Stereo and motion parallax cues in human 3D vision: Can they vanish without a trace?

Andreas M. Rauschecker
Medical Scientist Training Program, Neurosciences Program, Stanford University School of Medicine, Stanford, CA, USA

Samuel G. Solomon
Disciplines of Physiology and Anatomy & Histology, School of Medical Sciences, University of Sydney, Sydney, Australia

Andrew Glennerster
School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom, & Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom

In an immersive virtual reality environment, subjects fail to notice when a scene expands or contracts around them, despite correct and consistent information from binocular stereopsis and motion parallax, resulting in gross failures of size constancy (A. Glennerster, L. Tcheang, S. J. Gilson, A. W. Fitzgibbon, & A. J. Parker, 2006). We determined whether the integration of stereopsis/motion parallax cues with texture-based cues could be modified through feedback. Subjects compared the size of two objects, each visible when the room was of a different size. As the subject walked, the room expanded or contracted, although subjects failed to notice any change. Subjects were given feedback about the accuracy of their size judgments, where the "correct" size setting was defined either by texture-based cues or (in a separate experiment) by stereo/motion parallax cues. Because of feedback, observers were able to adjust responses such that fewer errors were made. For texture-based feedback, the pattern of responses was consistent with observers weighting texture cues more heavily. However, for stereo/motion parallax feedback, performance in many conditions became worse such that, paradoxically, biases moved away from the point reinforced by the feedback. This can be explained by assuming that subjects remap the relationship between stereo/motion parallax cues and perceived size or that they develop strategies to change their criterion for a size match on different trials. In either case, subjects appear not to have direct access to stereo/motion parallax cues.

Keywords: parallax, stereopsis, size constancy, virtual reality, feedback

Introduction

There is a growing consensus that when several sensory cues contribute to a percept such as the 3D shape of an object, the combination process is well described by a weighted linear summation of cues in which the weighting of each cue is determined by its reliability (Backus, Fleet, Parker, & Heeger, 2001; Buckley & Frisby, 1993; Jacobs, 2002; Johnston, Cumming, & Landy, 1994; Johnston, Cumming, & Parker, 1993; Landy, Maloney, Johnston, & Young, 1995; Richards, 1985; Taylor, 1962; Young, Landy, & Maloney, 1993; for recent quantitative analyses, see Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003). It has been argued that this combination may be "mandatory" for cues within one sensory modality because subjects appear to be unable to access the information from individual visual cues, at least for discriminations close to threshold (Hillis, Ernst, Banks, & Landy, 2002). Nevertheless, it is possible to change the weight applied to different cues in a matter of seconds (Triesch, Ballard, & Jacobs, 2002) by changing the reliability of those cues. It has also been shown that training can influence the relative weighting applied to visual and haptic cues (Atkins, Fiser, & Jacobs, 2001). Varying the task has been shown to alter subjects' responses even when the reliability of available cues remains the same. This is probably because the visual system computes quite different parameters depending on the task, rather than the effect being due to reweighting of cues (Bradshaw, Parton, & Glennerster, 2000; Glennerster, Rogers, & Bradshaw, 1996; Tittle, Todd, Perotti, & Norman, 1995).

Here, we investigate an "expanding room" environment that was presented using virtual reality in which, at first sight, subjects seem to ignore stereoscopic cues and motion parallax information altogether (Glennerster et al., 2006). The purpose was to determine whether, using feedback, we could train subjects to attend to the stereo/

doi: 10.1167/6.12.12 Received August 3, 2006; published December 19, 2006 ISSN 1534-7362 © ARVO

motion parallax cues and weight these more heavily in determining their responses. Unlike other paradigms (Hillis et al., 2002), the stereo/motion parallax signals were not near threshold, increasing the chance that subjects could learn to base their responses on those cues in isolation.

The subjective reports of people in the expanding room are that they are surrounded by a stable room, although its dimensions (as specified by stereopsis and motion parallax) change greatly in all directions (up to fourfold in this experiment) as they walk across the room. The impression of a stable scene is equally strong when the floor and ceiling are removed, showing that it is not simply because people assume a constant eye height (Ooi, Wu, & He, 2001). The relative disparities and equivalent motion parallax signals change by a factor of 4 (i.e., a 300% increase, substantially above a detection threshold of 10–20% for disparity increments; McKee, Levi, & Bowne, 1990); hence, they should be readily detected. Indeed, in our apparatus, subjects can detect the change in size of the room when they walk through a virtual wall from a small room into a large room, using only stereo/motion parallax cues. The situation in which stereo/motion parallax cues are apparently suppressed is when the room expands around them as they walk across it (the center of expansion is the cyclopean point) so that as objects get farther away, they also get larger. In this case, an assumption that objects and texture elements (such as the bricks that compose the walls of the room) remain the same physical size (and the same distance) conflicts with the stereo/motion parallax cues. This assumption (which leads to what we describe as a "texture-based cue") appears to dominate subjects' perception of the size of the room.

Despite observers' subjective reports on the apparent size of the expanding room, there is good evidence that stereo/motion parallax cues contribute to subjects' performance when they are asked to carry out certain tasks. For example, when they are asked to compare the sizes of two objects, one seen when the room is small, the other seen when the room is large, observers' matches are well described by a weighted combination of information from texture-based and stereo/motion parallax cues (Glennerster et al., 2006). The question we address in this article is whether subjects can be trained to bias their responses in this task toward the size signaled by stereo/motion parallax cues if they are given appropriate feedback. Figure 1 illustrates three different ways in which this feedback could operate: by changing the weight applied to each cue, by changing the interpretation of each cue (the mapping from a cue value to a size), or by changing the interpretation of the combined cues. These possibilities are discussed in detail in the Model section. We find that subjects do change their responses because of feedback but in ways that imply that the visual system has no direct access to stereo/motion parallax cues and cannot increase the weight applied to them.

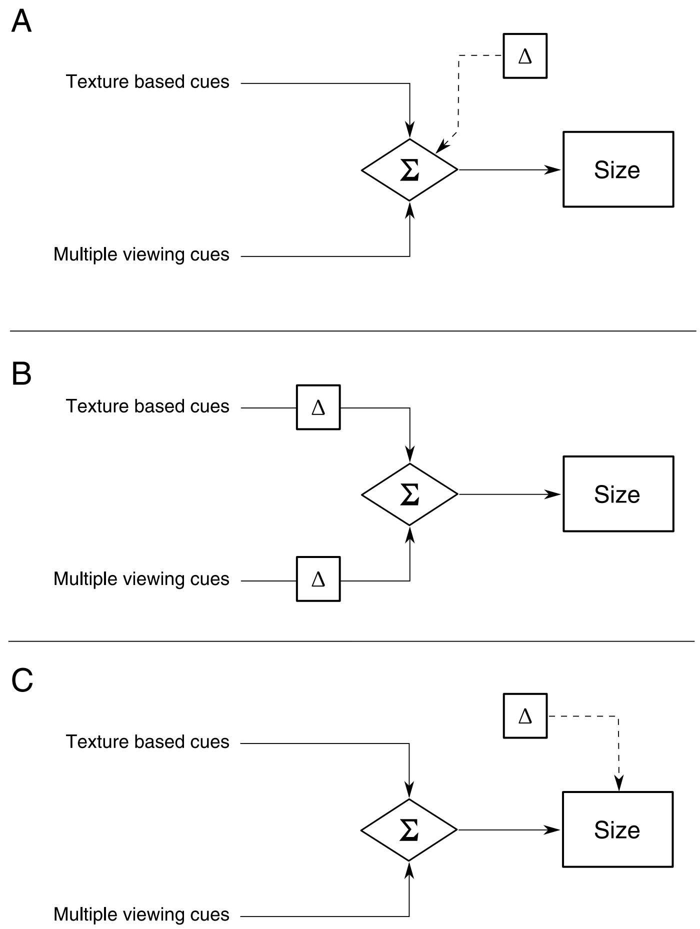

Figure 1. Illustration of the three models described in the Model section. Feedback could (A) modify (represented by Δ) the relative weight applied to texture versus stereo/motion parallax cues ("reweighting" model), (B) cause a change in the size estimates provided by individual cues ("remapping" model), or (C) encourage subjects to shift their criterion for a size match on different trials ("strategy" model).

