

Context Effects and Their Interaction with Development: Area Judgments

Gerd Gigerenzer

Universität Salzburg, Austria

Hans R. Richter

Universität München, West Germany

The perceptual constancy view posits that children perceive rectangular areas dependent on shape, whereas adults do not. Following Brunswik (1934a), we call this dependency the "limited perceptual constancy of area under variation of shape." A series of four experiments (with 150 children and 80 adults) using ordinal paired comparisons rejected the general equation $a_i = f(A_i, S_i)$ as a model of limited area constancy (a_i is the perceived area, A_i and S_i the physical area and shape of rectangle i). Children judged area dependent on shape, but dependent on the covariation C_{set} of area and shape in the experimental set of rectangles, rather than on the particular shape of the rectangle judged. A quantitative version of this covariation mechanism that seems to generate limited area constancy is proposed, and its relation to Parducci's range-frequency theory is discussed. Not one child's judgments could be predicted from a height + width rule (assuming identical linear psychophysical functions). We demonstrate and correct serious errors in Anderson and Cuneo's (1978a, 1978b) claims for a height + width rule in children. Our results support a three-step developmental model for paired comparisons of area: (1) centering on one side of the rectangle, (2) limited area constancy, that is, $a_i = f(A_i, C_{set})$, and (3) area constancy.

How are the areas of rectangular shapes judged by children, and how are they judged by adults? Beginning in 1916, this question was studied in Jena by Peters (1933), and later Brunswik took it up (1934a, 1934b) in Vienna. The conclusions

This article was completed while Gerd Gigerenzer was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California. We are grateful for financial support provided by the Spencer Foundation and by the Volkswagen Foundation, which supported parts of this work by an Akademie Stipendium to Gerd Gigerenzer. We would like to thank the children and the staff of the Montessori Institute, Munich, for their cooperation, Wolfgang Hell, Kathleen Much, Kurt Müller, Alan Parducci, Victor Sarris, and Colleen Surber for helpful comments and suggestions on earlier versions of this article, Peter Sedlmeier for his assistance in statistical analysis, and Fritz Wilkening for kindly sending us the reliabilities and data from his 1979 study.

Correspondence and requests for reprints should be sent to Gerd Gigerenzer, Institut für Psychologie, Universität Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria.

were that (1) perceived area is dependent on the shape of the rectangle, and (2) the degree and the direction of this dependency changes with age. This interaction between perceived area and shape has been theoretically treated within the framework of Gestalt psychology (Rubin, 1921; Werner, 1933) and within that of the “generalized problem” of perceptual constancy (Brunswik, 1934a), which in turn influenced Piaget’s approach to the evolution of perceptual constancies (Piaget, 1969). According to these views, perceived area was understood in German-speaking countries to be different from the mere “sum” of its parts, height, and width (Bühler, 1913; Granit, 1921; Hempstead, 1900; Leiser, 1916; Lenk, 1926; Rausch, 1949, 1952, 1964; Sander & Volkelt, 1962; Selinka, 1939); this nondecomposability is reflected, although in different ways, in the recent distinction between integral and separable dimensions in English-speaking countries (Garner, 1974; Monahan & Lockhead, 1977; Shepp, 1978; Shepp, Burns, & McDonough, 1980; Tighe & Shepp, 1983; Weintraub, 1971).

All these views share a lowest common denominator: Perceived area is dependent on shape. We call this dependency the “limited perceptual constancy of area under variation of shape,” following Brunswik (1934a), or, for short, “limited area constancy.” The classical constancy problem refers to the perceived constancy of a property of an object (such as size) under variation of external conditions (such as its distance to an observer). Brunswik’s “generalized problem” refers to the perceived constancy of a property of an object (such as area) under variation of other properties of the same object (such as its shape or color). How limited area constancy should be modelled, however, has not always been made clear, but most of the authors mentioned above seem to agree with the following general formulation:

$$a_i = f(A_i, S_i) \tag{1}$$

where a_i is the perceived area of a rectangle i with physical area A_i and shape S_i . The function f accounts for the dependency and is as yet unspecified. Although Equation 1 is the most common interpretation of limited area constancy, it is not the only one possible. For instance, Equation 1 is based on the specific assumption that the dependency can be explained by the shape S_i of a single rectangle i , without taking into account the distribution of shapes in the whole series of rectangles presented. Most of the authors mentioned above (with exceptions such as Shepp, Burns, & McDonough, 1980) view the development of area perception as an increase in perceptual constancy: Children perceive area as dependent on the shape (Equation 1), whereas adults show, at least approximately, perceptual constancy of area under variation of shape (Equation 2):

$$a_i = f(A_i) = f(A_i, S_i), \text{ for all } S_i \in S \tag{2}$$

To summarize, what we call the perceptual constancy view assumes that perceived area is dependent on shape in young children but is approximately

independent in adults. Development of area perception is, in this view, the increasing ability to abstract from “irrelevant” cues such as shape.

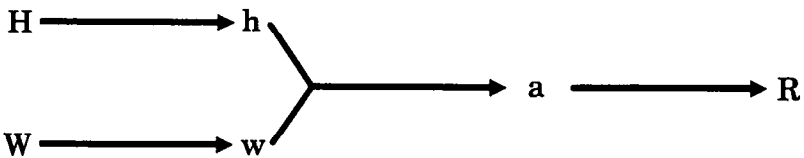
COGNITIVE ALGEBRA HYPOTHESES

The last decade has witnessed the rise of another point of view, however, which assumes that perceived area is arrived at by adding or multiplying height and width, the assumed perceptual “parts.” Anderson and Cuneo (1978a), Cuneo (1980), and Wilkening (1979, 1980) concluded from their experiments that children (5–6 years old and younger) use an adding rule to calculate perceived area from perceived height and width, whereas adults use a multiplying rule. Let us denote perceived height as h_i , perceived width as w_i , and perceived area as a_i ; then these hypotheses can be written as follows:

$$a_i = h_i + w_i \quad (\text{height} + \text{width rule}) \quad (3)$$

$$a_i = h_i w_i \quad (\text{height} \times \text{width rule}) \quad (4)$$

The theoretical framework is called cognitive algebra or information integration theory, and its methodology is called functional measurement (Anderson, 1981, 1982). In the theoretical framework, judgment of area is decomposed into three steps (Figure 1). First, the physical height H of a rectangle is transformed by a psychophysical function f_h into perceived height h ; similarly, physical width W is transformed by f_w into w . Note that h and w are assumed to be independent. Second, h and w are integrated into perceived area a by an integration rule I , that is, either by adding or multiplying. Third, a is transformed by the judgment



psychophysical
function

integration
rule

judgment
function

$$h = f_h(H)$$

$$w = f_w(W)$$

$$a = I(h, w)$$

$$R = g(a)$$

Figure 1. Theoretical framework of information integration theory (cognitive algebra).

function g into the judged area R on a rating scale. If one accepts this framework, developmental changes might be postulated either in f , I , or g , or in some combination thereof. The authors mentioned above prefer the interpretation of development as a change in the integration rule, from adding to multiplying.

Doubtless the most surprising conclusion was that young children used a height + width rule. This hypothesis implies that children are able to decompose area into its parts, height and width, but that they, because of some unknown innate developmental factor, use the wrong algebraic operation to integrate the parts again. Meanwhile, researchers using the same framework and methodology concluded that children integrate information of many kind by an adding rule (e.g., Butzin & Anderson, 1973; Cuneo, 1982; Kun, Parsons, & Ruble, 1974; Miller, 1982; Singh, Sidana, & Saluja, 1978; Singh, Sidana, & Srivastava, 1978; Wilkening, 1980).

