Perception, 2007, volume 36, pages 375 – 390

DOI:10.1068/p5311

Does an estimate of environmental size precede size scaling on a form-comparison task?

David J Bennett
Department of Cognitive and Linguistic Sciences, PO Box 1978, Brown University, Providence, RI 02912, USA; e-mail: David_Bennett@Brown.edu

Received 29 June 2004, in revised form 12 April 2006; published online 21 February 2007

Abstract. In a number of studies reaction time has been found to increase with increases in size ratio on a same – different form-comparison task. Bennett and Warren (2002, *Perception & Psychophysics* **64** 462 – 477) teased apart environmental and retinal size ratios by showing the forms along a (simulated) texture hallway, viewed monocularly, on a simultaneous form-comparison task; roughly equal effects of environmental and retinal size ratios were found. The current study enhanced scene-size information by showing the forms in simulated stereo and by adding texture to the forms themselves; special care was also taken to perceptually isolate the stimuli. In two experiments, with different kinds of forms, strong effects of environmental size ratio were found; no effect of retinal size ratio (same trials) was observed in either experiment. The results support the hypothesis that the (same-trial) form-size codings reflect 'all told' estimates of environmental size, and place constraints on modeling the functional architecture of the visual system.

1 Introduction

In a number of studies reaction time has been found to increase with increases in size ratio on same – different form-comparison tasks; simultaneous presentation (eg Bundeson and Larsen 1975) and successive presentation (eg Bundeson et al 1981) yield qualitatively similar results.[1] However, in these studies the stimuli were shown in isolation against a blank background, and so environmental and retinal size ratios were confounded. It is therefore not possible to say which controlled the rise in reaction time. Bennett and Warren (2002) prized apart environmental and retinal size ratios by presenting random forms along a simulated textured hallway, viewed monocularly, on a simultaneous form-comparison task. With the monocularly viewed textured hallways, the effects of environmental size ratio and retinal size ratio were about equal (environmental size ratios are the size ratios specified ideally by the scene-size information, eg the hallway texture). When the forms were shown against a vertical textured wall, preserving the same screen locations as used in the hallway conditions, retinal size ratio dominated, but there was still a significant effect of environmental size ratio; presumably this resulted from information carried by the heights in the picture plane of the forms (see Sedgwick 1986).

One plausible interpretation of this pattern of results, favored by Bennett and Warren (2002), is that prior to 'mental scaling'[2] and form comparison, subjects

[1] The typical finding is that there is a close to linear dependence of reaction time on size ratio, and this is the case in Bennett and Warren (2002) and in the present study. Cave and Kosslyn (1989) demonstrated that other functions also fit their (size-scaling) data, but they, too, found that reaction time depends close to linearly on size ratio.

[2] On the usual interpretation, 'mental scaling' involves the transformation of image-like representations. Not everyone agrees with this kind of explanation, or with similar interpretations in the 'mental rotation' literature (cf Pylyshyn 2003). The relevant literature is large here, but see Finke (1989), Farah (2002), and Kosslyn (1994) for appeals to empirical evidence in favor of 'mental imagery' interpretations; see also (especially) Tye (1991) for a defense of the conceptual coherence of this kind of account. Though no commitment is made to any specific or detailed account of the nature of imagistic representation, for the purposes of this paper, it is assumed that some kind of 'mental imagery' interpretation of the size-scaling experiments is correct, and questions are addressed, and the data interpreted, as the issues arise and look from within this broad framework.

pooled available environmental size information and made 'all told'[3] estimates of environmental size. On this interpretation of the results, the effect of 'retinal size ratio' reflected incomplete scene-size information (eg there was no stereo information) and/or conflicting information specifying a flat screen (accommodation, some parallax information, and the absence of a blur gradient). As Bennett and Warren (2002, page 467) put matters: "To the degree to which the displays are interpreted as flat, the perceived size ratio of the forms will approach retinal size ratio—so 'flat size ratio' is, on this account, a better phrase than 'retinal size ratio' ".

However, this is not the only possible interpretation of the pattern of results observed in Bennett and Warren (2002). Rock (1983) notes that there is a kind of visual receptiveness in which the visual angles of objects have some reflection in awareness: "although an object at varying distances does appear to be the same objective size, its changing visual angle is by no means without representation in consciousness. We are aware of, even if not attending to, the fact that at a greater distance the object does not fill as much of the visual field of view as it does when it is nearby" (page 254). The mode of perception in which we take note of this presentation of size that (partially) reflects visual angle Rock calls the 'proximal mode', which he contrasts with the 'world mode', in which we naturally and immersively attend to the environmental sizes and size relations of objects. Similarly, Gibson (1950, pages 26 – 43) distinguished an attitude of attending to the visual field, which requires an 'introspective or analytic' disengagement from the more familiar, everyday immersion in the 'visual world'. So one alternative interpretation of the results of Bennett and Warren (2002) is that the effect of 'retinal size ratio' reflected a partial tuning out of the environmental size information, as a result of taking up something like Rock's 'proximal mode' when engaged in the task. It is indeed conceivable that such 'attending to the visual field' may have speeded responses by bypassing at least some of the processing engaged in determining environmental size (although such 'proximal mode' sensing is also unnatural and requires an effortful visual reorientation, and so might, on these grounds, be subject to slow processing).

Is there any way to determine whether the 'all told' estimate of size interpretation is correct, as an interpretation of the Bennett and Warren (2002) results? The approach taken in the current study is to enhance the scene-size information by adding stereo and by adding texture to the forms themselves; special care was also taken to perceptually isolate the stimuli (see sections 2.1 and 3.1). If the 'all told' estimate of size interpretation of subjects' performance is correct, then, under these conditions, environmental size ratio should dominate, and there should be little if any effect of 'retinal' size ratio.

What, theoretically, rides on the truth or falsity of the 'all told' estimate of size interpretation? The answer places constraints on theories of the functional architecture of the visual system. So, for example, Kosslyn et al (1990) maintain that information about the depth and shape of facing surfaces is computed by the combined operation of specialized input modules and deposited in a 'visual buffer'. Kosslyn et al liken the information available in the visual buffer to the information available in Marr's 2.5D sketch (see also Kosslyn 1994, page 356). Many other researchers also posit a level of surface representation resulting from the combined operation of processes extracting depth and shape information (cf Landy et al 1995; see also Nakayama et al 1995).

[3] This just means that the size estimates take account of all available size information, as specified by scene-size information, and as specified by 'extra scene'-size information, when available. No specific view is assumed how this information is combined in arriving at a final size estimate; neither Bennett and Warren (2002) nor the current study was designed to determine this.