Methods

Subjects

Five male observers (23–31 years old) had normal or corrected-to-normal visual acuity. Two subjects were authors (S.G.S. and A.M.R.), and three were naïve to the purpose of the experiment (T.J.P., H.G.E., and J.H.P.).

Equipment

The virtual reality equipment used is described in detail elsewhere (Glennerster et al., 2006). Briefly, the virtual reality system consists of a head-mounted display, a head tracker, and a computer, which generate appropriate binocular images given the locations and pose of the observer's head. The Datavisor 80 (nVision Industries Inc., Gaithersburg, MD) head-mounted display unit presents separate 1,280 × 512 pixel images to each eye using CRT displays. In our experiments, each eye's image was 72° horizontally by 60° vertically with a binocular overlap of 32°, giving a total horizontal field of view of 112° (horizontal pixel size, 3.4 arcmin).

The location and pose of the head was tracked using an IS900 system (InterSense Inc., Burlington, MA). This system combines inertial signals from an accelerometer with a position estimate obtained from the time of flight of ultrasound signals. Four ultrasound receivers are attached to the tracker ("Minitrax"); more than 50 ultrasound emitters placed around the room send out a timed 40-kHz pulse sequence. The data are combined by the InterSense software to provide 6 *df* in the estimate of the tracker pose and location. These data are polled at 60 Hz by the image generation program. Because the offset of the tracker from the optic centers of each eye is known, the position and pose of the head tracker allow the 3D location of the optic centers to be computed. These are used to compute appropriate images for each eye. Binocular images were rendered using a Silicon Graphics Onyx 3200 at 60 Hz. We have measured the latency from movement of the Minitrax tracker to image change as 48–50 ms.

Stimulus and task

Subjects moved in a virtual room whose dimensions depended on the location of the subject in the real room. When the subject was on the left side of the room and standing within an unmarked viewing zone (0.5 × 0.5 m), a red "reference" cube was visible ahead of them, presented either 0.75 or 1.5 m away from the center of the viewing zone. (For the smallest room, the 1.5-m cube was 12.5 cm from the far wall.) Subjects were instructed to walk to their right until a comparison cube appeared (within a similar unmarked viewing zone close to the right wall, see Figure 2) and to signal, by pressing one of two buttons, which cube appeared larger. The comparison cube was also at 0.75 or 1.5 m. Leaving the first viewing zone caused the reference cube to disappear; thus, no simultaneous comparison of the two cubes was possible.

When either cube was visible, the room remained stable. However, in the region between the two viewing zones, the room size was directly related to the lateral component of the subject's location. On some trials, the room expanded, whereas on others, it contracted: The expansion factors were 0.25, 0.5, 1, 2, and 4. For example, the inset in Figure 2 shows how the room size changed with subject location when the expansion factor was 4. The point of expansion was the cyclopean point (halfway between the eyes) so that as objects got larger, they also got farther away. Thus, no single image would allow the observers to know whether they were in a large or in a small room. Only comparison of two or more views (and a knowledge of the separation of the optic centers from which the images were obtained) would reveal the size of the room. The dimensions of the nonexpanded virtual room were 3 m wide × 3.5 m deep. At this scale, the virtual floor was at the same level as the subject's feet. The smallest size of the room was 1.5 × 1.75 m and the largest was 6 × 7 m. The walls were textured with a brick pattern and the floor with regular tiles (see Figure 2). No other objects were present in the room.

Psychometric procedure

Measurements of 20 independent psychometric functions were randomly interleaved in one run of trials (two distances of the reference cube, two distances of the comparison, and five room expansion factors). Each psychometric function consisted of 40 trials, that is, a total of 800 trials in a run. Subjects were encouraged to take breaks about every 100–150 trials. (In all, 800 trials took approximately 3–4 hr, wherein 4.8 km walking is required, which could be spread across more than 1 day.) Other properties of the cubes were also randomized but did not define separate psychometric functions. These were the heights of the reference and comparison cubes and the size of the reference cube. If g is the room expansion factor on a given trial (0.25, 0.5, 1, 2, and 4), then the reference cube size was 0.75, 1, or 1.5 times a "standard" cube size (constant with respect to the virtual room) of $1/\sqrt{g} \times 10$ cm. The heights of the reference cubes were $1/\sqrt{g} \times 9$, 18, or 36 cm below eye level and the heights of the comparison cubes were $\sqrt{g} \times 9$, 18, or 36 cm below eye level.

On each trial, the size of the comparison object relative to the reference was determined according to a standard staircase procedure (similar to Cornsweet, 1962; Johnston et al., 1993; Levitt, 1971). The data for each psychometric function were gathered using four randomly interleaved staircases, two starting from a high value and two from a low one. The two staircases starting from a high value were a 1-down, 3-up staircase (i.e., one correct answer and the cube size would be made smaller, three errors and it would be made larger) and a 3-down, 1-up staircase (i.e., converging more slowly and steadily). The two staircases starting from a low value were the same but in reverse (1-up, 3-down and 3-up, 1-down). The step sizes reduced over the first six trials per staircase (step size was $6/N$, where N is the trial number on that staircase until $N \geq 6$, after which step size remained constant). The starting ranges were such that they included both a real size match (as specified by stereo/motion parallax) and a texture- or room-based match (equal size relative to the room). Because the experimental cue was a size ratio, the scale used in both the staircase and psychometric fitting was logarithmic. The size of the comparison object at the