This conclusion led to the postulate of a new general principle of cognitive development:

. . . the height + width rule is a manifestation of a primitive integration process that operates across many different situations. Various stimulus cues are evaluated as relevant, depending on the specifics of each situation. In each situation, however, the relevant cues are combined by the same adding-type integration. . . . Overall, it seems reasonable to think that a single general-purpose integration process underlies these diverse judgments (Anderson & Cuneo, 1978a, pp. 366–367).

In addition to the height + width and the multiplying rules, we will consider three other simple algebraic rules, discussed in the cognitive algebra tradition:

$$a_i = \max(h_i, w_i) \quad \text{(longer-side rule)} \quad (5)$$

$$a_i = h_i \quad \text{(height-only rule)} \quad (6)$$

$$a_i = w_i \quad \text{(width-only rule)} \quad (7)$$

The longer-side rule predicts that children use the longer of height and width when judging the area. Different from the height-only and width-only rules, this rule implies that children attend to both sides of the rectangle, but base their judgment solely on the larger one. Previous results concerning this perceptual strategy are contradictory. Verge and Bogartz (1978) concluded that children center on the longer side in an area-matching task, whereas Leon (1982) found no evidence of such a rule in ratings of areas.

The height-only and width-only rules were considered by Anderson and Cuneo (1978a) to represent Piaget's theory of centration. Since their data reject both rules, they claimed to have falsified Piaget's centration hypothesis. We shall

not consider these rules as representing Piagetian theory, since, when Piaget applied his “law of relative centrations” to height and width (1969, pp. 8–12), he gave quite a different theoretical account. Piaget referred to an overestimation of whatever portions of a figure are attended to, rather than to a selective response to only one aspect of the figure. The law of relative centrations predicts an overestimation of the height of a rectangle if it is paired with a shorter width and vice versa.

This means that Piaget’s law predicts a *dependency* between perceived height and width of rectangles. This dependency, however, differs from the height-only rule and the width-only rule as well as from the height + width rule and the multiplying rule, since all these cognitive algebra models are based on the assumption that h and w are independent. Therefore, we shall consider these 2 one-dimensional rules as hypotheses stemming from the cognitive algebra framework rather than from Piaget’s. Similarly, as has already been pointed out by Wilkening (1979), logical multiplication, which according to Piaget is necessary for conservation, is not the same as Equation 4 in the cognitive algebra framework. Piaget’s hypothesis, that judged height is dependent on width, is similar to Equation 1 rather than to Equations 6 and 7.

To summarize, the cognitive algebra view differs from the perceptual constancy view (as stated in Equation 1) in two fundamental respects: It assumes that perceived area is independent from shape in all stages of development, and it uses height and width rather than area as basic theoretical concepts, thereby assuming that area can be decomposed into independent height and width in all stages of development. Both views explain judged area exclusively by properties of the single rectangle (or pair of rectangles) to be judged, but do not include properties of the series of rectangles in which a rectangle to be judged is embedded.

There are serious methodological problems present in the early research in German-speaking countries, such as Peters’s work (1933), and such problems persist, although in different guises, in recent cognitive algebra research. In what follows, we shall first briefly review three major problems in the cognitive algebra tradition and add some new information. Second, we shall propose a research strategy, that we used in the series of experiments reported, and which eliminates some of these problems.

Problems with Previous Tests of Hypotheses

Power and Asymmetric Testing. One fundamental problem in cognitive algebra research is that the adding and multiplying hypotheses were accepted according to different standards—widely different—as will be seen. The use of a multiplying rule was determined by the presence of a significant (bilinear) interaction between height and width, while the use of the height + width rule was inferred from a *nonsignificant* interaction. It is well known in theory, although

sometimes forgotten in practice, that no conclusion can be drawn from a nonsignificant result unless it has been shown that the test had a large power (Sedlmeier & Gigerenzer, 1989).

The power of a test is the long-run probability of getting a significant result if the alternative hypothesis is true. The alternative hypothesis of interest, both for the perceptual constancy and the cognitive algebra views, is an interaction pattern. Note that the power increases with the effect size, which in turn depends on the perceptual mechanism *and* the design. If the dependency of perceived area from shape were on the order of Piaget's (1969, p. 11) dependency of height on width, around 6%, or on the order of Mach's illusion, from 2% to 5% (Richter, 1988), then such deviations from either parallel curves or linear fans need a powerful test to be detected, as illustrated by Gigerenzer and Murray (1987, p. 98).

What is the power of previous tests? We do not know, since previous studies, with the important exception of those of Anderson and Cuneo (1978a), from which the height + width rule was originally inferred, did not estimate the power. Nevertheless, there are two different statements in the literature, which cannot both be true. Bogartz (1978) calculated the predictions of both the height + width rule and the multiplying rule (assuming in both cases linear judgment functions) for Anderson and Cuneo's designs. Bogartz showed that the differences between both predictions were small compared to the standard errors of the means (only about $\frac{1}{3}$ of the standard error), and consequently argued that the experiments lacked the power to detect the interaction predicted by a multiplying rule. In their reply, however, Anderson and Cuneo (1978b) argued that the power was greater than .99. We shall resolve this contradiction.

We reconstructed both Bogartz's and Anderson and Cuneo's calculations and recalculated a power of only .09 for Anderson and Cuneo's (1978a) Experiments 2, 5, and 6, and .06 for Experiment 3. (These are the *same* four experiments for which Anderson and Cuneo claimed a power greater than .99.) This finding clearly supports Bogartz's conjecture. There are two reasons for these widely different power values. First, Anderson and Cuneo (1978b) pooled the data of all four experiments, that is, of four tests of interactions, into their calculation of a single power value. This artificially increased n by a factor of 4 and resulted in a large overestimation of the true power of each of the four tests. Second, they did not follow standard texts on power calculations (e.g., Cohen, 1977, pp. 369–377) for calculating the effect size of the interaction, but used a formula from Dixon and Massey (1969, p. 270), which is not appropriate for effect sizes of interactions, but rather for tests between two means. This choice increased the already inflated power value. Expressed as Cohen's f , the true effect size in Anderson and Cuneo's study (i.e., the difference between the predictions of the height + width and the multiplying rule) was .10, which is a "small" effect size in Cohen's terminology. Note that the sample effect sizes actually found were much larger, .22, .40, and .29 in Experiments 2, 5, and 6, respectively, but *none*

Table 1. Reliability of Ratings

Age Groups	Mean <i>r</i>	Standard Deviation	Mean error $1 - r^2$
5-6-year-olds	.67	.218	.51
7-9-year-olds	.79	.154	.37
Adults	.96	.025	.08

of these was significant! Thus the interaction tests were not sensitive enough to detect effect sizes that were even two or four times as large as that predicted by the multiplying rule or by any other interaction of similar size.¹

Therefore, the prior probability *not* to detect an interaction, if there is one, was more than 90% (beta), whereas that of the opposite error (alpha) was only 5%. The issue of power is, consequently, a question of *symmetry* (Sedlmeier & Gigerenzer, 1989): Anderson and Cuneo's tests led to widely different probabilities of correct acceptance for the height + width rule on the one hand, and for a competing interaction hypothesis, such as an interaction between area and shape, on the other.

