N90-22922

PARADOXICAL MONOCULAR STEREOPSIS AND PERSPECTIVE VERGENCE

J. T. Enright
Scripps Institution of Oceanography
La Jolla, California

SUMMARY

The question of how to convey depth most effectively in a picture is a multifaceted problem, both because of potential limitations of the chosen medium (stereopsis? image motion?), and because "effectiveness" can be defined in various ways. Practical applications usually focus on "information transfer," i.e., effective techniques for evoking recognition of implied depth relationships, but this issue depends on subjective judgments which are difficult to scale when stimuli are above threshold. Two new approaches to this question are proposed here which are based on alternative criteria for effectiveness.

Paradoxical monocular stereopsis is a remarkably compelling impression of depth which is evoked during one-eyed viewing of only certain illustrations; it can be unequivocally recognized because the feeling of depth collapses when one shifts to binocular viewing. An exploration of the stimulus properties which are effective for this phenomenon may contribute useful answers for the more general perceptual problem.

Perspective vergence is an eye-movement response associated with changes of fixation point within a picture which implies depth; it also arises only during monocular viewing. The response is directionally "appropriate" (i.e., apparently nearer objects evoke convergence, and vice versa), but the magnitude of the response can be altered consistently by making relatively minor changes in the illustration. The cross-subject agreement in changes of response magnitude would permit systematic exploration to determine which stimulus configurations are most effective in evoking perspective vergence, with quantitative answers based upon this involuntary reflex. It may well be that "most effective" pictures in this context will embody features which would increase "effectiveness" of pictures in a more general sense.

INTRODUCTION

One of the central issues involved in spatial display is the question, "What is the most effective way to convey three-dimensional depth in a pictorial representation?" This article deals only with a very restricted approach to that question, being confined to representations without stereopsis and without image motion; and so the problem addressed here should probably be rephrased, "What is the *third* most effective way of conveying depth in pictures?" Such rephrasing seems appropriate because there can be little doubt that the most effective representations of the third dimension are those which involve stereopsis; and that the second most effective way to convey a feeling for depth is through use of image motion: optical flow patterns, image shear, motion parallax and the like. When both stereopsis and image motion are excluded, one is dealing with no more than third best; and the rephrased question is in some ways like asking what is the best way to participate in a footrace, subject to the precondition that the runner's feet be tied together by his shoelaces.

Nevertheless, the question of how best to convey the third dimension in a static pictorial representation has been of central concern to artists for many hundreds of years; and the result of that interest is an organized body of technique, collectively known as perspective, to deal empirically with that problem. One might well ask, then, whether there is any hope for deriving new answers to this question—if thousands of artists, throughout their careers, have been experimenting for centuries with just this objective in mind. The honest reply is that this article has no new answers to offer, no new tricks to suggest. Instead, it focuses upon two interesting phenomena involving the perception of and response to depth in illustrations—phenomena which seem to me to have the potential of providing more quantitative answers to the question, "How can depth be more effectively represented?" These phenomena suggest research programs for the future, which would address this question within certain restricted contexts, and it is conceivable that the answers might be applicable to other, more general contexts as well. The hope is that such research might provide general, quantitative rules for optimizing the depth impression which is conveyed by the stimulus field in an illustration.

PARADOXICAL MONOCULAR STEREOPSIS

The first of the phenomena of interest here is a remarkable and relatively little-known sort of depth perception which was described by the French visual scientist, Claparède, in a brief article published in 1904; he christened this visual experience "paradoxical monocular stereopsis." The essence of Claparède's message is that if certain pictures which illustrate a three-dimensional scene—drawings, paintings or photographs—are carefully examined *with one eye covered*, a truly compelling sense of depth can sometimes be obtained, an effect nearly as striking as looking into a stereoscope. Once this sort of perception has been achieved, it can be sustained while continuing to inspect the picture, and one might suspect that it results simply from thinking about and focusing attention on the illustrated subject matter. It is easy to demonstrate, however, that something unusual is involved, because the moment that the other eye is opened, to see the picture binocularly, the anomalous 3-D effect vanishes; the picture flattens out just as suddenly and completely as when one *closes* one eye while looking into a stereoscope.