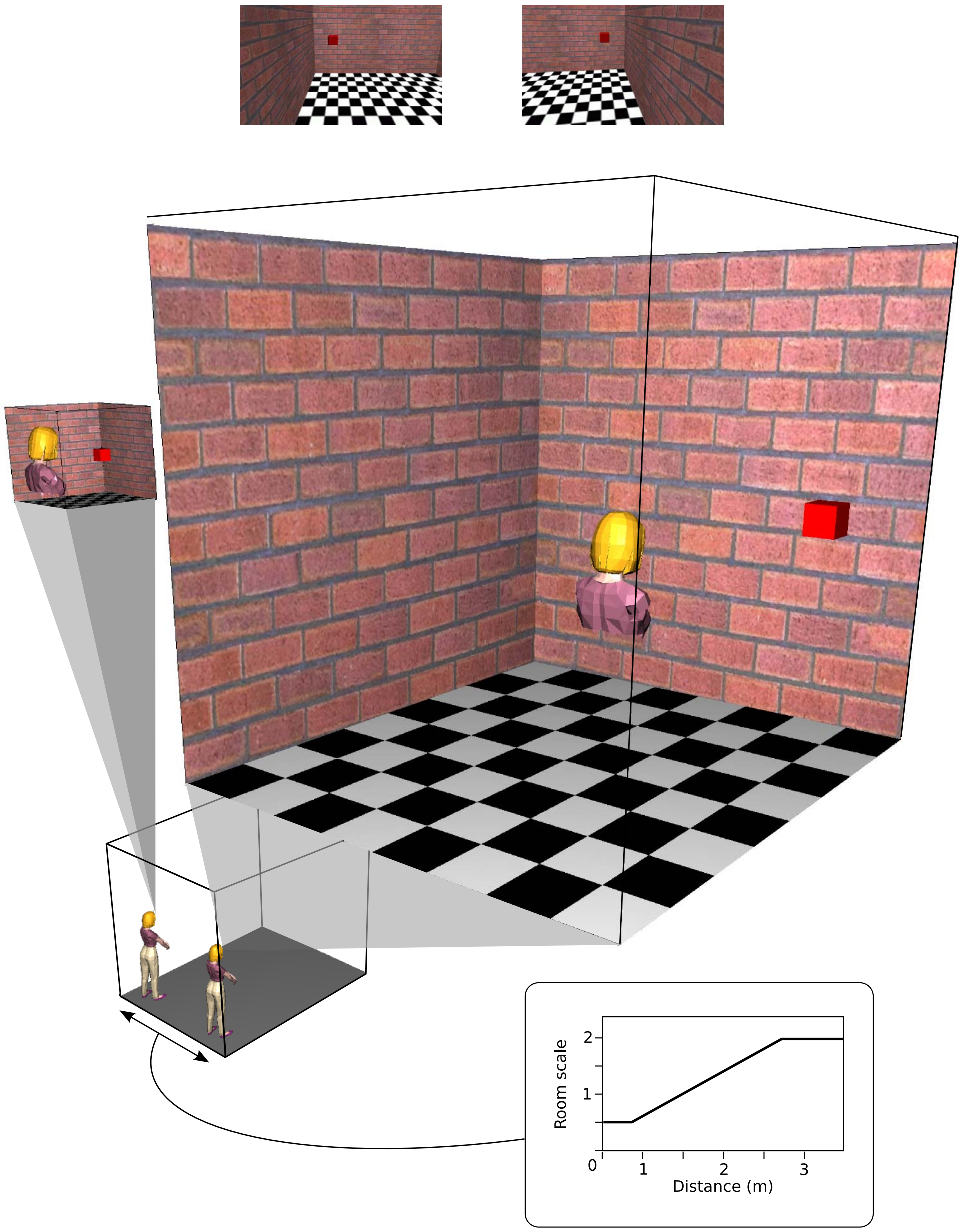

Figure 2. Illustration of the relationship between the virtual and physical rooms. As the observer moves from side to side within the physical room, the size of the virtual room changes. Here, the room expansion factor is 4, which means that the size of the virtual room gradually expands by a factor of 4 as the observer moves from left to right. The center of expansion is the cyclopean point; thus, any single view cannot reveal the changed size of the room (as can be seen from the images shown above). Subjects had to judge whether a cube that was visible when they were on the right side of the room (“comparison”) was larger or smaller than a cube that was visible when they were on the left side (“reference”). On other trials, the room could remain static or decrease in size. (Figure reproduced, with permission, from Glennerster et al., 2006, © Elsevier.)

point of subjective equality was obtained by fitting each psychometric function with a cumulative Gaussian by probit (Finney, 1971). In Figures 3 and 4, size matches show the bias (50% correct point). Error bars show the standard error of the mean. In some runs, feedback was given: This was the primary experimental manipulation. Details of the feedback given to subjects are described with each experiment. In Figures 3 and 4, the plots for feedback conditions show the results obtained for the last 400 trials within each run to allow the effect of feedback to show after at least 400 trials.

Information about distance

In our display, information about the distance of objects is available to the subject from multiple views: binocular views at any moment and changing views over time. This

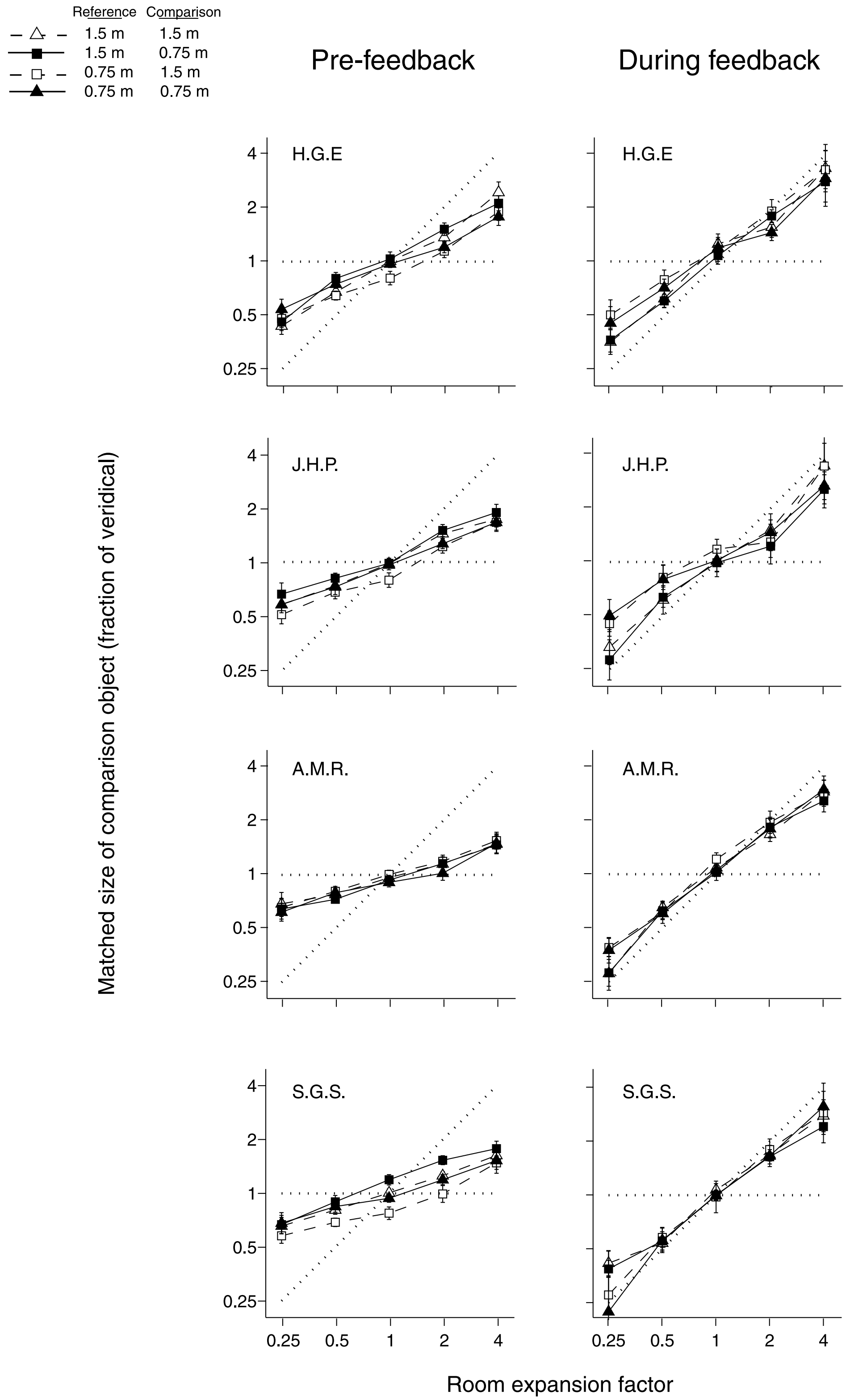

Figure 3. Results of Experiment 1 (texture-based feedback). The ratio of the physical size of the comparison object to the physical size of the reference object at the point of subjective equality (i.e., the subject's size match) is plotted against the expansion factor of the virtual room (on a log scale) for all four observers. Data are plotted separately for different distances of reference and comparison objects, as described in the key. The horizontal and diagonal dotted lines show the expected size matches using only stereo/motion parallax or texture-based cues, respectively. The left and right columns show matches made before and during feedback. The unusual pattern of pre-feedback data for subject S.G.S. arises from the fact that he carried out Experiment 2 before Experiment 1 (see text). Error bars show ±1 *SEM*.