On the Kosslyn et al (1990) account, size is held to be coded in retinal coordinates at the level of the visual buffer; this commitment (to retinal or angular coding of size) is supported by citing the widely known neurophysiological studies that are held to reveal a large number of retinotopically organized visual maps (see also Cave et al 1994; for discussion, see the 'General discussion' in Bennett and Warren 2002; see also Murray et al 2006). Elsewhere, Kosslyn maintains that mental imagery is associated with patterns of activity in such a visual buffer (Kosslyn 1980, 1987; Kosslyn and Thompson 2002), in line with evidence that neural machinery in visual cortex is exploited by at least some kinds of mental imagery (Farah 1990, 2002; Finke 1989; Kosslyn 1987; Kosslyn and Thompson 2002). Since mental imagery is held to be supported by activity in a buffer where size is coded by visual angle, one would expect only an effect of retinal size ratio in the form-comparison task (see also Cave et al 1994); in keeping with this, Kosslyn and coworkers hold that environmental size is determined downstream from the visual buffer, in parietal cortex (Kosslyn 1994; Kosslyn et al 1990). The Bennett and Warren (2002) results, with monocular viewing, already conflict with this prediction, and suggest that environmental size is at least partly coded for in any such visual buffer underlying mental imagery. If environmental size ratio dominates under stereo viewing, this would indicate that a full 'all told' estimate of environmental size fixes size codings at the point in visual processing that is associated with the mental imagery employed in doing the size-scaling task—and this would constrain theorizing about the functional architecture of the visual system in a fundamental way.

2 Experiment 1

Figure 1 shows an example of a stimulus used in experiment 1; actual viewing was in simulated stereo (see section 2.1 for details). The two chief differences between these displays and those used in Bennett and Warren (2002) are the texture on the forms and the stereo viewing. The texture on the forms could, in principle, facilitate any measuring out of the form sizes in terms of the texture units of the surrounding surfaces. The texture on the forms also provides elements internal to the forms available for stereo matching. In verging on the forms, the vergence, disparity, and visual-angle information available could, in principle, yield information about the environmental sizes of the forms. However, the main role played by the addition of stereo may well be to help 'break through' the screen, overwhelming the accommodation information, the slight parallax information, and the absence of a blur gradient (all indicating the presence of a flat screen).

In addition to the size information just briefly surveyed, there is also a kind of horizon-scaling information that subjects might utilize (for a theoretical discussion of horizon scaling, see Sedgwick 1980; for empirical studies, see Mark 1987; Warren and Whang 1987; Wraga 1999a, 1999b). Because there is no visible ground surface (unlike in the Bennett and Warren 2002 displays), it may not be possible to reliably perceptually locate the textured floor of the simulated scene relative to the actual ground surface of the laboratory apparatus. And so there may be no way to reliably determine the sizes of the forms, through horizon scaling, in body-scaled units. But it would be in principle possible to scale up the scene in 'floor to horizon' size units, without giving a body-scaled meaning to the 'floor to horizon' measure. This would require determining the location of the implicit horizon, but this could be done by extending the parallels defined by the texture elements of the floor of the simulated enclosure.

As in Bennett and Warren (2002), the basic approach was to vary both the 'retinal' (or 'flat screen') and environmental size ratios in a simultaneous same–different form-comparison task—where, as in the earlier experiments, 'environmental size ratio' refers

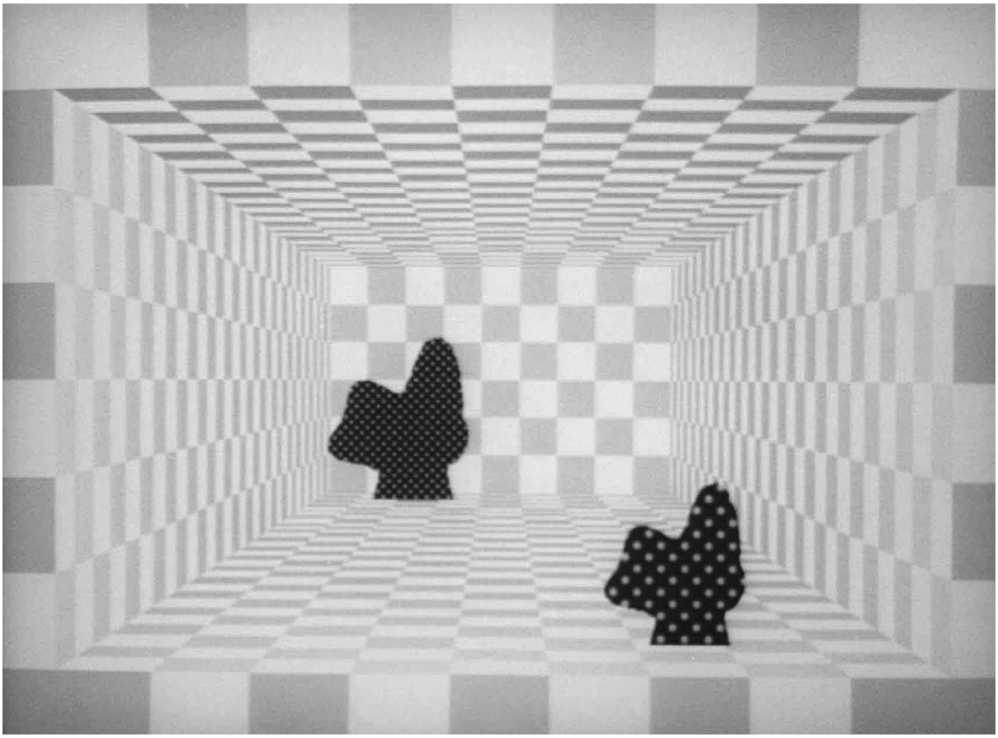

Figure 1. Black-and-white reproduction of an experiment 1 same-trial stimulus. The retinal size ratio is 1 : 1 and the environmental size ratio is 1 : 2.2. Actual viewing was in (simulated) stereo.

to the size ratio specified ideally by the scene-size information. The relative effects of 'retinal' (or 'flat screen') and environmental size ratios on reaction time can be gauged by comparing the main-effect slopes.

2.1 *Method*

2.1.1 *Subjects.* Thirty-nine subjects participated in the experiment, including seven (specifically recruited) subjects—graduate students and research assistants—who were experienced in running in vision-research experiments (although none had prior experience running in experiments by the author). Two of the thirty-two non-specifically recruited subjects were dropped as 2.5 standard deviation overall (same trial) reaction time outliers.

2.1.2 *Apparatus.* The displays were generated on a Silicon Graphics Onyx2 and displayed on a 19 inch monitor with a resolution of 1280 × 1040 pixels (though vertical resolution was halved in stereo mode). The screen edges were masked off with black poster board. Remaining reflections from the screen light were masked off by setting an adjustable riser so that its top was close to and just below eye level (in this configuration the riser itself perceptually disappeared). The arms of the subjects and the computer mouse were hidden from view beneath a table-like construction that supported the riser. The head was stabilized with a chin-rest and a padded head-rest, though some (very) slight side-to-side head movement was still possible. Distance to the screen was 75 cm.