High-quality, well-printed color photographs of outdoor scenes, of the sort found in magazines like *National Geographic* and *Arizona Highways*, often provide good material for demonstrating this sort of depth perception, but one of the most interesting aspects of paradoxical monocular stereopsis is how difficult it is to predict whether a given illustration will be effective in evoking the response. The compelling impression of depth is not simply a response to monocular viewing of all illustrations which show a three-dimensional scene, but to certain configurations of stimuli. The question therefore arises, "What is the most effective way to evoke paradoxical monocular stereopsis with an illustration?" This is, of course, a much more limited question than asking what is the most effective way to convey depth in a picture, but it may be more tractable. One has available the clear-cut criterion, "Does the (supplementary) depth impression flatten out, when switching over to binocular viewing?" Furthermore, although the best stimuli for paradoxical monocular stereopsis may not turn out to be fully congruent with the stimuli which are optimal for conveying a three-dimensional impression during binocular viewing, preliminary evidence suggests that if a picture is effective in evoking paradoxical stereopsis, it will at least give a satisfying and convincing impression of depth during binocular viewing.

A search of the published literature indicates that there have apparently been no systematic investigations of which kinds of pictures best evoke paradoxical stereopsis; and in fact, I have encountered less than a dozen references, in the entire 80-year interval since Claparède's (1904) initial description of the phenomenon, in which this sort of depth perception is even mentioned (e.g., Pirenne, 1970; Schlosberg, 1941; Ames, 1925; Streigg, 1923; and the references cited there). Qualitative preliminary testing indicates that there is good agreement among subjects, in the sense that certain pictures seem to be very effective stimuli for everyone, so the project of exploring stimulus optimization should be relatively easy to carry through, with a relatively modest number of subjects. And if the illustrations which are to be used were to be carefully selected, it seems very likely that an organized body of rules will emerge which characterize the optimal stimuli.

PERSPECTIVE VERGENCE

In the brief article in which Claparède (1904) described this unusual sort of depth perception, he also proposed an interesting hypothesis about the mechanisms responsible. He speculated that during monocular inspection of a picture, the covered eye would be free to make vergence movements which might correspond to the relative distances implied by the illustration (converging, then, for apparently near objects and diverging for more remote ones), just as changes in vergence accompany binocular inspection of a real, three-dimensional scene. He pointed out that vergence changes of this sort could not take place during binocular viewing of a picture because of the demand for fusion; and he further proposed that this sort of postulated vergence movement might be responsible for the compelling sense of depth evoked during monocular viewing. Apparently there has been no test of Claparède's hypothesis, nor even any restatement of it, in the subsequent 80 years; a recently initiated research program, however, has provided compelling evidence that Claparède was essentially correct in his speculation about eye movements (Enright, 1987a; Enright, 1987b). Vergence changes of the sort he postulated do, indeed, take place when inspecting a picture of a three-dimensional scene with one eye covered—though whether those eye movements are responsible for paradoxical stereopsis remains an open question, and one which will be much more difficult to investigate.

METHODS

The experimental equipment which was used in this eye-movement research is extremely simple, both in principle and in practice (Fig. 1). The subject sits with head held firmly in place by a bite board and headrest while two video cameras monitor eye position from somewhat below the line of sight. The output of the cameras is combined with an image splitter and recorded for subsequent analysis; the sum of the two distances between iris margins and the image-splitting line is an index for vergence state. The illustrations to be viewed are mounted at about 30 cm from the subject's eyes, and an obstruction is placed a few centimeters in front of the nondominant eye, at a level which hides the picture from that eye, but permits the camera to record eye position. While viewing the picture monocularly, the subject changes fixation at intervals of 2 to 3 sec, between points which are at different implied distances away. Single-measurement precision of the recording method is about 6 arcmin for each evaluation of eye position, and averaging results over repeated tests can further reduce the influence of random measurement error; but the between-trial variability within a given test session for a given subject and target is sufficiently large that a more precise monitoring technique could not appreciably improve the reliability of the estimates of average response; the variability in the eye movements from one refixation to the next limits precision of the estimates, as reflected in the standard errors.