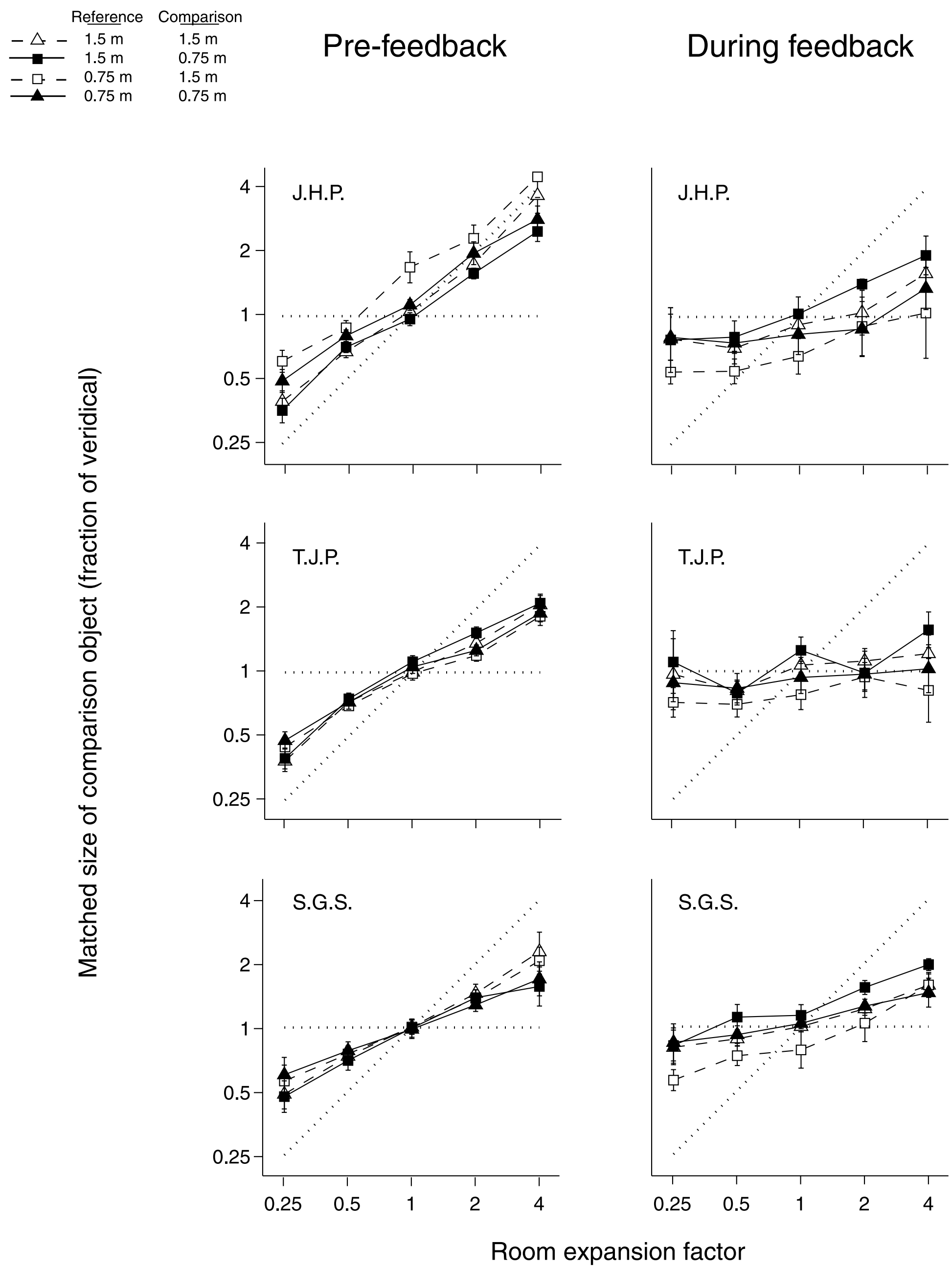

Figure 4. Results of Experiment 2 (stereo/motion parallax feedback). Data are plotted in the same way as in Figure 3.

is the information we describe as "veridical" in the rest of the article because it defines the scale and 3D structure of the scene when combined with information about the "baseline," that is, the distance between the optic center of the eye in two images. For stereo vision, this is the interocular separation; for motion parallax, this is the distance the eye has translated between two views observed at different times. Vergence, vertical disparities, and structure-from-motion are all correct and consistent for these stimuli and contribute to the veridical cues. What we call "texture-based" cues in this article are a set of cues that are consistent with the interpretation that the room remains a constant size. The assumption that bricks, tiles, or other objects remain a constant size contributes to this category of cue, as does the assumption that the height of the eye above the ground plane remains a constant value.

Results

Size matches without feedback

The left-hand column of Figures 3 and 4 shows the size matches that subjects made before they were given feedback in each experiment. For those subjects who had not received feedback before (i.e., this was their first experiment: H.G.E. in Experiment 1; T.J.P., S.G.S., and J.H.P. in Experiment 2), the data generally lie between the horizontal dashed line (a physical match as specified by stereo/motion parallax cues) and the diagonal (a match assuming that the size of texture elements and the room was constant). The lines show fits according to a weighted linear sum of the two cues (see the Model section). Glennerster et al. (2006) also used this simple model, where the weight applied to the stereo/motion parallax cue was smaller for more distant comparison cubes. Although a much narrower range of distances was used here, these data show the same trend (data for comparison distances of 1.5 m are closer to the diagonal, see the Model section). All subjects reported that the room appeared to be of constant size in all conditions, before and after feedback.

Experiment 1: Texture-based cue feedback

The goal of Experiments 1 and 2 was to determine whether subjects could change their responses in the size-matching task according to a criterion set by the computer.

In the first experiment, the "correct" size match for any given trial was a comparison object that was the same size as the reference object in relation to room size. Texture-based cues such as brick size signal the correct estimate of stimulus size in this experiment because the size of texture elements is scaled directly with room size. (Thus, if the reference cube is two bricks high, then the correct match of the comparison cube is also one that is two bricks high, regardless of what size the cubes appeared to be.) The goal of feedback in this experiment was to drive observers to a size match based on these texture cues. The experiment was slightly different for the two naïve observers (H.G.E. and J.H.P.) as compared with the two experienced observers (authors A.M.R. and S.G.S.). The naïve observers did not know what rule was being used by the computer to determine feedback, whereas the experienced observers did.

Figure 3 shows the results (right-hand column). If observers changed their responses according to the feedback, their data should lie along the diagonal (matched size scaled directly with the expansion factor of the room). As can be seen in Figure 3, observers' responses moved closer to a texture-based match during texture-based feedback, although never quite obtaining it. The subjects' success at minimizing their errors can be measured by the deviation from the texture-based match, where deviation is calculated as the sum of squared logarithmic distances between the observers' size matches and the matches expected from texture cues alone. Across all subjects, the average deviation decreased by a factor of 3.9 from pre-feedback to feedback blocks. By this criterion, observers A.M.R. and S.G.S., who knew the purpose of the experiment, were more successful (with an average decrease in deviation of 7.9 times) than naïve observers H.G.E. and J.H.P. (average, 2.6 times). The slopes of the data increased, whereas matches made in the static room (room expansion factor = 1) remained unaffected. This is the pattern that was expected if subjects weight the texture cue more heavily in determining their responses.

Experiment 2: Veridical (stereo/motion parallax) feedback

In this second experiment, the correct size match was one in which the comparison object has the same physical size as the reference object regardless of the room expansion factor. Stereo/motion parallax cues signal the correct estimate of stimulus size in this experiment. In Figure 4, this would be reflected as data lying along the horizontal dotted line (i.e., matched size is always equal to 1). As in Experiment 1, the naïve observers (J.H.P. and T.J.P.) were not told the rule by which the computer determined its feedback, whereas observer S.G.S. knew. However, in this case, knowledge of the feedback rule appeared not to help.