Age-Dependent Loss of Reliability. One possible source of low power is low reliability in children's area judgments. What do we know about the reliabilities? Studies conducted in the cognitive algebra framework, such as Anderson and Cuneo's (1978a), Butzin and Anderson's (1973), Cuneo's (1980), and Wilkening's (1979, 1980) made general claims that young children can use rating scales in a reliable way. *None* of these studies, however, reported the actual reliabilities. We replicated Wilkening's (1979) study (see Experiment 1) and calculated the reliabilities. Table 1 shows a strong, age-dependent loss of reliability. The average intraindividual error in area ratings, as measured by $1 - r^2$, drops from .51 in the youngest group to .37 in 7-9-year-olds to .08 in adults.

In order to check the objection that children's low reliabilities were a result of peculiarities of our replication, we asked Fritz Wilkening for the reliabilities of his 1979 study, which he kindly calculated and sent to us. His values for $1 - r^2$ are almost identical with ours for the 5-year-olds and 8-year-olds: .48 and .36.² The source of this large error in young children is unknown.

¹ The power of the height \times width interaction test in the experiments of Anderson and Cuneo (1978a) was calculated from the following information: (1) the degrees of freedom reported, ranging from $F(4,24)$ to $F(4,56)$; (2) the pooled mean square error reported by Anderson and Cuneo (1978b) (as an estimation of the population variance); and (3) the predictions of the multiplying hypothesis, as specified in Table 5 in Bogartz (1978). The effect size was calculated using Cohen's (1977, p. 371) formula for the standardized effect size measure of interaction. For example, for Experiment 6 in Anderson and Cuneo (1978a), Cohen's (1977) Table 8.3.15 gives a power of .09.

² Wilkening's 11-year-olds are not considered here, because we had no such group; their value was .22, which also supports the age-dependent decrease in reliability. The only difference occurred

We suggest that the rating scales used in previous research pose a task too difficult for young children: The child must be able to memorize and maintain a constant anchor and metric throughout all judgments (Gigerenzer, 1983a, 1983b; but see Wilkening, 1983). This reasoning would locate the source of error in the judgment task, and not in children's perception, that is, in g instead of a (see Figure 1). Whatever the source is, this large error has an important effect when combined with asymmetric hypothesis testing: Any interaction pattern (not only a diverging fan) is less likely to be detected in younger children than in older children and adults.

Number of Unknown Parameters: A Crucial Experiment Reevaluated. In the cognitive algebra framework the judgment functions, the integration rules, and the psychophysical functions are unknown parameters. In none of the studies mentioned above can these functions be independently estimated from the data. The data support only *combinations* of integration rules *plus* judgment functions, but not integration rules or judgment functions *per se*. For instance, it is well known that parallel curves do not imply the use of a height + width rule. Combinations consisting of an integration rule *and* a judgment function imply parallelism, but such combinations may or may not include a height + width rule. For example, a multiplying rule *plus* a logarithmic judgment function implies parallelism, too (for more examples, see Birnbaum, 1982, p. 451). Surber (1984) argued that it is routine for experimenters to ignore this issue and simply assume that the judgment function is linear throughout all ages.

As a solution, two-operation tasks were proposed, in which the combined area of two rectangles had to be estimated. We know of a single experiment using a two-operation task with 5-year-olds (Anderson & Cuneo, 1978a, Experiment 6) that attempted a crucial test between two competing hypotheses. These were the height + width rule *plus* linear judgment function and the multiplying rule *plus* logarithmic judgment function. The former predicted parallel curves. Since the interaction did not reach significance, the authors inferred that their "results imply that the judgments of single rectangles follow the height + width rule" (p. 353). This experiment is still cited as the strongest evidence available: "The height + width rule can hardly be called artifactual in light of these results" (Silverman & Paskewitz, 1988, p. 77). Let us examine this "evidence."

In this crucial experiment the interaction predicted by the competing hypothesis (multiplying rule *plus* logarithmic judgment function) was not quantitatively specified, as often is the case in asymmetric significance testing. The authors

with respect to Wilkening's adults, whose average value was .24, attributable to two highly unreliable subjects. Furthermore, our result is supported by Bogartz (1978, p. 382), who estimated that the error variance of the children's mean ratings in Anderson and Cuneo's Experiment 1 (1978a) was four times as large in 5-year-olds as in 11-year-olds, and by Lohaus (1986), who reported that the reliabilities of magnitude ratings (such as used by Wilkening, 1979) decreased from .71 in adults to .37 in 5-6-year-olds (Kendall's τ).

relied only on the *nonsignificance* of the interaction test—but recall that we estimated a power of only .09 for this test.

Gigerenzer and Murray (1987, p. 100) specified the predicted interaction. Their analysis revealed that (1) children's mean area ratings followed the predictions of the multiplying rule *plus* logarithmic judgment function rather than those of the height + width model, and (2) *even if mean ratings had coincided exactly with the predictions of the multiplying model, the interaction would not have reached significance*. Even then, the height + width rule *plus* linear judgment function would have been erroneously judged as supported. Thus, contrary to Anderson and Cuneo's claim, their data support the multiplying model and *not* the height + width rule *plus* linear judgment function. For a detailed exposition of this erroneous claim see Gigerenzer (1983b) and Gigerenzer and Murray (1987, pp. 91–103).

This sole extant crucial experiment epitomizes how nonspecification of the predictions of the alternative hypothesis in asymmetric significance testing, large age-dependent error, and low power can combine to elicit unwarranted inferences.

Symmetric Testing Based on Paired Comparisons

In an attempt to resolve these three problems, we adopted in the following experiments a symmetric testing procedure based on ordinal paired comparisons. The use of paired comparisons instead of area rating allowed us to dispense with all parametric assumptions about the judgment function, for example, the assumption of linear or logarithmic shapes. In an ordinal paired comparison task, the child simply has to point to the rectangle in a pair that appears to have a greater area (similar to Siegler, 1976, 1978). We have had to make assumptions about the psychophysical functions (see Figure 1), but, with one important exception, we have had to assume only *monotone* psychophysical functions. Thus the overall number of unknown parameters is reduced. In order to avoid the problems of asymmetric testing, we used a symmetric testing procedure, in which each hypothesis was tested against the data with the same probability of erroneous rejection. More precisely, we reformulated each hypothesis as a statistical null hypothesis and specified its predictions, as described in the next section.

For ordinal paired comparisons, the hypotheses above can be restated as follows:

$$\begin{aligned}
 a_i > a_j &\text{ iff } f_a(A_i) > f_a(A_j), \text{ and} \\
 a_i = a_j &\text{ iff } f_a(A_i) = f_a(A_j) \qquad \qquad \qquad \text{(area constancy),} \quad (8)
 \end{aligned}$$

where A_i, A_j are the physical areas of the rectangles i and j , respectively, and f_a is the psychophysical function. It is sufficient to assume that f_a is a strictly

monotone function in order to derive the prediction that ordinal paired comparisons follow physical area. Equation 8 gives the same predictions as the multiplying rule in the cognitive algebra view. But, area constancy does not mean multiplying. Limited area constancy implies that no strictly monotone function f_a exists, since the same physical areas can be perceived to have different areas, dependent on the shape. The predictions of the remaining hypotheses for paired comparisons are:

$$a_i > a_j \text{ iff } \max(f_h(H_i), f_w(W_i)) > \max(f_h(H_j), f_w(W_j)) \text{ (longer-side rule)} \quad (9)$$

$$a_i > a_j \text{ iff } f_w(W_i) > f_w(W_j) \quad \text{(width-only rule)} \quad (10)$$

$$a_i > a_j \text{ iff } f_h(H_i) > f_h(H_j) \quad \text{(height-only rule)} \quad (11)$$

$$a_i > a_j \text{ iff } f_h(H_i) + f_w(W_i) > f_h(H_j) + f_w(W_j) \quad \text{(height + width rule)} \quad (12)$$

If “=” is substituted on one side for “>”, it is substituted on the other side too, as in Equation 8. It is sufficient to assume that f_w and f_h are monotone functions in (10) and (11) respectively, and that f_w and f_h are identical monotone functions in (9), in order to derive the prediction that area judgments follow physical height or width. Predictions of hypotheses (8) to (11) are the same for any monotone function f . This would not be true, however, for the height + width rule, where we shall assume identical linear functions in order to derive predictions.