RESULTS

An excerpt from a longer recording is shown in Fig. 2, made while a subject changed fixation from the upper front corner to the upper back corner of the perspective drawing of a small box (target illustrated in Fig. 3). Concurrent with the recording, a three-position switch, which was connected to two tone generators, was activated by the subject to indicate the fixation point; the timing of those signals is shown as open and solid bars in Fig. 2. It is, then, quite clear that convergence occurred while fixating on the apparently nearer corner of the box, and divergence while fixating on the farther corner. A simple summary value for the typical vergence-change response can be obtained from such a recording, based on measuring one value of vergence state for each steady-state fixation, and then calculating differences between successive values; in this case, the average change in vergence, over 20 fixations, was 68 arcmin ± 8 arcmin. In Fig. 3, this summary value is shown for Subject 1, along with five other values for her, each with this same target, each recorded on a different day; and values of average vergence change are also shown there for another eight subjects with this target. Average vergence change, based on the method of calculation, could in principle also be negative (i.e., contrary to the perspective implication of the drawing); in fact, however, all 24 measured values are positive, and all except one of the results are statistically significant, most of them at the 0.01 level. In other words, the subjects all showed consistent vergence changes during changes in fixation point in this drawing; and those vergence changes corresponded in direction with the relative distances implied by the perspective of the drawing. For those who may be concerned about the reliability of this simple and unconventional method of recording eye movements, it is worth mentioning that the basic result of Fig. 3 has now been replicated for other subjects in two other laboratories, each of them using a fundamentally different and more familiar measurement technique. I have proposed (Enright, 1987a) that these oculomotor responses to pictorial representations be called "perspective vergence."

Before considering additional details of the responses which have been measured for other kinds of illustrations, it seems worthwhile to try to place perspective-vergence responses into some sort of broader context. A phenomenon which is now called "proximal vergence" has long been known to visual physiologists, an eye-movement response which has been attributed to "knowledge of nearness" (Maddox, 1893). Although vergence responses to perspective representations have not been previously studied, it is probably appropriate to consider perspective vergence to be a subcategory of "proximal vergence" (Hokoda and Ciuffreda, 1983). It is important, however, to distinguish between these responses and another subcategory known as "voluntary vergence": some trained subjects can cross or uncross their eyes at will, even in total darkness. Many lines of evidence indicate, however, that the eye-movement responses to perspective illustrations are instead the result of an involuntary reflex. It is conceivable—even likely—that training or an "act of will" might enhance the responses, but fully naive, untrained subjects also show comparable behavior in their first test session—even subjects who are fully unaware that convergence is the appropriate response to objects which are nearby. They show this response even though they are uninformed about the purpose of the experiment, even though they have no visual feedback or other clues to tell them whether vergence has changed—much less whether the response was "as intended." Perspective vergence is an automatic response to components of the visual stimulus field—truly a reflex. Furthermore, at least certain components of the stimulus field which evoke this kind of response are apparently not a reflection of learning or prior experience, but instead represent built-in constraints on the visual system—although it seems likely that "learning" may also play a role—that prior visual experience with our three-dimensional world may build upon and supplement those components which are "hard-wired" into the system. Because of the reflex nature of the responses, an evaluation of illustrations, in terms of the magnitude of the vergence responses evoked, represents something far more substantial than can be achieved by asking for subjective opinions about picture quality.

An experimental program has been initiated, designed to determine what features of an illustration enhance or inhibit this oculomotor response. The results of Fig. 4 summarize some of the kinds of data which have been obtained, with modest variations on the compositional theme of a single rectangular box. Despite the large inter-subject differences in response magnitude for a given picture, as shown in Fig. 3, there are remarkably consistent cross-subject *changes* in response magnitude for particular alterations in the picture; hence, the ratio of response for a given picture to the same subject's response for a standard, represents a reliable way of demonstrating the relative effectiveness of various representations in evoking perspective vergence. Doubling the size of the picture in all dimensions, for example, reliably led to an increase of about 50% in response magnitude (Fig. 4 vs. Fig. 4B); inverting the picture led to a reduction in response (Fig. 4A vs. Fig. 4C), with 7 of 9 subjects showing smaller vergence changes. A reduction in the inclination of the box (with only minor other modifications in line spacing) led to a drastic reduction in response magnitude (Fig. 4B vs. Fig. 4D); for 8 of the 9 subjects, the response was even smaller than that to the "standard" picture, which shows a box half the size (Fig. 4A). When a cross-hatched lid was superimposed upon a box which was in the relatively ineffective orientation, response magnitude increased for all 9 subjects (Fig. 4D vs. Fig. 4E), but when a similar lid was superimposed on a box with more effective orientation, it tended to reduce the response (Fig. 4A vs. Fig. 4F; 8 subjects out of 9). In all cases, there was remarkably good cross-subject agreement in the way in which a given change in the drawing affected magnitude of the response (details in Enright, 1987a).

