2, the results are compared to the mean of the low-level correlations previously measured using adjustment and Yes/No tasks. The correlation is high ($r = 0.90$). The slope of 0.66, however, points to a certain degree of viewpoint abstraction. First, we tested it against the hypothesis that there is no viewpoint dependence (slope $= 0$). The difference was significant ($p < 0.01$). Then, we tested it against the hypothesis that the viewpoint dependence is fully determined by the low-level correlations (slope $= 1$). Again, the difference was significant ($p = 0.04$).

The pixel analysis on this task yielded categorization slopes that were exactly equal to the low-level pixel correlations, as theoretically expected when no attempt to abstract viewpoint is made. This rules out the alternative explanation that task or stimulus differences are causing the behavioral results, rather than properties of the visual system.

The results of the χ^2 goodness-of-fit tests can be seen in Table 2. Smaller p-values indicate worse fits. No fits were rejected on a 0.05 significance level, meaning that

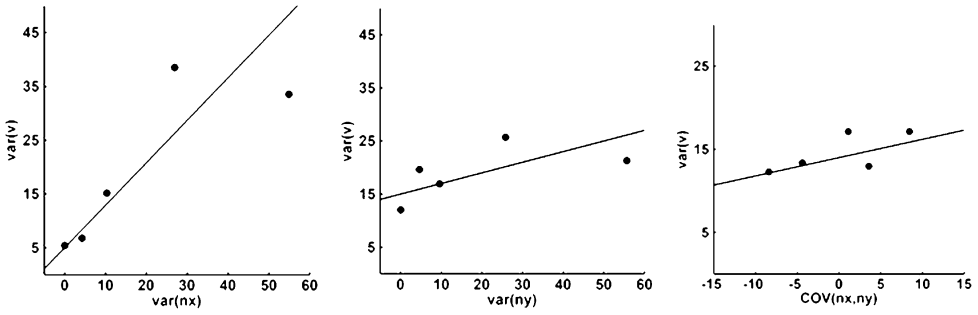

Figure 9. An example of the data, in this case the Bend dataset of subject MD. Note that the units used for variance are in terms of the stimulus matrix, and thus arbitrary. In the case of the X-dimension (Bend), there is a steep slope in the linear relationship between total answer variance and external variance, pointing to a high weight to that dimension. The viewpoint dimension on the other hand has a lower weight attached to it, as can be seen by the shallow slope in the middle panel. This will result in a viewpoint-dependent decision rule, as is confirmed by the external covariance data. The covariance fit also yields the sign of the slope: opposite to the slope of this linear fit, thus negative.

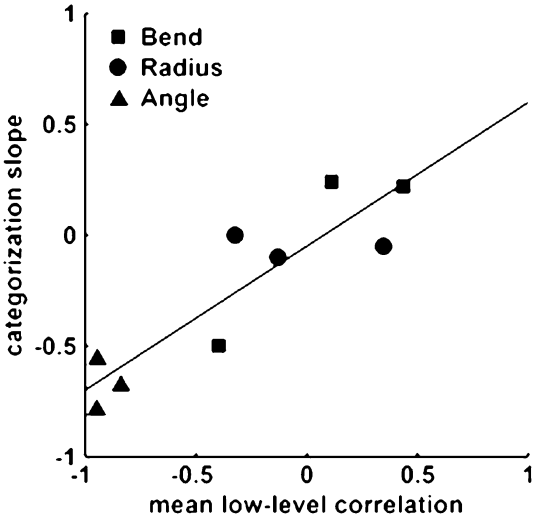

Figure 10. Scatterplot and linear fit of the mean behavioral low-level correlations (the mean of the adjustment and Yes/No data) versus the categorization slopes measured. A high correlation, but moderately correspondent slope can be seen.