The results of the series of four experiments surprised us and led to theoretical revisions. Each experiment was designed to test alternative explanations of the previous one.

EXPERIMENT 1

The first experiment contained (1) a replication of Wilkening's (1979, Experiment 2) study using area ratings, and (2) a study using paired comparisons and symmetric hypotheses tests. The emphasis of the paired comparison study was on strict comparability with Wilkening's experiment. Therefore, we asked the same children and adults for both ratings and paired comparison, and we used the same rectangles in order to avoid possible effects of the absolute size of rectangles.

Method

Subjects. Three groups were investigated: 5–6-year-olds ($n = 20$), 7–9-year-olds ($n = 23$), and adults ($n = 20$). The two groups of children were taken from a Montessori kindergarten and a Montessori school in Munich. The adults

were psychology students at the University of Munich. Sex was balanced in all groups. Because of illness, three 5–6-year-olds could not participate in the rating task, and three 7–9-year-olds could not participate in the paired comparison task.

Rating Scale Task

As mentioned above, materials and procedure used in the rating task exactly replicated those used by Wilkening (1979, Experiment 2).

Materials. Sixteen rectangles were generated from all possible combinations of four heights and four widths, of 4, 8, 12, and 16 cm. Rectangles were cut from wood 8 mm thick and wrapped in silver paper, so that they looked like chocolate bars. A real 4×4 cm chocolate bar, consisting of four 2×2 cm chocolate pieces, was used as an anchor. A board, 150 cm long, 6 cm wide, and 2 cm thick, was used as a rating scale. The board lay between the child and the experimenter and had on its reverse side, invisible to the child but in view of the experimenter, a centimeter scale.

Procedure. Each child was tested separately. First, the experimenter gave the child the real chocolate bar and asked him or her to unwrap it, break it into the four pieces and form a row of them, starting on the left side of the scale. Then the child was asked to indicate on the board where the row ended. Having completed this introduction, the child was allowed to eat the anchor stimulus. Thereafter, the child was shown the 16 “chocolate bars,” one after the other, and was told that he or she should imagine that each bar was broken into pieces of the same size he or she had seen before, and arranged in a row. Then the child was asked to point to a place on the board where the row would end. The child’s judgment was noted in centimeters by the experimenter. The 16 rectangles were presented in random order. The presentation was repeated in a different order, yielding a total of 32 judgments per child. The experimental procedure was the same for adults.

Paired Comparison Task

Materials. From the rectangles used by Wilkening (1979, Experiment 2), all pairs were generated that gave different predictions for the height + width rule and the multiplying rule. These 10 pairs are shown in Table 3. All rectangles were wrapped in silver paper, so that they looked like chocolate bars.

Procedure. Each child was tested separately. The child sat at a table and the experimenter placed one pair of chocolate bars on the table. The child was asked in which of the two bars there was more chocolate. The child could respond “equal,” but was not explicitly instructed to do so. All 10 pairs were presented in random order and the pairs were repeated six times in different orders, yielding a

total of 60 judgments for each child. Adults were tested under identical conditions. The order of rating scale task and paired comparison was varied among subjects in all groups.

Results: Area Ratings

Group Analysis. Figure 2 shows the mean ratings of areas in the three age groups. The group means are almost identical with those reported by Wilkening (1979); they show roughly parallel curves in 5–6-year-olds, and a diverging fan of curves in adults, from which the height + width rule in young children and the multiplying rule in adults were inferred (Anderson & Cuneo, 1978a). Means in 7–9-year-olds show the well-known in-between pattern. Main effects for height and width were $F(3,48) = 42.2$ and 41.8 in 5–6-year-olds, $F(3,66) = 51.5$ and 47.6 in 7–9-year-olds, and $F(3,57) = 239.0$ and 243.9 in adults. The bilinear component of interaction (Graesser & Anderson, 1974) was $F(1,22) = 8.05$ and $F(1,19) = 132.3$ in 7–9-year-olds and adults, respectively. The bilinear component was not significant ($p > .05$) in 5–6-year-olds, and the residual component was not significant in any of the three groups.

Single-Case Analysis. Inferences from group means to the processing of individual minds can be highly misleading, for instance, if individual differences exist within age groups. Individual analyses of variance (following Anderson, 1982, pp. 75–78) revealed that rating patterns strongly varied within age groups, and that many patterns in young children were unsystematic (Table 2). Inspection

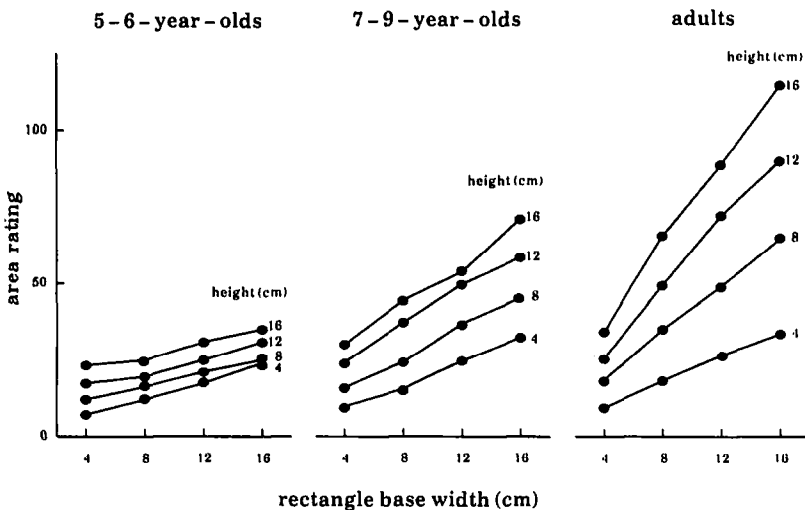


Figure 2. Group mean ratings of area. Each point corresponds to one rectangle. Each unit on the ordinate (rating scale) corresponds to 2 square cm (since pieces of chocolate 2 cm wide were used as anchors; see text).

Table 2. Single-Case Functional Measurement Analysis

Age Group	<i>n</i>	Significant main and interaction effects ^a				
		Height, Width, and Bilinear Interaction ^b	Height and Width	Height only	Width only	Others
5-6-year-olds	17	0	7	1	1	8
7-9-year-olds	23	4	11	3	2	3
Adults	20	12	3	0	0	5

^a*p* < .05^bDiverging curves only (see Bredenkamp, 1984).

of individual graphs confirmed this result. Only 7 out of 17 children in the 5-6-year-olds showed results in the single-case analysis of variance that corresponded to the group analysis. Mean intraindividual error variances in area ratings were strongly age-correlated, as shown above in Table 1. The average percentage of crossovers (from all possible crossovers in a 4 × 4 design) was 3% in adults, 15% in 7-9-year-olds, and 25% in 5-6-year-olds.

Results: Paired Comparisons

Paired comparisons showed no age-dependent loss in reliability. Almost half of the children and 3 adults produced consistent judgments in all 6 replications.