Stereo was simulated with Stereographics liquid-crystal goggles, essentially a fast shuttering system, alternating left-eye and right-eye views. The transmittance was given (by the company) as '32% typical' and the dynamic range was given (by the company) as 1500 : 1 (dynamic range is the ratio of the transmittance of a shutter when open to its transmittance when closed).[4] If the forms are shown without texture against a blank grey background, there is faint but readily detectable ghosting (with contrast approximately equated to the experiment displays, as measured by a photometer). When the forms are shown in the textured box, but without texture on the forms, there is some ghosting visible on the interior of the forms. But in the experiment displays there was texture on the forms and a textured background. And under these conditions there is very little visible ghosting, in keeping with observations by developers of the stereo goggles (Lipton 1987). When the left-eye and right-eye images of a form overlap in screen location, there is some slight ghosting visible internal to the form, corresponding to the overlap (even with the texture on the forms). But this requires close scrutiny to detect.

2.1.3 *Stimuli and design.* The checkerboard surround was grey and white; simulated lighting was used, resulting in a slight darkening with distance; the forms were dark-blue with white texture elements.

New random forms were generated on each trial, according to certain rules (described below). The forms were designed so that the difficulty of the same – different discrimination could be controlled: as a result of pilot studies, the difficulty was set so that the overall reaction times were reasonable (not far from 1 s), and yet subjects engaged in 'mental scaling' at the low size ratios (the pilot studies suggested that, if the same – different discrimination was made too easy, the 'mental scaling' response-time effect would be lost, beginning at the small size ratios—presumably because subjects then do the task entirely by looking for salient feature differences).

Construction of the stimulus forms is illustrated in figure 2. First, eight points were chosen within eight 25° 'pie slices', under the constraint that they could not be within 5.87° of the rays bounding the pie slices; as illustrated in figure 2, the points were also not allowed to be too close[5] to the 'origin' of the pie slices. Then, a curved outline was approximated that passed near these eight points; small, straight line steps were used to approximate the curve (splines were not used, though the outlines appeared smooth).[6] On same trials the forms had the same shape. On different trials one of the forms was constructed by randomly perturbing the initial eight points; new points were constrained to be outside of the smaller, inner circles illustrated in figure 2, but inside the larger circles. The new points were also constrained to lie within the

[4] There are two possible sources of ghosting. First, with some systems, images shown to one eye can be visible, to some degree, to the other, shuttered, eye (this is called 'shutter leakage'). But qualitative observations by the developers of the liquid-crystal goggles suggest that this does not occur with dynamic ranges greater than about 40 : 1 (Lipton 1987, 1991); no appreciable reduction in ghosting was observed by systematically increasing dynamic range past this point. Given the vagaries of informal, qualitative judgments, and difficulties in precisely measuring dynamic range (Lipton 1987), this is an approximate figure. But it is much lower than the dynamic range reported by the company or the goggles used in the current experiments. This leaves the second source as the plausible source of the ghosting that remained (see the text): phosphor persistence.

[5] The lines, in figure 2, crossing the base of each 'pie slice' determine an area, near the origin, that the points could not fall in; the percentage of the distance from the origin to the far edges of the pie slices at which this line crossed ranged from 34% at the sides to 26.4% for the two top pie slices.

[6] A partial account of how the curve was approximated is as follows. Consider the rays from the 'origin' to each of the eight points. Consider two such adjacent rays. To begin, another point was chosen that was linked to the origin by a ray midway in length between the lengths of the rays associated with the two bounding points, at an angle halfway between the two rays associated with the bounding points. Two more points were then chosen in a similar way, between the ray to the 'halfway point' and the original points on either side. If the points were connected at this point, there still would be sharp angles—points and valleys—associated with the original eight points; a somewhat ad hoc, several-step procedure (not described) was then used to 'round off' these points and valleys, by means of two short, straight-line segments.

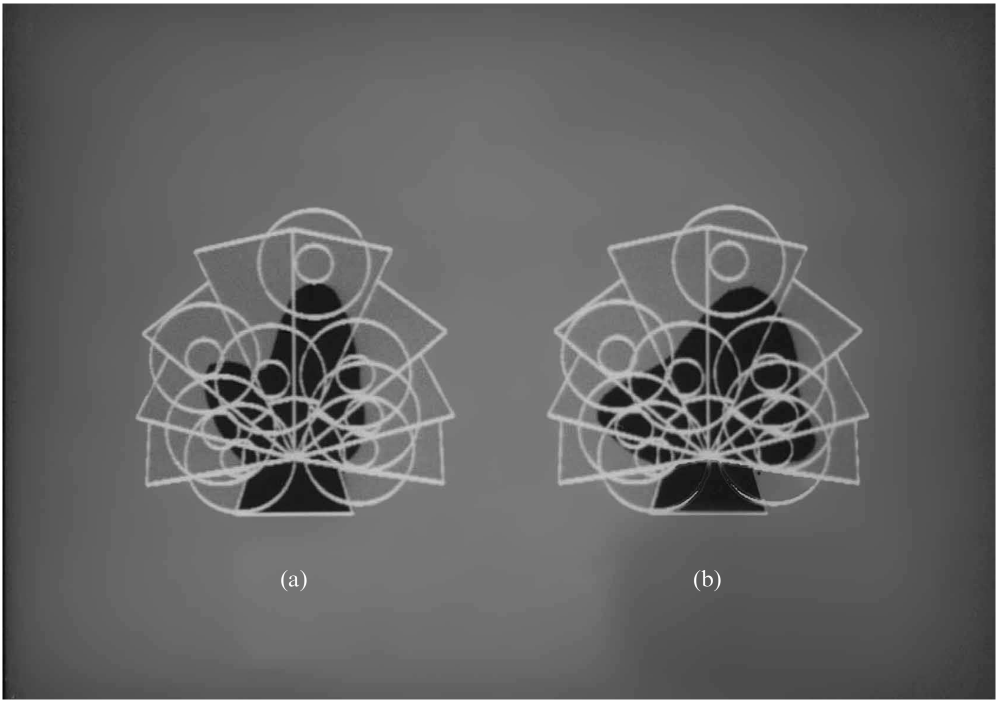

Figure 2. Figure construction: (a) original form; (b) new form arrived at by randomly deforming the original form (see the text for a fuller description).

same pie slice (and, as in the choice of the initial eight points, at least 5.87° from the bounding rays of each pie slice, and not too close to the 'origin').

To simulate stereo viewing, asymmetric viewing frustums were defined for each eye, with the dimensions adjusted depending on eye width. Eye width was measured for each subject by placing a clear plastic ruler on the bridge of the nose.

A two-factor crossed design was used, with four different retinal and environmental ratios, 1 : 1, 1 : 1.4, 1 : 1.8, 1 : 2.2 (standard : variable). There were thus 16 different combinations of retinal and environmental ratios. There were 12 tokens of each of the 16 combinations of size ratios for same trials, and 12 tokens for each of the 16 combinations for different trials, resulting in a total of 384 trials. Same and different trials were randomly intermixed.