Table 2.
Comparison of correlations and categorization slopes

Property	Subject	Mean correlation	Decision rule	Goodness of fit
Bend	BW	$r = 0.44$	$sl = 0.22$ $SE = 0.14$	$\chi^2 = 9.04$ $p = 0.43$
	PZ	$r = 0.11$	$sl = 0.24$ $SE = 0.16$	$\chi^2 = 12.27$ $p = 0.20$
	MD	$r = -0.40$	$sl = -0.50$ $SE = 0.17$	$\chi^2 = 14.08$ $p = 0.12$
Radius	BW	$r = -0.32$	$sl = 0$ $SE = 0.10$	$\chi^2 = 8.71$ $p = 0.46$
	PZ	$r = 0.35$	$sl = -0.05$ $SE = 0.08$	$\chi^2 = 4.77$ $p = 0.85$
	MD	$r = -0.13$	$sl = -0.10$ $SE = 0.17$	$\chi^2 = 13.11$ $p = 0.16$
Angle	BW	$r = -0.84$	$sl = -0.68$ $SE = 0.19$	$\chi^2 = 9.40$ $p = 0.40$
	PZ	$r = -0.95$	$sl = -0.79$ $SE = 0.13$	$\chi^2 = 10.49$ $p = 0.31$
	MD	$r = -0.94$	$sl = -0.56$ $SE = 0.21$	$\chi^2 = 6.90$ $p = 0.65$

linear functions could adequately describe the relationships between total answer variance and external (co)variance, as was theoretically expected.

It is clear that viewpoint independence could not be reached. More importantly, we found a strong positive correlation between low-level correlations and categorization rules, pointing to a linear dependence. The categorization slopes are moderated, however, compared to the low-level correlations. This indicates that the visual system can make some abstraction of viewpoint, even in briefly displayed 2-D images, and despite apparent confounds between viewpoint and intrinsic object property changes in the 2-D projection and early processing of these projections. The viewpoint dimension could not be fully separated, however, causing viewpoint dependence that is to be brought back to low-level correlations.

These results have some theoretical implications.

First, we have shown that viewpoint dependence found using metrically manipulated stimuli is not necessarily to be ascribed to properties of object recognition itself. Factors prior to object recognition were able to explain nearly all viewpoint dependence found in our object recognition experiment ($r = 0.90$). As such, these often-used stimuli may not be completely suitable for research on the workings of object recognition through its failure to attain viewpoint invariance. However, we did not base these studies on a specific theory on viewpoint dependence in object recognition, and neither did we mean to test one. Our data are compatible with both viewpoint-dependent and viewpoint-independent theories of object recognition. Likewise, we do not deny that other tasks and stimuli can and will yield viewpoint-dependent recognition performance that cannot be due to these low-level correlations we measured. We have shown them to be relevant though for abstract, metrically manipulated stimuli within a small range of viewpoints and property manipulations. Whether this can be generalized to less homogenous stimuli and larger viewpoint ranges, and how far the generalization could reach, is an empirical question. We do not think our models will continue to hold, because several of our assumptions (for instance, equal covariance matrices) will then be violated. However, the general tendencies could remain equivalent as long as the type of stimulus remains the same — metrically manipulated.

Second, we confirmed Tjan and Legge's (1998) finding that stimulus information limits the viewpoint independence of object recognition, and modelled and measured one possible specific cause for it. We also found confirmation for Tjan *et al.*'s (1995) suspicion that visual processing of 2-D image properties is likely to be an important bottleneck for object recognition as well, by finding higher correlations at this level than were present in the stimulus image. The finding that object recognition attains some viewpoint independence despite low-level correlations means that low-level constraints do not necessarily define a strict upper limit on viewpoint independence. Tjan and Legge (1998) already found a similar result using multiple-part objects, and hypothesized (in their view-rate hypothesis) that the visual system can select the most viewpoint-invariant components of an object when time for recognition is limited. It will then gain speed and lose overall accuracy — but the accuracy will be less dependent on viewpoint. We do not feel this explanation is appropriate here though, because our objects were abstract and homogeneous, and did not

consist of clearly identifiable components. Still, in a more general sense this explanation can still hold. The visual system could use its input information differently depending on the task by, for instance, strategically focusing more on information that allows separation of viewpoint, even if it is at the cost of some sensitivity to changes on the relevant object dimension. As an example, global shading information can be very informative to retrieve 3-D object structure, while of no special use when 2-D images are to be discriminated. But although optimally using available information may allow object recognition to perform better than is to be expected on the basis of low-level constraints, our results and previous research have shown that these constraints continue to play a role despite attempts to overcome them.