One other closely related kind of target has been tested, which is not shown in this figure; three-dimensional cardboard models of the boxes shown in Figs. 4A and 4D were constructed and photographed from 30 cm with illumination which produced a distribution of light and shadow, and prints of those photos, at appropriate scaling, were tested as targets. The rationale for this approach is that shading might enhance the resulting vergence changes. In these tests there was indeed a slight but significant increase in response for the box shown with suboptimal orientation (Fig. 4D), but no significant change—in fact a slight decrease—for the more optimally oriented box (Fig. 4A).

The vergence responses of this same group of 9 subjects have also been tested with a set of more complex pictorial representations: photographs which reproduce five classical paintings and an etching; and those experimental results have offered further hints about the kinds of stimuli which can be effective in evoking perspective vergence. By using a portrait by Rembrandt, for example, statistically significant vergence changes in the appropriate direction (nearly as large as those for the "small-box" drawing [Fig. 3]), were evoked in all 9 subjects by a change in fixation from the nose to the ear of the portrayed philosopher and back again, although no suggestion of *linear* perspective was evident in the picture, and the implied difference in distance between the fixation points was quite small (ca. 10 cm, at a distance of 2 to 3 m from the viewer). One landscape scene evoked strong responses in every subject tested, and another outdoor scene, in which linear perspective was conspicuous, did not lead to statistically significant results for *any* of the subjects. Again, then, there was very good cross-subject agreement, in terms of which artworks were effective stimuli and which were not.

DISCUSSION

The cross-subject consistency in terms of response magnitude demonstrates that in measuring perspective vergence we are dealing with relatively general characteristics of the oculomotor response system; but the experiments conducted so far do no more than define a few of the dimensions of the multidimensional coordinate system implied in the question, "What is the optimal stimulus for this response?" There seems to be clear non-additivity (a cross-hatched surface between fixation points enhances a response, or it does not, depending on context), which considerably complicates the exploration of these dimensions. Furthermore, it is by no means clear that the rules which might be derived from a line drawing of a cubical box can be generalized to other sorts of figures; nor do the available data define an optimum point in any stimulus dimension. Consider, for example, the conspicuous effect of tilt of the opening on responsiveness (Fig. 4B vs. 4D): while it seems clear that a 22° tilt (4B) is much more effective than an 11° tilt (4D), there is presumably a continuous function relating responsiveness to inclination in the illustrated box, with a maximum someplace between 0° and 90°; and it may well be that 22° is far removed from that optimum tilt. The necessary experiments to explore this dimension should be enlightening—but the existence of nonlinearities cautions against overgeneralization.

The consistently positive responses to the Rembrandt portrait demonstrate that the dimensions which must be explored in any complete attempt to define optimal stimuli go far beyond the systems of lines and angles which constitute linear perspective. The opportunity to explore the question of stimulus optimization offers exciting promise for the future, but it is self-evident that the available data do not even adequately define the dimensions of the problem. Beyond the issue of stimulus optimization, the intriguing possibility exists that perspective vergence responses may provide an objective metric for evaluating the general effectiveness of an attempt to convey depth in a picture: that oculomotor responsiveness may prove to be well correlated with subjective perceptual responsiveness to pictorial implications of depth. Such a correlation would be a necessary—but not a sufficient—condition for establishing the validity of Claparède's most interesting speculation: that perhaps vergence movement itself contributes to the perception of paradoxical monocular stereopsis.

ACKNOWLEDGMENT

Research supported by Grant Number BNS 85-19616 from the National Science Foundation.