Each of the five hypotheses defined in Equations 8-12 predicts which rectangle in a pair is judged to contain more chocolate, or that both are judged equal (Table 3). Take for instance the (12 × 12, 16 × 8) pair. The height-only rule (*H*)

Table 3. Predicted Responses for Experiment 1

Left Stimulus	Rectangle Pairs									
	12×12	12×8	8×12	8×8	8×16	4×16	8×8	4×16	4×16	4×12
Right Stimulus	16×8	16×4	16×4	16×4	12×12	12×8	16×4	8×12	8×8	8×4
Hypothesis										
H	R	R	R	R	R	R	R	R	R	R
W	L	L	L	L	L	L	L	L	L	L
H+W	Eq	Eq	Eq	R	Eq	Eq	Eq	Eq	L	Eq
L	R	R	R	R	L	L	R	L	L	L
AC	L	L	L	Eq	R	R	L	R	Eq	R
LC+	L	L	L	L	R	R	L	R	R	R
LC-	L	L	L	R	R	R	L	R	L	R

Note. *R* denotes right-side rectangle judged larger, *L* denotes left-side rectangle judged larger, and *Eq* denotes the two rectangles judged equal size. For convenience, predictions of the two limited area constancy hypotheses (*LC*⁺, *LC*⁻) are included; these were suggested post hoc by the judgmental patterns in Experiment 1 (see text).

Abbreviations for hypotheses are: height-only rule (*H*), width-only rule (*W*), height + width rule (*H+W*), longer-side rule (*L*), area constancy (*AC*), limited area constancy (*LC*⁺ and *LC*⁻). The rectangles' sizes are specified as height × width.

predicts that the right rectangle will be chosen, since it has the greatest height. The width-only rule (W) predicts the left rectangle, the height and width rule ($H + W$) predicts that the two rectangles are judged equal size, and so on. Each hypothesis gives a different pattern of predictions. Since each person judged the 10 pairs in 6 repetitions, each hypothesis can give between 0 and 60 correct predictions for each subject.³

Symmetric Testing. The problem of asymmetric hypothesis testing discussed above is that one of two hypotheses ($H + W$) was identified with the statistical null hypothesis, whereas the other was not. This resulted in a small probability of erroneously rejecting $H + W$ (alpha) and a very large probability of erroneously rejecting competing hypotheses (beta), as we have seen.

One way to avoid this problem is to reformulate *each* hypothesis as a null hypothesis and test each hypothesis using the *same* alpha level. We did so. To reformulate the deterministic hypotheses above as statistical ones, a probability (of false predictions) larger than zero had to be selected (zero corresponds to the deterministic hypothesis). We chose a probability of .10 for each hypothesis. Thus we tested the five null hypotheses that judgments follow area constancy (height + width rule, longer-side rule, etc.) with a probability of .90. Under each of these null hypotheses the number of false predictions should follow a binomial distribution with a mean of 6 and lie below 12, with probability .99. If it reached or exceeded this boundary, the null hypothesis was rejected at the .01 level of significance (one-tail test). This symmetric hypothesis testing procedure guarantees the *same* alpha level (probability of erroneous rejection) for each hypothesis tested.

Note that if a null hypothesis was rejected, this result implied that all "stronger" hypotheses with $p < .10$ were also rejected (including the deterministic model). All tests were performed at the individual level.

Results. In Table 4, a subject is classified under a particular hypothesis if his or her judgmental pattern deviates significantly from the patterns predicted by *all* hypotheses except for the hypothesis under which it is classified. None of the five hypotheses could predict the area judgments of the children. Overall, three judgment patterns, intra- and interindividually highly consistent, emerged; they define three subgroups whose group size changes over age. The longer-side rule

³ One problem in counting false predictions arose when a hypothesis predicts "equal," but a child used "greater" or "smaller" judgments only. To account for such a response style, we treated this situation as a forced choice task, where $p(\text{"greater"}) = .5$ is predicted. Each pair of "greater" and "smaller" judgments within the 6 repetitions was transformed into 2 "equal judgments." For instance, if a hypothesis predicted "equal," either 6 "equal" judgments or 3 "greater" and 3 "smaller" judgments or 2 "equal," 2 "greater," and 2 "smaller" judgments were counted as 6 correct judgments. Six "greater" judgments, however, were counted as 6 false predictions, and intermediate values were treated accordingly.

Table 4. Test of Predictions for the Stimulus Set in Experiment 1

Age	<i>n</i>	LC ⁺	LC ⁻	AC	H+W	L	H	W	Unclassifiable
5-6-year-olds	20	10	0	1	0	3	0	1	5
7-9-year-olds	20	7	1	8	0	1	0	0	3
Adults	20	0	4	15	0	0	0	0	1

Note. A subject is classified under a hypotheses (rule) if his or her judgments did not deviate significantly from the predictions of that hypothesis *and* did deviate significantly from the predictions of *all* other hypotheses (all $ps < .01$).

Abbreviations for hypotheses are: Limited area constancy (LC⁺ and LC⁻), area constancy (AC), height + width rule (H+W), longer-side rule (L), height-only rule (H), and width-only rule (W).

predicted the judgments of a small group of children perfectly (i.e., 60 correct predictions); these were all among the youngest in their groups. Area constancy predicted the judgments of a second group of subjects, whose group size increased with age, from 1 to 8 to 15. The most frequent pattern in children, however, could be predicted by none of the hypotheses. This pattern was shown by 17 children but no adults, and was highly intra- and interindividually consistent: In a total of $17 \times 60 = 1020$ judgments, there were only 7 deviations from the general pattern.

A post hoc analysis revealed that this pattern followed area constancy, except for 12 violations that all occurred in the same two pairs, (8×8 ; 16×4) and (4×16 ; 8×8). In these pairs, and in only these pairs, physical area was equal. Without exception, all of these 17 children judged the square to be larger. Surprised by this consistent overestimation of the square, we also reanalyzed the judgment patterns of the 5 adults whose judgments deviated significantly from area constancy. In four cases, the deviations were also due to these two critical pairs, but these adults judged the square to be smaller. Following this post hoc analysis, the symbol *LC* in Tables 3 and 4 denotes limited area constancy when physical areas are equal (i.e., in critical pairs), and *LC⁺* and *LC⁻* denote overestimation and underestimation, respectively, of areas with square shapes (as opposed to rectangular shapes).

The largest number of false predictions in both groups of children (79% and 84%, respectively) stemmed from the height + width rule (assuming identical linear psychophysical functions); here there was not a single child whose judgments did not differ significantly from the rule's predictions.

Discussion

The replication of Wilkening's study, with twice as many subjects in each of the three age groups tested, showed almost identical rating patterns on the group average level. As in Wilkening's study, mean ratings of the youngest group cover only a small range of the response scale and exhibit roughly parallel curves. On the individual level, however, patterns were highly variable and idiosyncratic in young children, and their reliability was low. For the same children and adults,

paired comparisons revealed clear-cut individual differences within age groups and no sign of an age-dependent loss of reliability (see also Lohaus & Trautner, 1986). It may be objected that high intraindividual reliability could simply be the consequence of an insensitivity of the pairs; that is, they could be too "easy" in the sense that every person judges the same rectangle to be the larger one. This is clearly not the case, since judgments systematically differ among individuals: Already in 5-year-olds, four significantly different judgmental patterns exist, predicted by the hypotheses *AC*, *L*, and *W*, and the post hoc pattern *LC*⁺. The frequency of these patterns increases or decreases systematically with age groups, which indicates their developmental significance.