In conclusion, both information content of the 2-D images and their early feature discrimination determine the low-level correlations that we measured. For the metrically manipulated stimuli we used, the low-level correlations can account for nearly all viewpoint dependence measured at the object recognition level. However, they do not pose a strict upper limit to object recognition, since the object recognition system manages to partly abstract viewpoint despite them.

Acknowledgements

This research was supported by an interdisciplinary research grant from the University Research Council (IDO/98/002) and by the Concerted Research Effort Convention GOA of the Research Fund K.U. Leuven (GOA/2005/03-TBA). We would like to thank Bert Willems and Peter Claessens for their valuable help in setting up and discussing these studies, and two anonymous reviewers and the editors for their helpful comments.

REFERENCES

1. Ashby, G. F. and Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli, *J. Exper. Psychol.: Human Perception and Performance* **14**, 33–53.
2. Ashby, G. F. and Townsend, J. T. (1986). Varieties of perceptual independence, *Psychol. Rev.* **93**, 154–179.
3. Autodesk, Inc. (1993). *Autodesk 3D Studio* (Release 3). Sausolito, CA.
4. Biederman, I. (1987). Recognition-by-components: a theory of human image understanding, *Psychol. Rev.* **94**, 115–147.
5. Biederman, I. and Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming, *Perception* **20**, 585–593.
6. Biederman, I. and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance, *J. Exper. Psychol.: Human Perception and Performance* **19**, 1162–1182.
7. Biederman, I. and Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: reply to Tarr and Bülthoff (1995), *J. Exper. Psychol.: Human Perception and Performance* **21**, 1506–1514.
8. Brainard, D. H. (1997). The psychophysics toolbox, *Spatial Vision* **10**, 443–446.
9. Bülthoff, H. H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition, *Proc. Nat. Acad. Sci. USA* **89**, 60–64.

10. Burgess, A. and Barlow, H. B. (1983). The precision of numerosity discrimination in arrays of random dots, *Vision Research* **23**, 811–820.
11. Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Annals of Statistics* **7**, 1–26.
12. Foster, D. H. and Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views, *Proc. Roy. Soc. London: Series B* **269**, 1939–1947.
13. Gold, J., Bennett, P. J. and Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers, *Vision Research* **39**, 3537–3560.
14. Gourevitch, V. and Galanter, E. (1967). A significance test for one parameter isosensitivity functions, *Psychometrika* **32**, 25–33.
15. Hayward, W. G. and Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition, *J. Exper. Psychol.: Human Perception and Performance* **23**, 1511–1521.
16. Hayward, W. G. and Williams, P. (2000). Viewpoint dependence and object discriminability, *Psychol. Sci.* **11**, 7–12.
17. Landy, M. S., Maloney, L. T., Johnston, E. B. and Young, M. J. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision Research* **35**, 389–412.
18. Maddox, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization, *Perception and Psychophysics* **63**, 1183–1200.
19. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies, *Spatial Vision* **10**, 437–442.
20. Rock, I. and DiVita, J. (1987). A case of viewer-centered object perception, *Cognit. Psychol.* **19**, 280–293.
21. Stankiewicz, B. (2002). Empirical evidence for independent dimensions in the vsiual representation of three-dimensional shape, *J. Exper. Psychol.: Human Perception and Performance* **28**, 913–932.
22. Stephens, M. A. (1986). Tests based on EDF statistics, in: *Goodness-of-fit Techniques*, R. B. D'Agostino and M. A. Stephens (Eds), pp. 195–234. Marcel Dekker, New York.
23. Tanner, W. P. (1956). Theory of recognition, *J. Acoustic. Soc. Amer.* **28**, 882–888.
24. Tarr, M. J. (1995). Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects, *Psychonomic Bull. Rev.* **2**, 55–82.
25. Tarr, M. J. and Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993), *J. Exper. Psychol.: Human Perception and Performance* **21**, 1494–1505.
26. Tarr, M. J., Williams, P., Hayward, W. G. and Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent, *Nature Neurosci.* **1**, 275–277.
27. Tjan, B. S. and Legge, G. E. (1998). The viewpoint complexity of a object-recognition task, *Vision Research* **38**, 2335–2350.
28. Tjan, B. S., Braje, W. L., Legge, G. E. and Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise, *Vision Research* **35**, 3053–3069.
29. Vanrie, J., Willems, B. and Wagemans, J. (2001). Multiple routes to object matching from different viewpoints: mental rotation versus invariant features, *Perception* **30**, 1047–1056.
30. Wagemans, J., Van Gool, L. and Lamote, C. (1996). The visual system's measurement of invariants need not itself be invariant, *Psychol. Sci.* **7**, 232–236.
31. Wichmann, F. and Hill, N. J. (2001). The psychometric function I. Fitting, sampling, and goodness of fit, *Perception and Psychophysics* **63**, 1293–1313.
32. Willems, B. and Wagemans, J. (2001). Matching multicomponent objects from different viewpoints: mental rotation as normalization? *J. Exper. Psychol.: Human Perception and Performance* **27**, 1090–1151.