The simulated height of the standard form was always approximately 0.15 m in environmental coordinates (as specified, ideally, by the scene-size information); the vertical visual angle of any given form ranged from about 1.5 deg to about 7.5 deg. The simulated distance to the near texture elements was 1.83 m and the simulated distance to the back wall was 4.84 m. The simulated distance to the mask, the fixation bar, and the average midpoint of the two forms (when separated in depth) was 3.176 m. The midpoint (in depth separation) of the forms was randomly jittered from trial to trial to minimize any effect on reaction times of any differences or idiosyncrasies in the way the surrounding texture met or aligned with the forms; when the two forms were shown in the same depth plane, the distance to the depth plane was randomly jittered.

2.1.4 *Procedure.* At the beginning of each experimental session subjects were led into a dark room, guided by a dim flashlight. The screen was uncovered only after subjects were settled in the chin- and head-rests and were wearing the stereo goggles. At this point, an empty, textured surround was shown (like the display shown in figure 1, but minus the two forms).

The sessions took approximately 40 min. There were 110 practice trials, with feedback about correctness, followed by the 384 experimental trials, in six blocks of 64 trials each, without feedback. Subjects were instructed to ignore differences of size in judging whether the two forms were the same or different, and that "we will be analyzing how fast you respond on those trials where you answered correctly, so answer as quickly as possible while still remaining accurate". "Same" responses were entered by pressing the right mouse button, and "different" responses were entered by pressing the left mouse button (with the two index fingers).

Each trial began with a black fixation bar (1.85 deg in height) shown sitting vertically in the middle of the hallway for 1 s. Then a blank grey field was shown for 200 ms (to prevent apparent motion between the fixation bar and the forms), followed by the forms sitting in the hallway. A pattern mask consisting of a grid of squares of randomly varying lightness (with 3.6 deg sides) was displayed as soon as a response was recorded. Subjects clicked the middle mouse button to begin a new trial.

2.2 *Results*

The average overall error rate was 4.2%; incorrect responses were not included in the analysis. Outliers were eliminated by the following rule: average reaction time and standard deviation for the same trials (for each subject) were computed and outliers over 3 standard deviations (from the overall same-trial average for that subject) were dropped; the same procedure was then followed for the different trials. This led to 1.75% of the trials being dropped.

The results are shown in figures 3 and 4. All of the analysis-of-variance (ANOVA) *p* values reflect a Greenhouse–Geisser correction for within-subjects designs (Greenhouse and Geisser 1959; the first method [of two] they present, as implemented in SPSS). A two-way repeated-measures ANOVA on same trials revealed a strong effect of

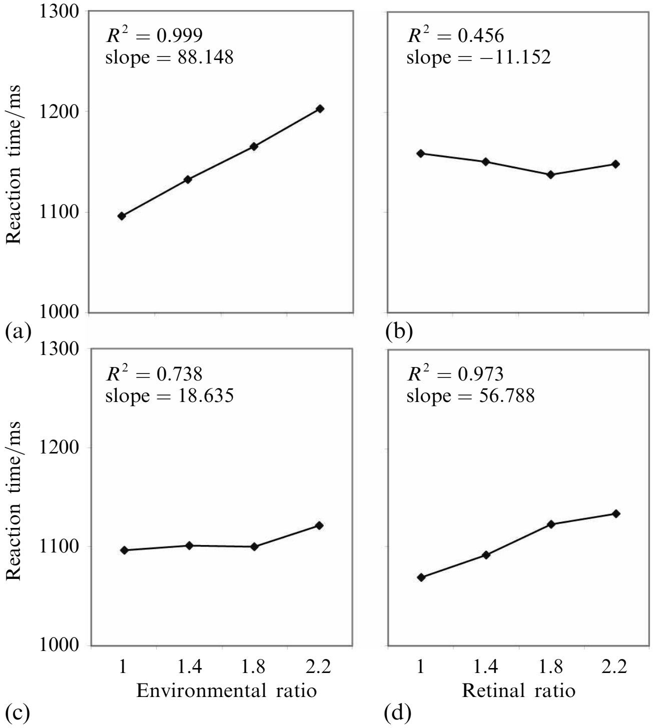

Figure 3. Experiment 1: (a) environmental size ratio, same trials; (b) retinal size ratio, same trials; (c) environmental size ratio, different trials; (d) retinal size ratio, different trials.

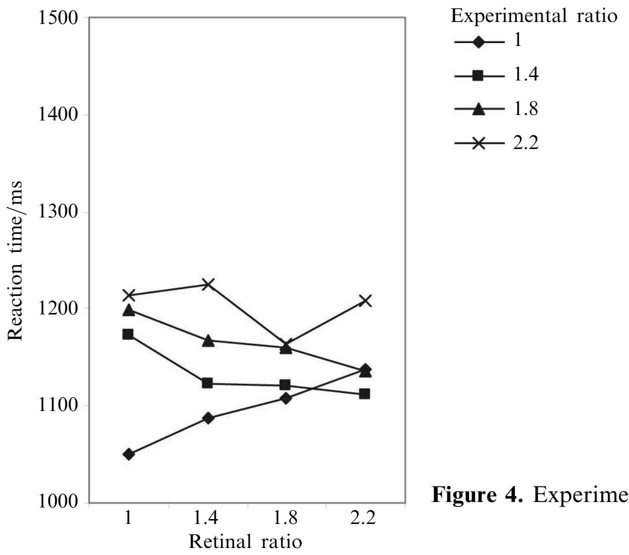

Figure 4. Experiment 1; full trial types, same trials.

environmental size ratio ($F_{3,108} = 19.393$, $p < 0.001$), but no effect of retinal size ratio ($F_{3,108} < 1$). There was also a significant environmental × retinal interaction ($F_{9,324} = 2.926$, $p = 0.01$); as discussed below, this may reflect an influence on reaction time of separation in depth. As indicated in table 1, there is no evidence that a speed versus accuracy trade-off accounted for the same trial pattern of reaction times. The slight numerical rise in error rate with retinal size ratios was not significant ($F_{3,108} = 1.96$, $p > 0.1$). The results for the seven specially recruited experienced subjects were qualitatively similar to the results for the other subjects, and so their data were pooled together with those of the other subjects. For the seven specially recruited subjects, the environmental slope was 101.81, and the retinal slope was −6.45.

Table 1. Same-trials error rates.

Experiment	Environmental ratio			Retinal ratio				
	1 : 1	1 : 1.2	1 : 1.8	1 : 2.2	1 : 1	1 : 1.2	1 : 1.8	1 : 2.2
1	1	1.32	1.7	1.78	1.22	1.68	1.24	1.68
2	1.37	1.37	1.34	1.69	1.49	1.37	1.34	1.57

The pattern on different trials was very different, indicating the operation of a different kind of process than the process underlying (at least primarily) the same-trial results (see the discussion below). There was no effect of environmental size ratio ($F_{3,108} < 1$), but there was a significant effect of retinal size ratio ($F_{3,108} = 8.315$, $p < 0.001$), though the slope was substantially less than the same-trial environmental-size-ratio slope. The different-trial environmental × retinal interaction was not significant ($F_{9,324} = 2.001$, $p = 0.063$).