REFERENCES

Ames, A., Jr. (1925): The illusion of depth from single pictures. J. Opt. Soc. Amer., 10, 137-148.

Claparède, E. (1904): Steréréoscopie monoculaire paradoxale. Annales d' Oculistique, 132, 465-466.

Enright, J. T. (1987a): Perspective vergence: oculomotor responses to line drawings. Vision Res., 27, 1513-1526.

Enright, J. T (1987b): Art and the oculomotor system: perspective illustrations evoke vergence changes. Perception, 16, 731-746.

Hodoka, S. C. and Ciuffred, K. J. (1983): Theoretical and clinical importance of proximal vergence and accommodation. In Vergence Eye Movements: Basic and Clinical Aspects. (Ed. by C. M. Schor and K. J. Ciuffreda) pp. 75-97. Butterworths, London.

Maddox, E. E. (1893): The Clinical Use of Prisms; and the Decentering of Lenses. Second Edition. Bristol, John Wright and Sons.

Pirenne, M. H. (1970): Optics, Painting and Photography, p. 95. Cambridge Univ. Press.

Schlosberg, H. (1941): Stereoscopic depth from single pictures, Amer. J. Psychol., 54, 601-605.

Streiff, J. (1923): Die binoculare Verflachung von Bildern, ein vielseitig bedeutsames Sehproblem. Klin. Monatsbl. Augenheilkunde, 70, 1-17.

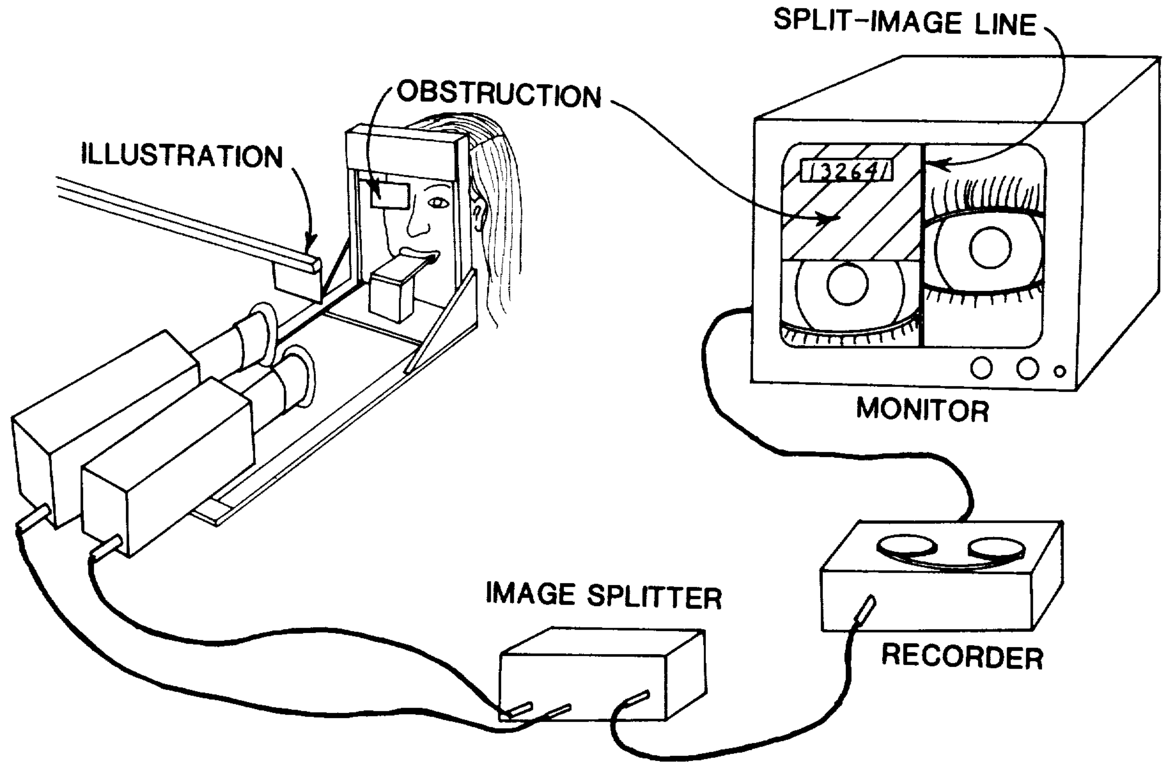

Figure 1.– Diagram of the equipment and setup used for recording eye position while viewing illustrations.