Our post hoc analysis of paired comparisons found the first trace of an influence of shape on perceived area in the following two deviations from area constancy (*S* denotes shape defined as shorter side/longer side):

$$a_i > a_j \text{ if } A_i = A_j \text{ and } S_i > S_j \quad (LC^+) \quad (13)$$

$$a_i > a_j \text{ if } A_i = A_j \text{ and } S_i < S_j \quad (LC^-) \quad (14)$$

We might defend the height + width rule by arguing that children gave different weights to perceived height and width, or that different psychophysical functions exist for height and width. Since different weights and psychophysical functions cannot be empirically separated in this context (Schönemann, Cafferty, & Rotton, 1973), we shall examine both objections as the question whether different psychophysical functions should be assumed.

Data from paired comparisons provide a means for checking this objection. Assume for a moment that there are different functions *f* for height and width. Then a 90° rotation of a nonsquare rectangle should change perceived area, whereas a rotation of a square should not. The set of rectangles contained, for each pair, a corresponding pair where rectangles had been rotated 90°, such as (12 × 12, 16 × 8) and (8 × 16, 12 × 12). Analysis of individual area judgments in children showed, however, that for all pairs, judgments were independent of a 90° rotation (except for one child who centered on width). Thus, for the present material, there is no evidence of different psychophysical functions.

EXPERIMENT 2

We had three objections to the conclusions derived from paired comparisons in Experiment 1. The first objection was that the *LC*⁺ and *LC*⁻ patterns were determined ex post facto. Therefore, we replicated the experiment. Second, the systematic deviations were detected in two critical pairs only. We added two other pairs that were sensitive for an overestimation of square shapes in younger children and for an underestimation in adults. Third, the bad fit of the height +

width rule in paired comparisons could be due to children's misunderstanding the instructions as a forced-choice task—not allowing judgments of equality. This could have an effect on the number of correct predictions of the height + width rule, since 8 out of the 10 pairs constructed from the rectangles used by Wilkening (1979, Exp. 2) gave equality predictions if the height + width rule were true (as opposed to 2 if area constancy—or multiplying *plus* identical linear functions f —were true, and 0 if the centering rules were true). Although we took into account such response styles in the evaluation (Note 3), there might still be some effect due to the instructions. We now explicitly instructed half of the subjects in all age groups that areas may be equal, whereas the other half received the same instructions as in Experiment 1.

Method

Subjects. Sixty new subjects were investigated; age groups were the same as in Experiment 1 ($n = 20$ in each group). The children and adults were sampled from the same institutions as in Experiment 1.

Materials and Procedure. We constructed two additional rectangle pairs that differentiated between the AC and the LC^+ hypotheses in children and the AC and LC^- hypotheses in adults. Pair Number 11 was 10×6.4 , 4×16 , which is a modification of the critical pair (8×8 , 4×16) in the original set. If children overestimated more squarelike shapes, and not only perfect squares, then they should judge the 10×6.4 rectangle to be the larger one. If they judged physical areas correctly, independent of shape, they should judge both equal; if they followed the height + width rule, they should select the 4×16 rectangle. Pair Number 12 was 16×8 , 10.7×10.7 , which has the same ratio of areas (.89) as the 16×8 , 12×12 pair in the original set, but now the square is physically smaller. If overestimation of square shapes is not restricted to physically equal rectangles, but has a stronger effect, then younger children should judge these two “chocolate bars” to be equal or even judge the square to be the larger.

If the adults underestimate the area of the more squarelike shape, then they should judge the 4×16 rectangle in pair Number 11 to be the larger, as in the old pairs (8×8 , 4×16 , and 16×4 , 8×8). The second new pair presented to the adults was different from the children's pair Number 12. Whereas for children the original pair (16×8 , 12×12) was modified into a pair where the more elongated rectangle had a larger physical area, adults received pair Number 12', 16×8 , 10×14.4 , where the more elongated rectangle had the smaller area. In all three pairs (the original and Numbers 12 and 12'), the ratio of areas was kept constant (.89). If adults underestimated the area of the more squarelike shape, then they should judge the 16×8 rectangle to be the larger one.

The procedure was identical with that in Experiment 1, except that, as mentioned above, half of the people in each group were asked “Which bar contains

more chocolate?”, whereas the others were also explicitly instructed that bars may be equal, and that in these cases the answer “equal” should be given. Each subject judged 12 pairs of rectangles in 6 repetitions, so the number of judgments was 72.

Results

Analysis of paired comparison followed the same procedure as in Experiment 1. Again, there was no age-dependent loss of reliability, and the pairs were sensitive enough to produce clear-cut individual differences within and between age groups. Explicit instruction for equality judgments increased the overall number of such judgments in all age groups, but this increase had no effect on the reliability of judgments or on the results of hypothesis testing. Therefore, both instruction conditions are aggregated in Table 5.

Results could be replicated when we used more critical pairs. We found the LC^+ pattern again in 16 children and in none of the adults. Centering rules and area constancy predicted the other two main patterns, which showed the same age-dependent change in frequency as in Experiment 1. The height + width rule again gave the largest number of false predictions in both groups of children, and we did not find a single subject whose judgments did not depart from it significantly.

In contrast to Experiment 1, more children centered on height than on the longer of the two sides. Two of these children showed an impressive performance when, after Experiment 2, we gave them the rectangles again and explicitly told them about the difference between height and area. But they continued to judge “amount of chocolate” by height only. When we rotated these rectangle pairs (90°) in front of the children, they changed their judgment and pointed to the other rectangle, which now had the greater height. These children did not seem to grasp the concept of invariance of area with respect to rotation.

We now made another check of whether the bad fit of the height + width rule (assuming the same linear psychophysical functions) was due to the larger number of equality predictions. We considered only those pairs of rectangles where dominance predictions (as opposed to equality predictions) were made, and reanalyzed Experiments 1 and 2. When equality predictions were not evaluated, the mean number of false predictions by the height + width rule was slightly reduced, but still large: 64.1% in 5–6-year-olds and 65.4% in 7–9-year-olds.

Table 5. Test of Predictions for the Stimulus Set in Experiment 2

Age	<i>n</i>	LC+	LC-	AC	H+W	L	H	W	Unclassifiable
5–6-year-olds	20	7	1	0	0	1	4	0	7
7–9-year-olds	20	9	2	3	0	1	1	0	4
Adults	20	0	6	10	0	0	0	0	4

Discussion

How can we explain the direction of the illusion? Why did young children overestimate square shapes instead of underestimating them? Peters (1933), who reported the same effect, did not explain why children judged a square to be the larger, but offered an explanation for why adults showed the opposite illusion. Older children and adults anticipate the effect of shape on area and try to avoid it, resulting in some cases in an overcompensation, that is, in an underestimation of the square. Brunswik (1934a, 1934b) gave a different account of the perceptual interaction between area and shape in his treatment of what he called the "generalized problem of constancy." According to Brunswik, the perceptual system, if one is instructed to judge area, generates "in-between objects" (*Zwischengegenstände*), that is, a compromise between area and shape. But neither Brunswik's "in-between objects" nor his "cue learning" principles (on their relationship, see Gigerenzer & Murray, 1987, pp. 70–74) can explain the direction of the illusion. The same holds for the more recent studies of Felfoldy (1974), Shepp (1978), and Shepp, Burns, and McDonough (1980).