2.3 *Discussion*

2.3.1 *General conclusions.* On the same trials there was a strong effect of environmental size ratio, but no effect of retinal size ratio. The different-trial pattern was very different, and, as discussed below, this somewhat complicates the overall interpretation of the results. But the same-trial reaction-time pattern suggests the use of representations that reflect 'all told' estimates of environmental size.

The completely flat same-trial 'retinal' (or 'flat screen') slope is perhaps surprising. If one takes up Rock's 'proximal mode' in viewing the stimuli, it is clear that retinal or angular size had a substantial effect on the relative amounts of visual field that the forms appear to take up. This is to be expected, in the same way that the diminishing angular size of cars parked at greater distances has some reflection in awareness. One might have expected at least some reflection of retinal or angular size ratio (per se) in the underlying size codings, and so in the reaction times. Alternatively, even with the care taken in perceptually isolating the stimuli, one might have expected that the slight, residual information specifying the screen of the computer—a solid object, at 75 cm distance—to have led at least some subjects to partially flatten out the displays. This might perhaps especially have been expected with the experienced subjects, who had much experience with computer-generated images; it might have been thought that this would lead them to be less 'taken in' by the stimulus information specifying environmental size. Thus, accommodation, the absence of a blur gradient, and the (very) slight residual parallax information all specify a flat screen (for discussion, see Watt et al 2005). But there is no evidence in the data that retinal or angular size ratio per se, or the (slight) residual flat-screen information, had any effect on the underlying size representations, even with the experienced subjects. Watt et al (2005) obtained broadly complementary results. They had subjects estimate the slant of simulated planes with and without the computer screen rotated to the same slant as the depicted plane. Under monocular viewing there was an effect of screen slant, but under stereo viewing there was no effect of screen slant. So it seems that with at least some stimuli and tasks, and with care taken to perceptually isolate the stimuli, stereo information can override residual information specifying a flat screen.

There are, however, also reasons why the results are not so surprising. A sensitivity to environmental size makes sense functionally, at least for some tasks: we need to know environmental size if we are to grasp objects, pass through openings unimpeded, or determine where and how to sit or step. And, given the results of experiment 1, it is plausible that the representations tapped play some role in guiding motor behavior in interacting with objects in the immediate environment (see also the related research discussed in section 4). There was also very little information remaining specifying a flat screen. Accommodation information, for example, may well have been simply overwhelmed, and dropped from consideration as widely discordant (see Landy et al 1995). And, with heads quite well stabilized, there was little parallax information available to subjects.

2.3.2 *Separation in depth.* An effect of separation in depth would lead to an interaction of a specific sort. Consider the rightmost column of four data points in figure 4, detailing the full-trial-type results. The upper right point corresponds to trials where the retinal and environmental ratios were both 1 : 2.2, and so the two forms are at the same simulated distance. As environmental ratio gets smaller (keeping the retinal ratio at 1 : 2.2) the separation in depth of the forms increases. Thus, if there was an effect of separation in depth, the right sides of the bottom lines should curve up; similar reasoning associates an effect of separation in depth with a curving up of the upper left sides of the lines. And there is some suggestion of such a pattern, reflected in a significant interaction. Although the evidence is not that clear or completely consistent, this aspect of the results hints at a mandatory 'translation in depth' (see also the corresponding results for experiment 2 below, in figure 7, where the pattern is perhaps a little clearer). There was no evidence of an influence on reaction time of separation in depth in the full-trial-type graphs of Bennett and Warren (2002), presumably because of the weaker depth information in the monocular displays used. The Bennett and Warren (2002) displays also included a salient visible horizon, with the forms sitting on a ground plane, and this may have encouraged subjects to estimate size directly, through horizon-ratio scaling, at no point coding for form distances. Pringle and Uhlarik (1982) looked at speeded judgments of sameness or difference of size—a task they suggest may engage a size-scaling transformation—and found an effect of separation in depth.

There is a weak negative correlation of -0.33 between separation in depth and the main-effect size ratios, retinal and environmental.[7] So it is possible that an effect of separation in depth slightly suppressed the effects of environmental and retinal size ratios, and, specifically, masked a small underlying contribution to size coding of retinal size or size ratio. Especially since the correlation is weak, it is doubtful that any contribution of separation in depth to reaction time had much of a suppressing effect, if it had any effect at all. However, strictly, because of this small negative correlation, all that can be definitely concluded is that scene environmental size information at least largely controlled underlying size codings, with at best little contribution of retinal size and/or of the slight residual information specifying a flat screen—and quite likely no contribution from either of these latter two sources. The conclusion above that there is 'no evidence' of an effect of retinal or angular size per se, or of the slight residual flat-screen information, still seems warranted.

2.3.3 *Different trials.* Different-trial performance was strikingly different from same-trial performance, with a positive retinal-size-ratio slope and no evidence of an environmental-size-ratio slope (though the retinal-size-ratio slope was much shallower than the same-trial environmental-size-ratio slope). On same – different tasks, where there is a same-trial reaction-time effect on a dimension irrelevant to the task, performance on different trials (when reported) is stimulus-dependent in ways that are not well understood (Farrell 1985, especially pages 429 – 430 and 434 – 435); and interpretations of the data offered in the present section will be rather speculative. The fact that the retinal-size-ratio slope was considerably flatter than the same-trial environmental-size-ratio slope suggests that subjects did not always need to 'mentally scale' in order to determine that the forms differed in shape; the comparison might have instead proceeded by detecting a salient difference, working off non-imagistic features or part-based representations. However, the positive retinal slope on the different trials suggests that, on at least some of the trials, the "different" response did result from imagistic, mental scaling, though with size coded in terms of retinal or angular size. One possibility is that different trial mental scaling works off imagistic representations of form pieces or parts, and it may be that, in order for an environmental size coding to be achieved, forms must be considered as wholes, situated in a wider, size-specifying environment.

The flat different-trial environmental slopes do, however, raise questions about how the more piecemeal or part-based process that, plausibly, underlies "different" responses, ie interacts with the more global or whole-form scaling that, plausibly, underlies "same" responses. The expectation would be that "different" responses would at least occasionally result from the more global or whole-form scaling—otherwise why would the visual system need to undertake the apparently slower, full(er)-form scaling process?

[7] As indicated in section 2.1, when the two forms were not in the same depth plane, the midpoint of the forms—in depth separation—was randomly jittered. The correlation between separation in depth and the size ratios was determined by assuming the average midpoint.

It was incorrectly claimed in Bennett and Warren (2002) that there was no correlation between separation in depth and the retinal and environmental ratios (this was the present author's responsibility). However, the error here has no bearing on the interpretation of the Bennett and Warren (2002) results, for two reasons: first, in Bennett and Warren (2002) there was no evidence in the full-trial-type graphs of the distinctive kind of interaction that would be associated with an effect on reaction time of separation in depth; second, in the fuller information displays (in the earlier study) the effects of environmental and retinal size ratio were about equal, so any small effect of separation in depth would have no bearing on inferences about the relative strength of the effects of environmental and retinal size ratios.