Figure 4 shows the data during the feedback run (right-hand column). For all subjects, feedback about the veridical size of the cube (as specified by stereo and motion parallax) caused size matches to become less dependent on the room expansion factor (matches moved toward the horizontal, as expected). The improvement can again be quantified by the difference between observers' size match and that expected for stereo/motion parallax cues alone. The average deviation was a factor of 4.1 smaller than it was prior to feedback. One might assume that this reduction in errors implies that the subjects had access to the stereo/motion parallax information in the stimulus and changed their responses accordingly. However, a distinctive pattern of the data, which is present for all three subjects, suggests that this is not the case. When the reference and comparison objects were presented at different distances, matches spread away from the horizontal line that defines physical size matches. When the reference object was at 1.5 m and the comparison object was at 0.75 m, all observers produced a larger matched size (i.e., comparison objects were judged as smaller than they should have been) than when the reference was at 0.75 m and the comparison was at 1.5 m, where all observers produced a smaller matched size than they should have. This led to increased error rates from pre-feedback to feedback blocks for certain conditions, most clearly when the room expansion factor was 1 (a static room) with reference and comparison objects at different distances: In this condition, the deviation was, on average, 1.63 times larger during feedback than in pre-feedback. In other words, size constancy had been lost in a static room because subjects no longer matched similar-sized objects presented at different distances, despite being given appropriate feedback that should help them to do so.

(The failure of size constancy is in the opposite direction to that predicted by retinal size matching.) This trend continued in the third block of trials (postfeedback; not shown), where the average deviation was 2.04 times larger than that in the pre-feedback block.

The unexpected pattern of results with veridical feedback suggests that observers could not simply ignore texture-based cues. It seems that whatever mechanism or strategy that allows the subjects to reduce their errors during feedback causes, as a by-product, increased errors in particular conditions. In the following section, we will explore two types of models that could explain this pattern of responses. Briefly, one model assumes that rather than reweighting cues, in which case the combined output would be constrained to lie somewhere between the values signaled by the two cues, the visual system can instead "remap" one or both cues to signal a quite different size of the comparison (or reference) cube. The other model assumes that subjects use a conscious strategy to change their criterion on different trials. Because they cannot identify the trials with different expansion factors, which would allow them to respond to the feedback appropriately, they instead use other parameters, such as the perceived distance of the reference and comparison cubes. These signals correlate to a limited extent with the room expansion factor and, hence, allow them to reduce their errors overall but at the cost of introducing systematic biases.

Model

Figures 3 and 4 show that feedback, whether it is in relation to texture-based or stereo/motion parallax cues, causes observers to adjust their responses and reduce the errors they make. What perceptual mechanisms underlie this? In this section, we consider a simplified model of cue combination in 3D visual perception, illustrated in Figure 1 and based largely on that of Landy et al. (1995). We assume that the size-matching task requires three stages of analysis: first, compatible ("promoted") estimates of size are made independently for texture-based and stereo/motion parallax cues; second, these estimates are combined linearly to form a unified sensory representation; third, this sensory representation is filtered in relation to the observer's task to form the report. Feedback could exert its influence at any or all of these stages.

Size matches before feedback

Our task required observers to match the size of two objects, with information available from texture-based and stereo/motion parallax cues. The model postulates that for each object, the observer perceives size, S, as a summation of independent texture-based (T) and stereo/motion parallax (P) estimates. Thus, perceived size,

$$\begin{aligned} S &= w_T T + w_P P \\ w_T + w_P &= 1, \end{aligned} \tag{1}$$

where w_T and w_P are the weights given to texture-based and stereo/motion parallax cues at one distance from the observer, respectively. The estimate of size from stereo/motion parallax cues, P, is always veridical, and the estimate of size from texture-based cues, T, is always inversely proportional to the size of the room (all texture elements scale with the room). The weights of texture-based and stereo/motion parallax cues are allowed to vary with distance reflecting the decline in precision of stereo/motion parallax cues at larger distances. We assume that the weights of texture-based and stereo/motion parallax cues at a given distance are independent of the size of the room.

Equation 1 can give us predictions for the perceived size (S) of the reference cube on the left side of the room (S_L) and for the comparison cube on the right side of the room (S_R). In our task, the size match, M, is the inverse of the ratio of perceived size of the two objects[1]:

$$M = [S_R/S_L]^{-1}. \tag{2}$$

The left-hand panels of Figures 3 and 4 show that for all observers, the size match depended on the size of the room (the size match is approximately proportional to the room expansion factor) and, to a lesser extent, on the distance of objects from the observer (for most expansion factors, the size match depends on the relative distance to the two objects). The model in Equations 1 and 2 was able to accommodate the pattern of size matches in each case (e.g., lines in Figure 5, left-hand plot, and pre-feedback fit parameters given in Table 1), with two parameters allowed to vary, namely, the relative weights of texture-based and stereo/motion parallax cues at the two object distances tested. For every observer, size matches could be best accounted for by assuming that stereo/motion parallax cues are weighted more at closer distances: Average w_P was 0.57 (SD = 0.12, n = 5) at a distance of 0.75 m and was 0.44 (SD = 0.19) at 1.5 m ($p < .05$, paired Student's t test). The relationship between viewing distance and weighting given to texture-based or stereo/motion parallax cues in this task has already been discussed by Glennerster et al. (2006).

Incorporating the influence of feedback

We consider three ways in which feedback could exert its influence, as illustrated in Figure 1. First, there might be a reweighting of mechanisms in the process of