APPENDIX

This appendix will detail how the decision rule is determined, why it is equivalent to a linear combination of dimensions, and why its slope should be equal to the low-level correlation if no separability of dimensions exists (given the assumptions described in the text).

Suppose a feature space with two dimensions X and Y, and in it two two-dimensional Gaussian distributions $S1$ and $S2$, with equal covariance matrices but different means. A subject observes a perceptual effect s, and must attribute it to one of the two distributions using a decision rule. The likelihood ratio LR is then the ratio of the likelihoods of s belonging to each S, and relates monotonically to the decision variable v that a subject will use to make a decision.

$$LR = \frac{f(s(x, y)|S1)}{f(s(x, y)|S2)}, \quad \text{where } [S1 \sim N^2(\mu_{x1}, \mu_{y1}, \sigma_x, \sigma_y); \; S2 \sim N^2(\mu_{x2}, \mu_{y2}, \sigma_x, \sigma_y)]$$

$s(x, y)$ is the perceptual effect of the stimulus on feature dimensions X and Y

μ_{x1} is the mean of $S1$ on dimension X; likewise for μ_{x2}, μ_{y1}, μ_{y2}

σ_x is the standard deviation of $S1$ and $S2$ on dimension X; likewise for σ_y

ρ is the correlation between dimensions X and Y in $S1$ and $S2$ $z_{x1} = \frac{x-\mu_{x1}}{\sigma_x}$; likewise for z_{x2}, z_{y1}, z_{y2}.

$$\iff \quad LR = \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{\frac{-[z_{x1}^2+z_{y1}^2-2\rho z_{x1}z_{y1}]}{2[1-\rho^2]}\right\}}{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{\frac{-[z_{x2}^2+z_{y2}^2-2\rho z_{x2}z_{y2}]}{2[1-\rho^2]}\right\}} \tag{A1}$$

$$\iff \quad v \approx \ln(LR) = \frac{-[z_{x1}^2 + z_{y1}^2 - 2\rho z_{x1}z_{y1}]}{2(1-\rho^2)} + \frac{z_{x2}^2 + z_{y2}^2 - 2\rho z_{x2}z_{y2}}{2(1-\rho^2)} \tag{A2}$$

$$\iff \quad 2v(1-\rho^2) = \frac{-\mu_{x1}^2 + \mu_{x2}^2 - 2x(\mu_{x2}-\mu_{x1})}{\sigma_x^2} + \frac{-\mu_{y1}^2 + \mu_{y2}^2 - 2y(\mu_{y2}-\mu_{y1})}{\sigma_y^2} - 2\rho\left[\frac{-\mu_{x1}\mu_{y1} + \mu_{x2}\mu_{y2} - x(\mu_{y2}-\mu_{y1}) - y(\mu_{x2}-\mu_{x1})}{\sigma_x\sigma_y}\right] \tag{A3}$$

$$\iff \quad 2v(1-\rho^2) = \frac{-2x\Delta_x}{\sigma_x^2} + 2\rho\left(\frac{x\Delta_y}{\sigma_x\sigma_y}\right) - \frac{2y\Delta_y}{\sigma_y^2} + 2\rho\left(\frac{y\Delta_x}{\sigma_x\sigma_y}\right) + c, \tag{A4}$$

where

$$c = \frac{-\mu_{x1}^2 + \mu_{x2}^2}{\sigma_x^2} + \frac{-\mu_{y1}^2 + \mu_{y2}^2}{\sigma_y^2} - 2\rho\left(\frac{-\mu_{x1}\mu_{y1} + \mu_{x2}\mu_{y2}}{\sigma_x\sigma_y}\right)$$