One possibility is that the same-trial mental scaling, that engages more global or holistic form representations, is at least primarily a secondary process, serving to check or verify that forms do share a common shape, at least with the stimuli used in the current study (as reported below, the different-trial pattern was similar for the different kind of stimuli used in experiment 2). The pattern here is reminiscent of an aspect of some mental-rotation results. Handedness information—sufficient to do the task—appears to be available, in some form, prior to mental rotation, which is nonetheless still performed (for this interpretation, see Corballis 1988, page 119; for the data that it is based on, see Corballis et al 1976, page 529; Cooper and Shepard 1982, page 84).

3 Experiment 2

Experiment 2 was similar to experiment 1, except that forms with spiky outlines were used, instead of the smooth-edged forms used in experiment 1 (see figure 5). If global, imagistic representations are employed in doing the task (at least on the same-trials) there is no clear reason to expect a qualitative difference in results with the change in the kind of form outline used. Still, given that the striking, flat 'retinal'-size-ratio (same-trial) slope is, for a couple of reasons, somewhat surprising (see above), it seems wise to test for the robustness and generality of the results in an experiment by using stimuli with a different kind of form outline. This was the rationale for experiment 2.

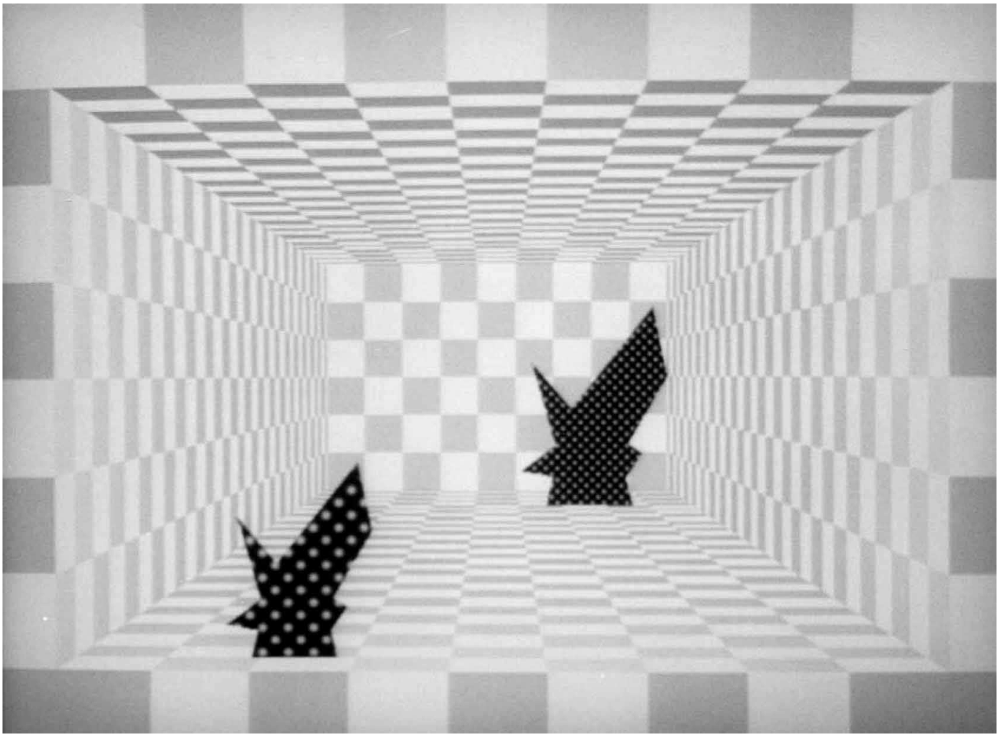

Figure 5. Black-and-white reproduction of an experiment 2 same-trial stimulus. The retinal size ratio is 1 : 1 and the environmental size ratio is 1 : 2.2. Actual viewing was in (simulated) stereo.

3.1 *Method*

3.1.1 *Subjects.* Thirty-five subjects completed the experiment, including three (specially recruited) subjects who were experienced at running in vision-research experiments (although none had prior experience running in experiments by the author).

3.1.2 *Stimuli and design.* The stimuli and the design were the same as in experiment 1, except that forms with spiky outlines were used, instead of the forms with smoothly varying outlines used in experiment 1. The spiky forms were constructed by connecting the original eight points chosen, one per 'pie slice', by straight lines (instead of approximating them by an—approximated—smooth outline). On different trials, one of the forms was constructed by randomly perturbing the vertices of the other form, in the same general way as described in section 2.1 and illustrated in figure 2 (only now the eight perturbed points were directly joined by straight lines).

3.1.3 *Procedure.* The procedure was the same as in experiment 1.

3.2 *Results*

The average overall error rate was 3.4%; incorrect responses were not included in the analysis. Outliers were also eliminated, by the same procedure as in experiment 1; this resulted in 2% of the trials being eliminated.

The results are shown in figures 6 and 7. As in experiment 1, there was a strong effect of environmental size ratio on the same trials ($F_{3,102} = 23.202$, $p < 0.001$), but no effect of retinal size ratio ($F_{3,102} < 1$). There was also a significant environmental × retinal interaction ($F_{9,306} = 3.806$, $p = 0.002$); as in experiment 1, this may reflect an influence of separation in depth. As indicated in table 1, there is no evidence that a speed versus accuracy trade-off accounted for the same-trial reaction-time patterns. The results for the three specifically recruited experienced subjects were qualitatively similar to the results for the other subjects, and so their data were pooled together with the data for the thirty-two non-specially recruited subjects (for the three specially recruited experienced subjects, environmental slope was 72.5, and retinal slope was 11.17).

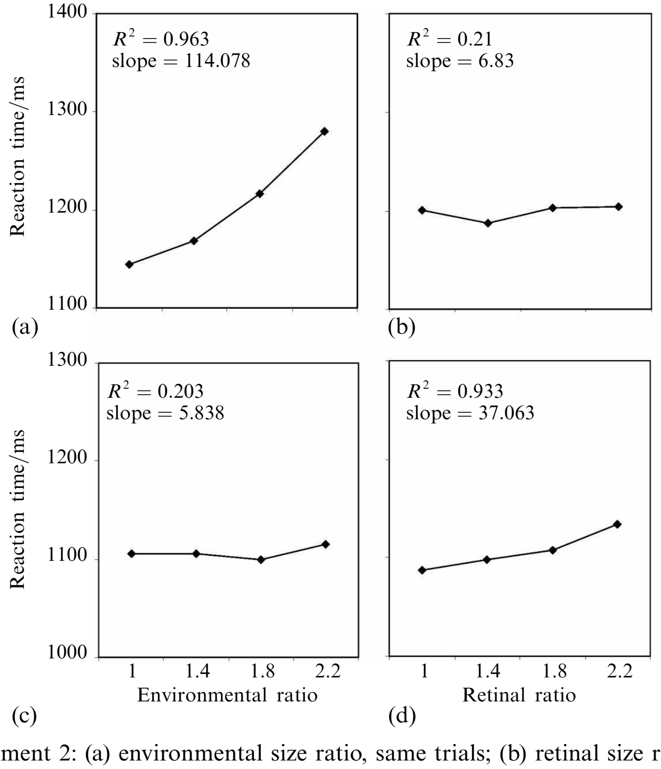

Figure 6. Experiment 2: (a) environmental size ratio, same trials; (b) retinal size ratio, same trials; (c) environmental size ratio, different trials; (d) retinal size ratio, different trials.