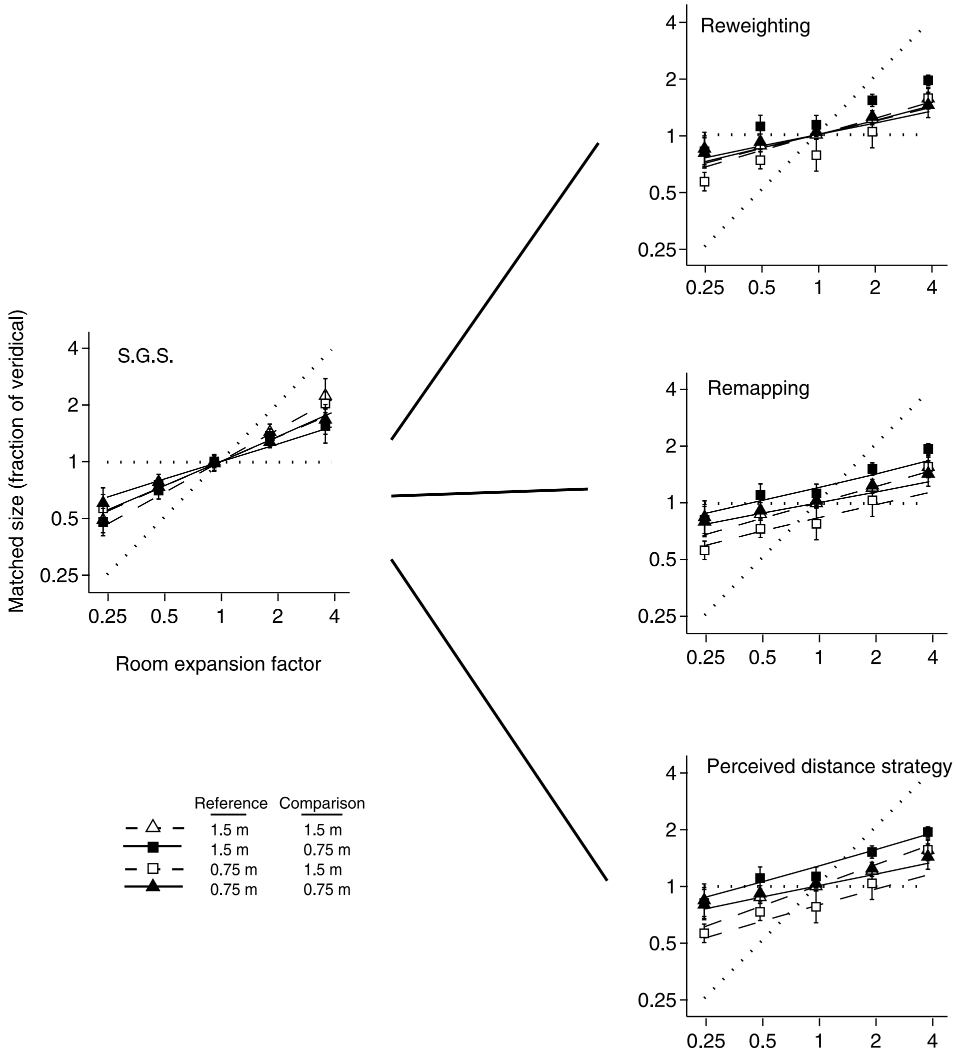

Figure 5. The three models illustrated for one data set. Data for one observer is shown before (left) and after (right) stereo/motion parallax feedback (Experiment 2). The pre-feedback data are fitted with a weighted linear summation model (see the Model section). Of the three models, the remapping model fits the data most closely (see Figure 6 and the Model section for details).

combining the two size estimates (Figure 1A), that is, allowing the relative weights of cues to vary in the same way as described above for the model applied to the pre-feedback data. Second, feedback could cause a remapping between the input and size estimate for either texture or stereo/motion parallax cues or both (Figure 1B). Finally, the observers may adopt a strategy to minimize errors by varying their criterion for an equal-sized match from trial to trial (Figure 1C). In the following, we will discuss each of these models and concentrate on the responses under veridical feedback, which proved the best at differentiating the model predictions (Figure 5).

For each model, best fitting predictions were obtained by minimizing the mean square distance between the logarithms of the size match, M, and the model's predictions. In the case of the remapping and strategy models, cue weights at each of the two object distances were estimated from the pre-feedback size matches. The most obvious way to reduce errors under feedback is to increase the use of the target cue by reweighting its input to the size estimate. We describe this model first.

Cue reweighting

If observers could learn to ignore the irrelevant cues, they would eliminate any errors. In the case of texture-based feedback, observers' responses should then lie along the diagonal, and most observers tended toward this (Figure 3). The reweighting model is able to account for this well, by reducing the weight of the stereo/motion parallax cues contributing to the size estimate (w_P) from 0.66 to 0.29 at 0.75 m and from 0.59 to 0.21 at 1.5 m (n = 4, see Table 1). In Figure 3, this results in a rotation about the origin toward the diagonal. The best fit for this and all subsequent models was found by minimizing the squared error between the model predictions and the observer's responses.

For veridical feedback (Experiment 2), the reweighting model was less successful. The top right panel in Figure 5

	Pre-feedback (all models)		During feedback: Reweight		During feedback: Remap		During feedback: Strategy
	w_P (0.75 m)	w_P (1.50 m)	w_P (0.75 m)	w_P (1.50 m)	P (0.75 m)	P (1.50 m)	k
Veridical feedback							
S.G.S.	0.63	0.46	0.77	0.69	1.94	2.95	0.33
T.J.P.	0.52	0.41	0.97	0.85	5.61	8.46	0.55
J.H.P.	0.39	0.19	0.82	0.68	4.87	12.28	0.54

	Pre-feedback (all models)		During feedback: Reweight		During feedback: Remap		During feedback: Strategy
	w_P (0.75 m)	w_P (1.50 m)	w_P (0.75 m)	w_P (1.50 m)	T (0.75 m)	T (1.50 m)	k
Texture feedback							
S.G.S.	0.71	0.65	0.1	0.31	12.21	7.26	−0.45
A.M.R.	0.7	0.72	0.26	0.18	8.26	9.56	−0.49
H.G.E.	0.59	0.4	0.36	0.2	1.82	4.06	−0.27
J.H.P.	0.62	0.58	0.44	0.17	2.62	4.8	−0.39

Table 1. Parameters of the models used to fit the data in Figures 3, 4, 5, and 6. Columns 2 and 3 show for each observer the estimates of the weight of veridical cues (w_P, Equation 1) in size judgments before feedback, at viewing distances of 0.75 and 1.5 m. The other columns show parameter values for each of the three models fit to the responses during feedback: For the reweighting model, these are new values of w_P at both distances; for the remapping model, they are the size estimates from texture-based cues (T) or those from veridical cues (P) at both distances. The strategy model is defined in Equation 3. In this case, one parameter, k, applies to the data for comparison cubes at both distances.

shows for one observer the size matches made during veridical feedback (symbols) and the corresponding predictions of the model (lines). The model is unable to account for the vertical spreading of responses. This can be best understood for the static room, where any amount of reweighting will still generate the same size match, because $T = P = 1$ in Equation 1 and w_T and w_P always sum to 1.

Cue remapping

Adams, Banks, and van Ee (2001) found that a remapping of binocular disparity cues provided a simple explanation for perceptual adaptation to slant, and we asked if it can also explain the impact of feedback in the size-matching task (Figure 1B). We define remapping as a change in each size estimate (T or P in Equation 1), but in the model, it is only the T/P ratio that matters. Arbitrarily, we have assumed in the following that the remapping applies only to the cue for which feedback was given. We allow size estimates to vary independently at the two object distances. The key difference between this and the previous model is that for the reweighting model, changing the weights applied to one cue must be accompanied by reciprocal changes in the weights to the other cue. This is not the case for remapping: Changing the size estimate for one cue source has no impact on that for the other cue.

As with the reweighting model, remapping can predict the rotation of the data toward the diagonal under texture-based feedback (data in Figure 3, right-hand column); it accomplishes this by increasing the size estimate T at both object distances—by an average of 6.4 ($SD = 2.3$, $n = 8$, see Table 1). There was no difference between the increases at the two object distances ($p = .29$; paired t test on the ratio between the changes at the two distances). Consequently both S_L and S_R tend toward the product $w_T T$. Perceived size is thus dominated by the texture-based component, and the size match tends toward the room expansion factor, g.

The remapping model is also capable of explaining observers' responses under veridical feedback. Its predictions are illustrated in the middle-right panel of Figure 5. As in the case of texture feedback, the model can account for any rotation of the data about the origin by increasing the stereo/motion parallax size estimate, P, at both distances. It can also account for vertical spread by increasing P more at the farther distance (a separate mapping is allowed at each distance). On average, the stereo/motion parallax size estimate, P, at a distance of 0.75 m was increased by a factor of 4.1 ($SD = 1.9$, see Table 1), and at 1.5 m, it was increased by 7.9 ($SD = 4.7$, $p < .05$).[2]

Strategic scaling of size matches

Error minimization during feedback may reflect neither reweighting nor remapping at early stages of visual processing. In the case of texture-based feedback, the experienced observers knew the rule for feedback and could apply a strategy accordingly (Figure 1C). They could, for example, relate object size to the size of the nearest wall bricks, ignoring the perceived sizes S_L and S_R when making their match. This strategy would nevertheless be equivalent in our model to increasing the weight of texture-based cues. For veridical feedback, there is no such obvious strategy that would lead to the correct size match, but there are others that would nevertheless help, more often than not, to give the correct answer. If observers could learn to distinguish the size of the room they were in, they would be able to choose an appropriate size match and so provide perfect responses in the feedback conditions. No observer reported being aware of the change in room size, but any cue that covaries with room size, even with a moderate correlation, would provide a valuable signal.