We tried earlier to explain this puzzling fact by a "good Gestalt illusion" in young children (Gigerenzer, 1984; Richter & Gigerenzer, 1984). This hypothesis assumes a tendency to overestimate simple and symmetric geometric figures. For instance, it predicts that children will also overestimate the area of a circle compared with an ellipse of equal size. Our thesis was that the limited perceptual constancy stated in Equation 13 was due to a good Gestalt illusion, whereas the limited perceptual constancy stated in Equation 14 was due to adults' anticipation of the perceptual illusion and overcompensation for it. Thus we assumed that Equation 13 reflected a principle of perceptual development, whereas Equation 14 described the interference of a cognitive process. In the third experiment we tested the hypothesis of a good Gestalt illusion using a new and larger set of rectangle pairs to control for effects due to the specific sets of rectangles used before.

EXPERIMENT 3

Our conclusions from Experiments 1 and 2 were that (1) area constancy increases with age, and (2) limited area constancy is due to a good Gestalt illusion in children and an overcompensation in adults. We now considered the objection that these results were dependent on the specific set of rectangle pairs derived from Wilkening's experiment. In particular, the squarelike shapes were never physically smaller than the more elongated shapes in the 10 pairs of Experiment 1, and in all but one pair (Number 12) in Experiment 2. Children with rather well-developed area constancy (as opposed to centering strategies) might learn during the experiment that the square is in most cases the larger one and use the square shape as a cue for larger area in critical pairs. In order to test this objection, we constructed a new and larger set of 30 rectangle pairs, in which the

number of physically greater areas was balanced between more elongated and squarelike shapes. This set also contained 10 pairs with physically equal areas but differing shapes, and thus 10 more critical pairs for testing the interaction between perceived area and shape. Since this set contained not a single pair for which the height + width rule predicted equality, it also provided a strong test for the objection that the failure of the height + width rule could be due to the larger number of equality predictions in Experiments 1 and 2.

Method

Subjects. Two groups participated in the experiment, 50 children aged 4–6 and 40 adults aged 18–40. The children were from a kindergarten in Munich; the adults were psychology students.

Materials. Rectangles were generated in a 6×13 (shape \times area) matrix, with shapes of $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, and $\frac{1}{1}$, and area ranging from 48.06 cm² to 136.55 cm² in steps of constant relative differences (8.9%). Thirty pairs were selected with the restriction that each pair (where physical area was not equal) had a “matched pair,” and that physical differences within a pair were not unreasonably large. “Matched pairs” were defined in the following way, using the symbol A_s for the area of the rectangle with the more squarelike shape in a pair, and A_r for that of the less squarelike. Two pairs are called “matched” if (1) the rectangles in the first pair have the same shapes as those in the second pair; and (2) the ratio A_s/A_r in the first pair is equal to the ratio A_r/A_s in the second pair. Thus in each matched pair there is one pair where the more squarelike rectangle is physically larger, and one pair where it is smaller. An example is 7.5×9.9 , 5.8×11.7 and 8.1×10.9 , 6.9×13.9 , where the shapes in each pair are $\frac{3}{4}$, $\frac{1}{2}$, and A_s/A_r is $74.25/68.06 = 1.09$ in the first pair, and A_r/A_s is $96.42/88.30 = 1.09$ in the second pair (shapes are exact, measures for height and width are rounded). The 12 pairs used in Experiment 2 were added, and thus the entire set consisted of 42 pairs.

Procedure. For the children, the procedure was the same as in Experiment 2, except that each rectangle pair was judged only once. For adults, rectangle pairs were projected onto a screen, and the subjects were given a questionnaire with response alternatives for each pair. This procedure allowed us to test adults in a group session and was introduced for purely economic reasons. Pretests using both rectangles wrapped in silver paper (as in Experiments 1 and 2) and slides showed that the results were quite similar in adults, but not in children. Therefore, only adults were tested in a group session using slides, but children were tested individually, using rectangles wrapped in silver paper and the procedure used in Experiments 1 and 2.

Table 6. Test of Predictions for the Stimulus Set in Experiment 3

Age	<i>n</i>	LC ⁺	LC ⁻	AC	H+W	L	H	W	Unclassifiable
4-6-year-olds	50	13	7	6	0	13	0	0	11
Adults	40	1	1	35	0	0	0	0	3

Results

Hypothesis testing and classification followed the procedures of Experiments 1 and 2. Table 6 shows two main results. First, in contradiction to the assumed good Gestalt illusion, children showed both overestimation (LC^+) and underestimation (LC^-) of square shapes, whereas in adults both effects disappeared except for two cases. Second, although the area constancy hypothesis now predicts 13 equality judgments (3 for the set of pairs used in Experiment 2, and 10 for the new set) as opposed to only 8 equality predictions by the height + width rule, there was still not a single child who followed the predictions of the height + width rule. We now have direct evidence that the failure of the latter cannot be attributed to the issue of equality predictions.

Discussion

Results support the perceptual constancy view: There is limited perceptual constancy in children and perceptual constancy in adults. But results contradict our previous explanation, the hypothesis of an age-dependent switch in the influence of shape on perceived area. The hypothesis of a good Gestalt illusion in children and an overcompensation in adults cannot explain the result in Table 6. It seems that the positive correlation between area and squarelike shapes in Experiments 1 and 2 influenced children to use shape as a cue for area, and to overestimate square shapes. In the third experiment, where a new set of pairs with uncorrelated areas and shapes was introduced, limited perceptual constancy in children is divided into LC^+ and LC^- patterns. The greater number of overestimations of squares can be attributed to the fact that we have added the 12 original pairs to the 30 uncorrelated ones, resulting in a slightly positive correlation between area and shape in the whole set used.

It seems that we now have strong evidence for the existence of limited area constancy in children, but that the direction of the dependence between area and shape calls for an explanation different from a good Gestalt illusion. Results of Experiments 1-3, seen together, even contradict Equation 1 as a general framework of limited perceptual constancy. The dependency between perceived area and shape seems to be due to the actual correlation between shape and area in the set of rectangles presented, rather than to the shape S_i of the particular rectangle judged. Thus the dependency, but not the direction of the dependency, seems to be a characteristic of age.

If this were true, we should be able to reverse children's stable overestimation

of squares in Experiments 1 and 2 into an equally stable underestimation if we reverse the correlation between shape and area in the experimental set.

EXPERIMENT 4

We next considered the judgments of the two critical pairs (equal physical area) in Experiment 1 as the dependent variable and the correlation between area and shape in the remaining 8 pairs as the independent variable. We replaced the original 8 pairs by 8 "matched" pairs, which had the same shapes and the same ratio of areas, but the more elongated rectangle was now always the larger one. Thus the correlation between squarelike shapes and area was now negative instead of positive, as in Experiment 1. The critical pairs were the same as in Experiment 1. If the mechanism of limited perceptual constancy is that children learn about the shape-area correlation during the experiment and use it as a cue for area, then we should now find that the consistent overestimation of square shapes disappears and turns into an equally consistent overestimation of elongated rectangles.

Method

Subjects. Twenty children aged 4–6 from a Munich kindergarten served as subjects.

Materials. The same materials as in Experiment 1 were used, except that the 8 pairs with unequal physical area were replaced by "matched" pairs. We kept the size of the more squarelike rectangle constant and enlarged the area of the second rectangle, leaving all shapes and the ratio of areas constant.

Procedure. Same as in Experiment 1, except that only 4 replications were made.

