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 Gerhardtstein, 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