For example, if the reference and comparison objects had always been presented at 0.75 m, the perceived distance of either cube would have covaried with the room expansion factor on each trial (because the distance of the cube relative to the room would have differed). This would have enabled subjects to change their criterion for an equal size match from trial to trial. Because the cubes were presented at two different distances, perceived distance of the comparison cube was not a wholly reliable correlate of the room expansion factor, but it would nevertheless provide subjects with a useful signal to determine how to adjust their criterion to reduce their errors. In fact, this was not the strategy that best fitted our data: It predicts, for example, that in the static room, performance should depend on the distance of the comparison cube alone. Instead, for all of our subjects, the relative distance of the reference and comparison cubes was a better predictor of size matches made in the static room. (Subjects also reported that they used the relative perceived distance of the two cubes to help decide how to adjust their size matches.) We calculated the predictions of subjects using this strategy, as will be shown in the following.

We assume that feedback has no effect on the perceived sizes of the reference or comparison cubes, S_L and S_R. Instead, feedback allows the subject to change their matched size, M, from trial to trial according to a cue, in this case, the relative perceived distance of the comparison and reference cubes, D_R/D_L. To fit the data, we allowed the exponent, k, applied to this ratio to vary:

$$M = [D_R/D_L]^k \cdot [S_R/S_L]^{-1}. \tag{3}$$

The perceived distances of the reference and comparison cubes, D_L and D_R, were calculated from the pre-feedback data. In fitting the model, the exponent k was allowed to vary between 0.1 and 2.

The fit of the strategy model to one data set is shown in the bottom-right panel of Figure 5. The main characteristics are similar to the remapping model, that is, a vertical spread of the lines fitting data for conditions in which the reference and comparison cubes were at different distances and lines with different slopes through the origin for conditions in which the two cubes were at the same distance. For all three observers, the exponent k in Equation 3 was less than 1 (0.33, 0.55, and 0.54). This was the only free parameter because the D_R/D_L ratio was obtained from the size matches of each subject's pre-feedback data.

Comparison of model performance

Figure 6 plots the residual error of each model for all subjects and for both types of feedback (Experiments 1 and 2). Errors in the reweighting and strategy models have been normalized by the error for the remapping model in each condition to aid comparisons between the models. The remapping model fits the data best in all cases. It should be remembered that the reweighting and remapping models each have two free parameters, whereas the strategy model only has one; hence, a slightly worse fit would be expected in the latter case. However, we compared the Akaike Information Criterion (AIC; Akaike, 1974) returned for each model during veridical feedback (Experiment 2): For two observers, the remapping model minimized AIC (evidence ratios of 9.4 and $>10^6$, respectively), and for the other observer, the strategy model was best, although the evidence ratio was lower (4.8).

We also measured, in separate blocks of trials, size matches following veridical feedback. No feedback was given during these blocks (see Figure 6C). Subjects were asked to match the sizes of the reference and comparison cubes as they perceived them and to ignore any strategy they may have adopted during feedback. The trends that were apparent during feedback were nevertheless more prominent in these postfeedback trials, with greater vertical spread. This suggests that, if subjects used a cognitive strategy of the sort proposed here, then at least some aspects of the strategy must have become automatic and difficult to "switch off" by the end of the feedback run.

Discussion

Despite our best attempts, we have found no convincing evidence that subjects have direct access to information from stereopsis and motion parallax when performing our size-matching task. Subjects appear unable to change their responses appropriately when given feedback about the size of objects as specified by these cues. Disparity and motion parallax are clearly used in the task, but these cues appear to be combined with others before the stage at which they can be used to determine the subject's

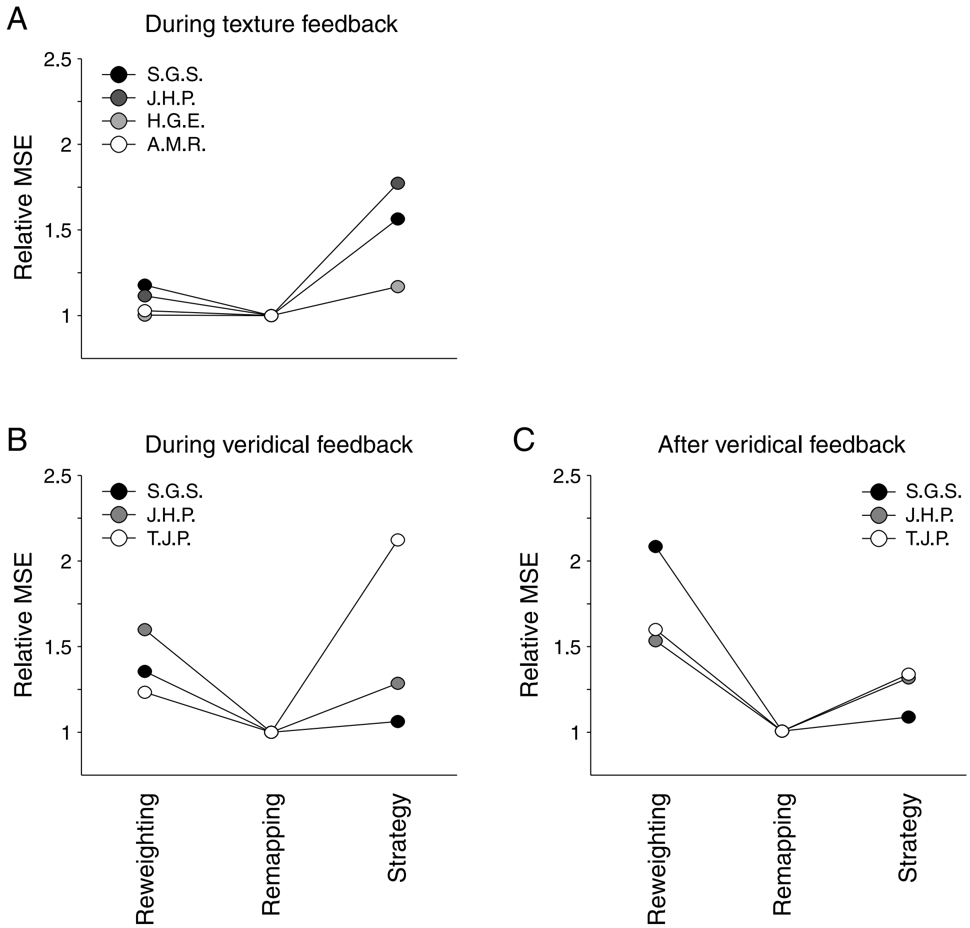

Figure 6. Performance of the three models compared. The residual mean squared errors (MSEs) for each model (see Figures 1 and 5) are shown for the texture-based feedback (A) and the stereo/motion-parallax-based feedback (B). The "strategy" model for observer A.M.R. performed so poorly that it was excluded from Panel A. To help comparisons across conditions, we normalized the MSE for the remapping model. Panel C shows MSEs for data collected without feedback, measured after the run in which stereo/motion-parallax-based feedback was given.

response. This contrasts with the result we found for texture-based cues (which we have used as shorthand to mean the information provided about the distance and size of objects if it is assumed that texture elements do not change size over time). Here, subjects were able to change their responses in accordance with the feedback provided, especially in the case of subjects who knew in advance the rules on which the feedback was based. This is what would be expected if subjects could choose to base their responses predominantly on texture-based cues. Accordingly, a reweighting model provides a good account of the data (Figure 6A). The fact that the same was not true for stereo and motion parallax cues indicates that these cues must be combined with others "mandatorily" (Hillis et al., 2002) without the subject having access to the individual cues.