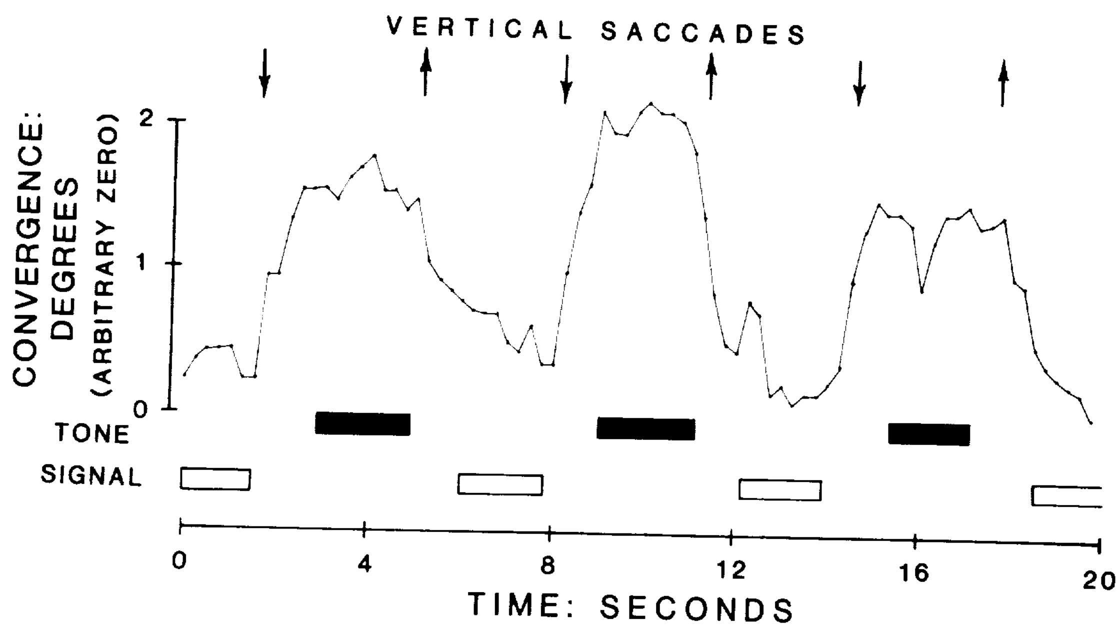

Figure 2.– Excerpt from a recording made while Subject 1 alternated monocular fixation between apparently nearer and apparently farther topside corners in a line drawing of a small cubical box (picture shown in Fig. 3 and as "Standard" in Figure 4). Bars beneath graph correspond to the timing of tone signals; solid bars represent fixation on "near" corner, open bars represent fixation on "far" corner. (Reprinted with permission from Vision Res. 27, J. T. Enright, "Perspective Vergence: oculomotor response to line drawings," Copyright 1987, Pergamon Journals Ltd.)