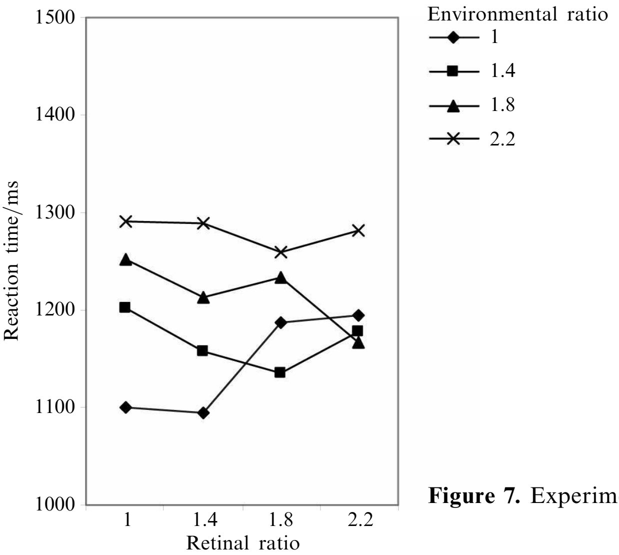

Figure 7. Experiment 2; full trial types, same trials.

Once again, the pattern of results for the different trials was very different than the same-trial pattern of results. There was no effect of environmental size ratio ($F_{3,102} < 1$), but there was a significant effect of retinal size ratio ($F_{3,102} = 4.552$, $p = 0.006$), though the slope was much shallower than the same-trial environmental-size-ratio slope. The different-trial environmental × retinal interaction was not significant ($F_{9,306} = 1.31$, $p > 0.1$).

3.3 *Discussion*

The results were qualitatively similar to those obtained in experiment 1. On the same trials there was a large effect of environmental size ratio, and no effect of retinal size ratio—though, as in the experiment 1, the different-trial pattern was very different. As in experiment 1, this pattern of results accords with the hypothesis that the representations used in arriving at the same-trial responses embody 'all told' estimates of the environmental sizes of the forms.

Once again, there was, as well, some suggestion of an effect of separation in depth, reflected in a significant interaction (see figure 7 and the earlier discussion). Because of the presence of the weak negative correlation of separation in depth with retinal and environmental size ratio, interpretation of the results must be slightly tempered, for the reasons discussed earlier. But, once again, the evidence suggests that scene environmental information played, at the least, the dominant role in fixing underlying size codings, for the representations implicated on the same-trial responses—with at best little contribution from retinal or angular size ratio per se and/or from the residual information specifying a flat screen, and probably no contribution from either of these other possible sources.

4 General discussion

In both experiment 1 and experiment 2 there were strong (same-trial) effects of environmental size ratio, but no effects of retinal size ratio. This suggests that the size codings embodied in the representations underlying mental scaling, on the same trials, were at least largely, and probably entirely, controlled by scene-size information; neither residual information specifying a flat screen, nor retinal size ratio per se, seemed to have any effect on (same-trial) response times (as indicated by the flat retinal-size-ratio slopes). The results support the conclusion favored by Bennett and Warren (2002) that prior to (same-trial) mental scaling 'all told' estimates of environmental size were determined, pooling available environmental-size information. If the representations used are deployed in a 'visual buffer' of the sort posited by Kosslyn and coworkers, then an 'all told' estimate of environmental size has been made by this point in visual processing—contrary to the commitments of Kosslyn and coworkers (see also Murray et al 2006).

Results that appear to be particularly closely related to the results obtained in the current study have been reported in two previous studies.

Spatial frequency discrimination has been widely used in studying spatial vision. Burbeck (1987) pointed out that, since the sine-wave gratings are invariably shown at the same distance, retinal frequency (cycles per degree) and object frequency (say, cycles per centimeter) are confounded. In Burbeck (1987) these were teased apart by presenting the gratings on two different monitors, located at different distances. Perhaps most telling, subjects appeared unable to learn to do the task by comparing retinal frequencies. All told, retinal frequencies did not appear to play any role in subject's responses.

Biederman and Cooper (1992) and Jolicoeur (1987) found that reaction times are lower in an old–new memory task when the sizes of the test and study forms agree. Milliken and Jolicoeur (1992) prized apart retinal and environmental sizes by presenting stimuli on computer screens that were moved about. Their results suggest that it is entirely an estimate of environmental size that is represented in memory.

Further, two studies (Aks and Enns 1996; Ramachandran 1989) have shown an effect of information specifying environmental size on the deployment of attention in visual search—suggesting at least some quite early, pre-attentive sensitivity to environmental size. Bennett and Warren (2002) also review a number of additional studies, not focused on the detection of size, indicating that the visual system is, in general, geared to detect surface and scene properties, with contributions likely beginning quite early in visual processing.

Finally, Wexler et al (1998) and Wohlschläger and Wohlschläger (1998) found evidence that the representations underlying 'mental rotation' are implicated in the close-in guidance of action. Although these studies did not concern 'mental scaling', they did explore the role of a similar kind of high-level spatial processing. And, as discussed above in considering the results of experiment 1, it makes functional sense—for the purposes of guiding action—for the visual system to code at least some representations of shaped surfaces or objects in terms of environmental size. (However, size may not be represented at all in those representations that, at least directly, underlie *classification*—see Biederman and Cooper 1992; Cooper et al 1992; see also Milner and Goodale 1996.)

In sum, the results of experiments 1 and 2, taken together with these other, complementary studies, strongly indicate that the visual system is geared to detecting and coding for environmental size, in the exercise of the visual capacities that underlie a variety of different tasks.

5 Conclusions

Though the very different pattern of results on the different trials somewhat complicates interpretation, the overall results of experiments 1 and 2 suggest that the coding of size in the representations that underlie the same-trial responses reflect 'all told' estimates of environmental size. With the stimuli used, the (same-trial) size codings were apparently entirely controlled by the scene-size information, with retinal or angular size per se, and/or the (slight) residual flat-screen information, playing at best little and probably no role in fixing the underlying size codings. These results place constraints on theorizing about the functional architecture of the visual system.

Acknowledgments. Thanks to Bill Warren for laboratory support and advice, to Philip Fink for help with data collection and analysis on experiment 2, and to Zili Lui for advice on an aspect of the data analysis. Part of this work was supported by a postdoctoral research grant to the author by the Institute of Research in Cognitive Science, University of Pennsylvania.