Finding a lack of appropriate response to feedback is surprising. Experiments in which subjects change their responses to minimize feedback errors are much more common and present problems of interpretation, for example, in determining whether feedback has changed subjects' perceptions or whether it has simply changed their response criteria. Here, however, when feedback was correlated with stereo and motion parallax cues, the results were quite different.

Instead of the pattern expected from a reweighting of cues, subjects all produced a distinctive and, at first sight, perplexing pattern of biases in their size matches. Size constancy even in the normal, nonexpanding room was disrupted. We have presented two kinds of model that could explain this pattern of responses. According to one ("remapping"), the relationship between the distance signaled by stereo/motion parallax and perceived size is changed (Adams et al., 2001). This predicts a pattern of responses similar to our data, including a loss of size constancy in the static room. However, there are some problems with this type of explanation. To fit our data, the remapping between stereo/motion parallax signals of distance and size was quite extreme (see Table 1). It would be straightforward to test whether such remapping had occurred in a general way and, hence, could have affected other tasks. One informal piece of evidence that this is not the case is that subject S.G.S. maintained the same pattern of biases after more than a month. During this time, any recalibration between stereo/motion parallax and perceived size that might have occurred during the experiment is likely to have disappeared or at least substantially diminished. In fact, the biases of observer S.G.S. had barely altered over this period (compare postfeedback data of S.G.S. in Figure 4 with his pre-feedback data in Figure 3).

The second model that could explain the peculiar pattern of our feedback data (Figure 4) is that subjects used a strategy to try and identify the trials on which the room was expanding and those on which it was contracting. If they used this information to shift their criterion in the size-matching task (e.g., choosing apparently smaller cubes as a match when they had evidence that the room was large on that trial), then it could help explain why subjects lose size constancy in the static room. The model we have used (Figure 6), based on the relative perceived distance of the reference and comparison objects, does well at fitting the data, but it is by no means the only such strategy that would be effective at reducing feedback errors.

There are two advantages of models based on strategies to shift the criterion for a size match. First, there is no need to propose large recalibration in the size or distance signaled by cues such as stereo and motion parallax. Second, the strategies are similar to those that subjects report they are following during the experiment. They describe choosing matches around a smaller criterion size on some trials and a larger one on other trials. The methods by which subjects distinguish such trials are likely to vary, but they commonly report using the perceived distance to the reference and comparison cubes to help disambiguate the trials on which they should use different matching criteria.

It might be argued that with a more careful experimental design, we could have eliminated any correlation between the perceived distance or size of objects and the expansion factor of the room to prevent subjects from using any such strategies. We do not believe this to be the case. Our results suggest that subjects have access both to the perceived size of objects (a combination of stereo/motion parallax and texture cues) and to the size of the object with respect to the elements in the room such as bricks (equivalent to the "texture cue" alone). As a consequence, it is not possible to devise a set of stimuli that are equivalent for each room expansion factor. A similar argument has been made by Hillis et al. (2002), who found evidence for "mandatory fusion" of disparity and texture cues in signaling the slant of a surface. Although successful in explaining much of the data, the mandatory fusion model failed where it predicted very large thresholds. The authors' explanation was that another cue (in their case, the apparent homogeneity of the texture elements) was used by the subjects and that in their paradigm, as in ours, it was not possible to eliminate both cues simultaneously.

Hillis et al. (2002) and Backus (2002) used an odd-one-out task to find evidence of metamerism and, hence, mandatory fusion of cues. (Metamerism is the inability to discriminate mixtures of stimuli when the components alone would be easily distinguishable, such as confusing yellow with a mixture of red and green light.) In our experiment, when the room is not static, it seems that texture-based and stereo/motion parallax cues are not fused mandatorily in the way that has been found close to threshold by Backus and Hillis et al. Instead, our data suggest that observers have access to two distinct signals: texture-based and a combination of texture-based and stereo/motion parallax cues. The second of these, we suggest, gives rise to the perceived size of the object.

Finally, one might ask *why* subjects are able to pick out the texture-based cues and change their responses accordingly, whereas they are not able to do the same for stereoscopic and motion parallax cues. It is likely that part of the answer lies in the fact that subjects can make two independent judgments about the distance to an object, as discussed above. The distance of the object relative to the size of the room correlates with what we have called texture-based cues (by definition), and this information remains apparent to subjects even when other cues are varied. It is not obvious that there is an equivalent perceptual judgment in our experiment that varies only with the stereoscopic and motion parallax distance cue, independent of other cues. In this sense, it is understandable that subjects can respond correctly to the texture-based cues. Nevertheless, the absence of an appropriate response to stereo/motion parallax feedback is surprising, particularly given the high sensitivity of observers and robust response of visual cortical areas to these cues in isolation.

Conclusion

Our results show that the visual system cannot reweight stereo/motion parallax cues in the way that it apparently can do for texture-based cues. We conclude that stereoscopic and motion parallax cues about the distance of objects become inextricably combined with other cues so that, by themselves, they can no longer be used to influence a subject's responses.

Acknowledgments

This study was supported by the Wellcome Trust and the Royal Society. We thank Andrew Welchman and James Hillis for helpful comments on the manuscript.

Andreas M. Rauschecker and Samuel G. Solomon contributed equally in this study.

Commercial relationship: none.
Corresponding author: Andrew Glennerster.
Email: a.glennerster@rdg.ac.uk.
Address: School of Psychology and Clinical Language Sciences, University of Reading, Reading RG6 6AL, United Kingdom.

Footnotes

[1]M is the inverse of the S_R/S_L because if an object appears smaller when the room is made larger, then the observer should increase the size of the comparison object in the larger room to match perceived size. We assume that changing the size of the comparison object leads to proportional changes in perceived size.

[2]Observers S.G.S. and J.H.P. participated in both texture-based and veridical feedback experiments. Both observers completed the veridical feedback trials before undertaking the texture-based feedback trials; the pre-feedback size matches of S.G.S. (Figure 3) are similar to those obtained under veridical feedback (Figure 5), 1 month earlier. To obtain model fits in this case, we estimated simultaneously the pre-feedback cue weights and the impact of feedback on size estimates.

References

Adams, W. J., Banks, M. S., & van Ee, R. (2001). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience, 4,* 1063–1064. [PubMed] [Article]

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723.

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41,* 449–461. [PubMed]

Backus, B. T. (2002). Perceptual metamers in stereoscopic vision. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Neural information processing systems* (14th ed., pp. 1223–1230). Cambridge, USA: MIT Press.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86,* 2054–2068. [PubMed] [Article]

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research, 40,* 3725–3734. [PubMed]

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research, 33,* 919–933. [PubMed]

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *American Journal of Psychology, 75,* 485–491. [PubMed]

Finney, D. (1971). *Probit analysis* (3rd ed.). Cambridge: CUP.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research, 36,* 3441–3456. [PubMed]

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology, 16,* 428–432. [PubMed]

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science, 298,* 1627–1630. [PubMed]

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12), 967–992, http://journalofvision.org/4/12/1/, doi:10.1167/4.12.1. [PubMed] [Article]

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6,* 345–350. [PubMed]

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34,* 2259–2275. [PubMed]

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research, 33,* 813–826. [PubMed]

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43,* 2539–2558. [PubMed]

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. [PubMed]

Levitt, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(Suppl. 2), 467+. [PubMed]

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30,* 1763–1779. [PubMed]

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature, 414,* 197–200. [PubMed]

Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society America A, 2,* 343–349. [PubMed]

Taylor, M. M. (1962). Figural after-effects: A psychophysical theory of the displacement effect. *Canadian Journal of Psychology, 16,* 247–277. [PubMed]

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 663–678. [PubMed]

Triesch, J., Ballard, D. H., & Jacobs, R. A. (2002). Fast temporal dynamics of visual cue integration. *Perception, 31,* 421–434. [PubMed]

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research, 33,* 2685–2696. [PubMed]