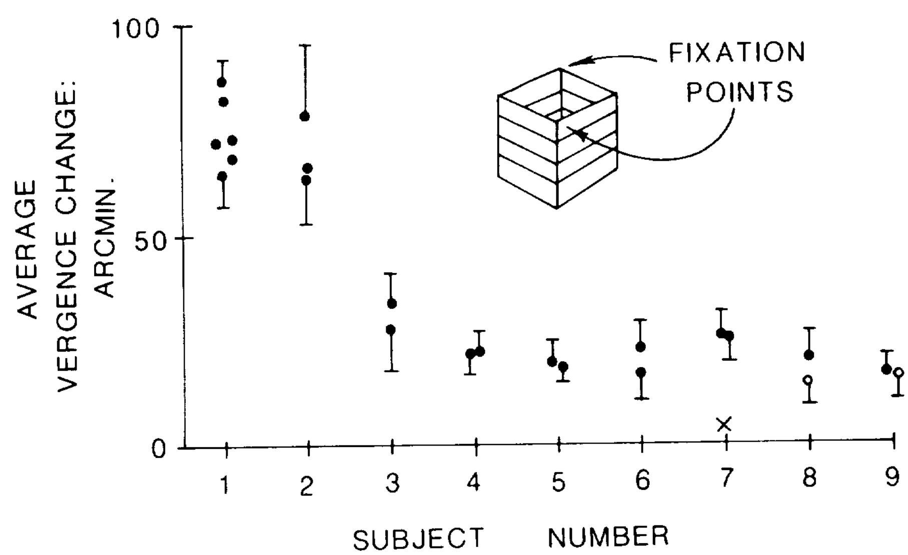

Figure 3.– Summary of average vergence changes made by 9 subjects in conjunction with changes in fixation on the line drawing of a small cubical box; each point represents average value during a separate test session, with standard errors based on N of 10 (20 changes in fixation).

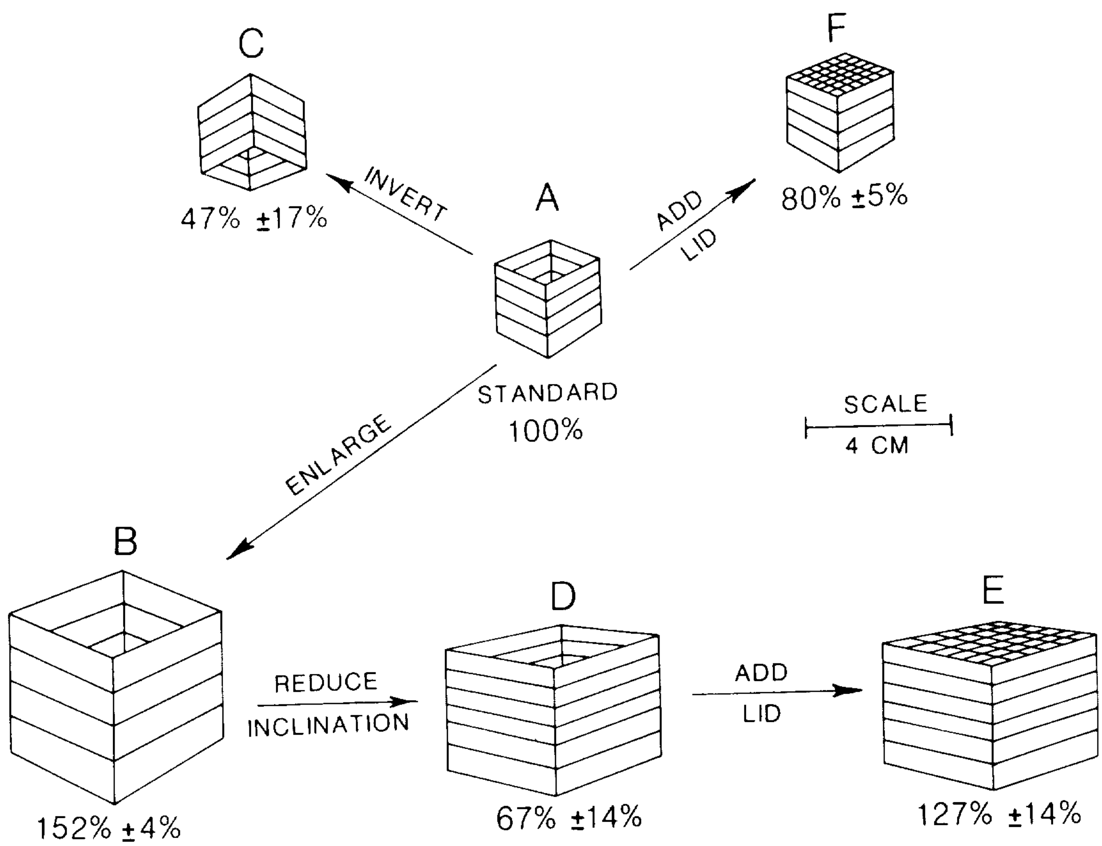

Figure 4.– Cross-subject values, and their standard errors, for 100 times the ratio: "average vergence change for a given drawing," divided by the same-subject value of "average vergence change for 'standard' illustration." N = 3 for part B, N = 9 for all other parts.