References

Aks D J, Enns J, 1996 "Visual search for size is influenced by a background texture gradient" *Journal of Experimental Psychology: Human Perception and Performance* **22** 1467 – 1481

Bennett D J, Warren W, 2002 "Size scaling: Retinal or environmental frame of reference?" *Perception & Psychophysics* **64** 462 – 477

Biederman I, Cooper E, 1992 "Size invariance in visual object priming" *Journal of Experimental Psychology: Human Perception and Performance* **18** 121 – 133

Bundeson C, Larsen A, 1975 "Visual transformations of size" *Journal of Experimental Psychology: Human Perception and Performance* **1** 214 – 220

Bundeson C, Larsen A, Farrel J E, 1981 "Mental transformations of size and orientation", in *Attention and Performance IX* Eds J Long, A Baddeley (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 279 – 294

Burbeck C, 1987 "Locus of spatial frequency discrimination" *Journal of the Optical Society of America A* **4** 1807 – 1813

Cave K, Kosslyn S M, 1989 "Varieties of size specific visual selection" *Journal of Experimental Psychology: General* **118** 148 – 164

Cave K R, Pinker S, Giorgi L, Thomas C E, Heller L M, Wolfe J, Lin H, 1994 "The representation of location in visual images" *Cognitive Psychology* **26** 1 – 32

Cooper L A, Shepard R N, 1982 *Mental Images and their Transformations* (Cambridge, MA: MIT Press)

Cooper L A, Schachter D L, Ballesteros S, Moore C, 1992 "Priming and recognition of transformed three-dimensional objects: Effects of size and reflection" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18** 43 – 57

Corballis M C, 1988 "Recognition of disoriented shapes" *Psychological Review* **95** 115 – 123

Corballis M C, Zbrodoff N J, Roldan C E, 1976 "What's up in mental rotation?" *Perception & Psychophysics* **19** 525 – 530

Farah M, 1990 "The neural basis of mental imagery" *Trends in Neurosciences* **12** 461 – 470

Farah M, 2002 "The neural bases of mental imagery", in *The New Cognitive Neurosciences* 2nd edition, Ed. M S Gazzaniga (Cambridge, MA: MIT Press) pp 965 – 974

Farrell B, 1985 "'Same – different' judgments: A review of current controversies in perceptual comparisons" *Psychological Bulletin* **98** 419 – 456

Finke R A, 1989 *Principles of Mental Imagery* (Cambridge, MA: MIT Press)

Gibson J J, 1950 *The Perception of the Visual World* (Boston, MA: Houghton Mifflin)

Greenhouse S W, Geisser S, 1959 "On methods in the analysis of profile data" *Psychometrika* **24** 95 – 112

Jolicoeur P, 1987 "A size congruency effect in memory for visual shape" *Memory & Cognition* **15** 531 – 543

Kosslyn S M, 1980 *Image and Mind* (Cambridge, MA: Harvard University Press)

Kosslyn S M, 1987 "Seeing and imaging in the cerebral hemispheres: A computational approach" *Psychological Review* **94** 148 – 175

Kosslyn S M, 1994 *Image and Brain: The Resolution of the Imagery Debate* (Cambridge, MA: MIT Press)

Kosslyn S M, Flynn R A, Amsterdam J B, Wang G, 1990 "Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes" *Cognition* **34** 203 – 277

Kosslyn S M, Thompson W L, 2002 "Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience", in *The New Cognitive Neurosciences* 2nd edition, Ed. M S Gazzaniga (Cambridge, MA: MIT Press) pp 975 – 985

Landy M S, Maloney L T, Johnston E B, Young M, 1995 "Measurement and modeling of depth cue combination: In defense of weak fusion" *Vision Research* **35** 389 – 412

Lipton L, 1987 "Factors affecting 'ghosting' in time-multiplexed plano-stereoscopic CRT display systems", in *Proceedings of SPIE* **761** *True Three-Dimensional Imaging Techniques and Display Technologies* Eds D F McAllister, W E Wobbins, pp 75 – 78

Lipton L, 1991 "Selection devices for field-sequential stereoscopic displays: a brief history", in *Proceedings of SPIE* **1457** *Stereoscopic Displays and Applications II* Eds J O Merritt, S S Fisher, pp 274 – 282

Mark L S, 1987 "Eyeheight-scaled information about affordances: A study of sitting and chair climbing" *Journal of Experimental Psychology: Human Perception and Performance* **13** 361 – 370

Milliken B, Jolicoeur P, 1992 "Size effects in visual recognition memory are determined by perceived size" *Memory & Cognition* **20** 83 – 95

Milner A D, Goodale M A, 1996 *The Visual Brain in Action* (Oxford: Oxford University Press)

Murray S, Boyaci H, Kersten D, 2006 "The representation of perceived angular size in human primary visual cortex" *Nature Neuroscience* **9** 429 – 434

Nakayama K, He Z J, Shimojo S, 1995 "Visual surface representation: A critical link between lower-level and higher-level vision", in *An Invitation to Cognitive Science* volume 2, 2nd edition, Eds D N Osherson, S M Kosslyn (Cambridge, MA: MIT Press) pp 1 – 70

Pringle R, Uhlarik J, 1982 "Comparative judgments of distal size: A chronometric analysis" *Perception & Psychophysics* **32** 178 – 186

Pylyshyn Z, 2003 *Seeing and Visualizing: It's Not What You Think* (Cambridge, MA: MIT Press)

Ramachandran V S, 1989 "Is perceived size computed before or after visual search?" paper presented at the meeting of the Psychonomic Society, Atlanta

Rock I, 1983 *The Logic of Perception* (Cambridge, MA: MIT Press)

Sedgwick H A, 1980 "The geometry of spatial layout in pictorial representation", in *The Perception of Pictures* volume 1 *Alberti's Window* Ed. M A Hagen (New York: Academic Press) pp 33 – 90

Sedgwick H A, 1986 "Space perception", in *Handbook of Perception and Human Performance* volume 1 *Sensory Processes and Perception* Eds K L Boff, L Kaufman, J P Thomas (New York: Wiley) pp 21.1 – 21.57

Tye M, 1991 *The Imagery Debate* (Cambridge, MA: MIT Press)

Warren W H, Whang S, 1987 "Visual guidance of walking through apertures: Body-scaled information for affordances" *Journal of Experimental Psychology: Human Perception and Performance* **13** 371 – 383

Watt S J, Akeley K, Ernst M O, Banks M S, 2005 "Focus cues affect perceived depth" *Journal of Vision* **5** 834 – 862

Wexler M, Kosslyn S M, Berthoz A, 1998 "Motor processes in mental rotation" *Cognition* **68** 77 – 94

Wohlschläger A, Wohlschläger A, 1998 "Mental and manual rotation" *Journal of Experimental Psychology: Human Perception and Performance* **24** 397 – 412

Wraga M, 1999a "The role of eye height in perceiving affordances and object dimensions" *Perception & Psychophysics* **61** 490 – 507

Wraga M, 1999b "Using eye height in different postures to scale the heights of objects" *Journal of Experimental Psychology: Human Perception and Performance* **25** 518 – 530

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 36 2007

www.perceptionweb.com