is constant and $\Delta_x = \mu_{x2} - \mu_{x1}$; $\Delta_y = \mu_{y2} - \mu_{y1}$

$$\iff \quad v = x\left[\frac{-\Delta_x\sigma_y + \rho\Delta_y\sigma_x}{\sigma_x^2\sigma_y(1-\rho^2)}\right] + y\left[\frac{-\Delta_y\sigma_x + \rho\Delta_x\sigma_y}{\sigma_y^2\sigma_x(1-\rho^2)}\right] + c. \tag{A5}$$

Thus, a decision rule whose boundary connects points of equal likelihood is a linear combination of dimensions with weights $\omega_x = \frac{-\Delta_x\sigma_y+\rho\Delta_y\sigma_x}{\sigma_x^2\sigma_y(1-\rho^2)}$ and $\omega_y = \frac{-\Delta_y\sigma_x+\rho\Delta_x\sigma_y}{\sigma_y^2\sigma_x(1-\rho^2)}$.

To use an optimal decision rule, this rule must connect all points with a likelihood ratio of 1, implying a $\ln(LR)$ of 0 (Ashby and Gott, 1988):

$$\iff \quad 0 = x\left[\frac{-\Delta_x\sigma_y + \rho\Delta_y\sigma_x}{\sigma_x^2\sigma_y(1-\rho^2)}\right] + y\left[\frac{-\Delta_y\sigma_x + \rho\Delta_x\sigma_y}{\sigma_y^2\sigma_x(1-\rho^2)}\right] + c \tag{A6}$$

$$\iff \quad y = \left[\frac{\sigma_y^2\sigma_x(1-\rho^2)}{\Delta_y\sigma_x - \rho\Delta_x\sigma_y}\right]\left[\frac{-\Delta_x\sigma_y + \rho\Delta_y\sigma_x}{\sigma_x^2\sigma_y(1-\rho^2)}\right]x + \left[\frac{\sigma_y^2\sigma_x(1-\rho^2)}{\Delta_y\sigma_x - \rho\Delta_x\sigma_y}\right]c \tag{A7}$$

$$\iff \quad y = \left[\frac{\sigma_y}{\Delta_y\sigma_x - \rho\Delta_x\sigma_y}\frac{-\Delta_x\sigma_y + \rho\Delta_y\sigma_x}{\sigma_x}\right]x + c', \tag{A8}$$

$$\text{where } c' = c\left[\frac{\sigma_y^2\sigma_x(1-\rho^2)}{\Delta_y\sigma_x - \rho\Delta_x\sigma_y}\right]$$

$$\iff \quad y = \left[\frac{-\Delta_x\sigma_y^2 + \mathrm{COV}(x,y)\Delta_y}{\Delta_y\sigma_x^2 - \mathrm{COV}(x,y)\Delta_x}\right]x + c' \quad \text{since } \rho = \frac{\mathrm{COV}(x,y)}{\sigma_x\sigma_y}. \tag{A9}$$

Thus, this decision rule is linear, and its slope can be computed from the weights attributed to the dimensions. Suppose the situation of our object categorization experiment: the prototype objects are in the same view (dimension X), but differ on physical object property dimension Y by an amount we will arbitrarily fix at 1 here. This is valid because units of physical stimulus difference are always arbitrarily defined, and because they cancel out in the equation. Likewise, we arbitrarily equate both internal variances to 1. Note that while their exact value is arbitrary, their being equal is not. This was accomplished by calibrating both dimensions beforehand to be on the same psychological scale. The slope of the categorization rule will then be equal to the low-level correlation if the above dimensional weights are used:

$$\textit{slope} = \frac{-\Delta_x\sigma_y^2 + \mathrm{COV}(x,y)\Delta_y}{\Delta_y\sigma_x^2 - \mathrm{COV}(x,y)\Delta_x} = \frac{0 + \rho.1.1.1}{1 - \rho.1.1.0} = \rho. \tag{A10}$$

The slope of the categorization rule can, however, deviate from the low-level correlation, if both dimensions are separable and their weights can be adjusted to the task (e.g. ignore viewpoint).