Table 7. Dependency of Children's Area Judgments on the Correlation Between Area and Shape in the Experimental Set (4–6-year-olds)

	Rectangle Set	
	Exp. 1	Exp. 4
Overestimation of squares (LC^+)	10	0
Underestimation of squares (LC^-)	0	12

Results and Discussion

As in experiment 1, children gave very reliable judgments. A child was classified as LC^+ if he or she always judged the square in the two critical pairs the larger, or if he or she made no more than one deviation ("equal" judgments were counted as deviations). The same procedure held for classifying children as LC^- . Table 7 shows that overestimation of squares disappeared completely, and that children now *underestimated* squares. Inter- and intraindividual consistency of this response pattern was almost identical to the opposite pattern in Experiment 1; 8 of the 12 classified children showed not a single deviation. This result demonstrates that the limited perceptual constancy found in children derives its direction from the particular shape-area correlation in the experimental set rather than from some innate, age-dependent, perceptual principle.

GENERAL DISCUSSION

A Contextual Theory of Limited Area Constancy

As mentioned in the introduction, there exists a long tradition from Bühler (1913) and Rubin (1921) to Rausch (1964) and Shepp (1978) that emphasizes the dependency of perceived area on shape. The results reported here, however, reveal that there are two ways to model this limited area constancy, corresponding to two distinct meanings of "shape" as context. The first view, which we followed earlier when we postulated a "good Gestalt illusion" in young children (Gigerenzer, 1984; Richter & Gigerenzer, 1984), attempted to understand the dependency on the conceptual basis of Equation 1. This view, like the cognitive algebra view, focuses on the properties of a single rectangle i :

$$a_i = f(A_i, S_i) \quad (15)$$

Present results, however, suggest that it is not the shape of a particular rectangle that influences perceived area, but the covariation of shape and area in the set of rectangles. Let us denote this covariation with C_{set} ; we then get the following revision of Equation 1:

$$a_i = f(A_i, C_{\text{set}}) \quad (16)$$

For paired comparisons, C_{set} can be defined in a simple way as the relative frequency of pairs in which the more squarelike shape has the larger area:

$$C_{\text{set}} = n_s / (n_s + n_r), \quad (17)$$

where n_s and n_r are the number of pairs in which the square or the rectangle, respectively, has the larger area. In Experiments 1-4 the values for C_{set} were 1.00, .89, .62, and .00, respectively. One important result of Experiment 3 was

that if the covariation between area and shape in the rectangle set is decreased from 1.00 (Experiment 1) and .89 (Experiment 2) to .62, this change did not result in a corresponding decrease of limited perceptual constancy and an increase of perceptual constancy (*AC* patterns). Rather, the percentage of *LC*⁺ and *LC*⁻ in children remained constant throughout all four experiments, ranging between 40% and 60%. This result suggests that (1) these children use the covariation cue in all cases, and (2) they infer the covariation from the area-shape relationships in the first few pairs. Since we presented the pairs in random order in all experiments, which varied from person to person, some children may have inferred from the first pairs that greater area covaries with a more elongated shape, whereas others may have experienced more squarelike shapes with greater area.

Thus the value of C_{set} should predict the proportions of children overestimating and underestimating square shapes. Table 8 shows that C_{set} is an excellent predictor of the relative number of children overestimating square shapes. The proportion of *LC*⁺ patterns among all limited constancy patterns (*LC*⁺ and *LC*⁻) is almost perfectly related to the C_{set} value of the experiment:

$$C_{\text{set}} = n^+ / (n^+ + n^-) \quad (18)$$

where n^+ and n^- are the numbers of *LC*⁺ and *LC*⁻ patterns, respectively. Equation 18 proposes a quantitative version of the covariation principle in Equation 16.

The conceptual change from Equation 1 to Equation 16 relates our understanding of limited area constancy to range-frequency theory (Parducci, 1974). As in range-frequency theory, judgments depend on the distribution of a variable in the whole experimental set of stimuli. But in range-frequency theory this variable is the variable judged, whereas in the present analysis it is the covariation between the variable judged (area) and a second variable of the same stimulus (shape). Moreover, range-frequency theory deals with contextual effects in category rating, but the covariation principle formulated in Equations 16-18 deals with context effects in ordinal paired comparisons. Both views are alike in that contextual effects are seen as an integral part of perceptual judgment and of

Table 8. Covariation Between Shape and Area in the Four Experiments and Relative Frequency of 4-6-year-olds Overestimating Squares

	n	C_{set}	$\frac{n^+}{n^+ + n^-}$
Experiment 1	20	1.00	1.00
Experiment 2	20	.89	.89
Experiment 3	50	.62	.65
Experiment 4	20	.00	.00

its development, rather than as a nuisance (Stevens, 1971) or something that should be avoided (Poulton, 1979). Contextual effects have often been dismissed as laboratory artifacts that merely distort the “true” subjective values or psychophysical functions. On the contrary, we conjecture that in many situations, subjective values and judgments are inferred from contextual information, a view proposed more than 50 years ago in Brunswik’s (1934a) “multidimensional psychophysics.” Range-frequency theory and our covariation hypothesis might be combined in some ways; for example, the frequency principle might be generalized to a two-dimensional frequency distribution that specifies the covariation between two variables.

For instance, the question whether the covariation principle also applies to children’s ratings of area could be answered by this generalization. Imagine two (or more) sets of rectangles with different covariations between area and shape. For instance, in one set rectangular shapes tend to be larger, but in the other set square shapes are. A “critical pair” would be any two rectangles with equal area (predicted by some hypothesis under consideration), but varying shape. If the covariation principle applies to ratings, children should rate either the rectangular shape or the square shape to be larger, whichever covaries with larger area. The degree of covariation in the set of rectangles should predict the proportion of children that give higher ratings to square shapes, in analogy to Equations 16-18.

Perceptual Constancy and Cognitive Algebra: Towards a Rapprochement

The perceptual constancy view and the cognitive algebra view are not irreconcilable. The problems that lead to contradictory predictions are the currently prevalent formulations. For instance, there is nothing in the cognitive algebra framework that precludes substituting area and shape for height and width in Figure 1, which would transform the cognitive algebra view into a perspective that is fully compatible with Equation 1, although not with Equation 16. Moreover, Anderson (1981, 1982) himself has done much work on the dependency of perception on “irrelevant” cues, and it is rather atypical that his developmental account of area perception does not deal with context effects. Thus, there is some potential for a rapprochement.

In any case, the popular developmental thesis that young children use a height + width rule (Anderson & Cuneo, 1978a; Cuneo, 1980; Wilkening, 1979) is not supported by ordinal paired comparisons. The evidence against this view is:

1. All 130 children aged 4-9 in Experiments 1-3 (Experiment 4 was not designed to test for the height + width rule) showed judgments that deviated significantly from the predictions of the height + width rule (based on identical linear functions f).
2. Among the cognitive algebra hypotheses considered, physical area provided by far the best overall predictions in all age groups. In the cognitive algebra

framework, this is equivalent to saying that the multiplying rule *plus* identical linear functions f gave the best overall predictions, even in 5–6-year-olds.

3. Area was judged to be dependent on shape by children (and children's comments show that they realize this dependency).⁴ Furthermore, some of the younger children consistently center.

With respect to the height + width rule, however, we are in no better position than in the cognitive algebra research cited above: In both cases conclusions are conditional on the assumption of linear functions for either psychophysical or judgment functions. What we reject is the joint assumption of a height + width rule and identical linear psychophysical functions for height and width. Note that in the range of lengths used in the present study, say between 0 and 30 centimeters, a linear function is not an unjustifiable assumption. This would imply the rejection of the height + width rule itself. Empirical evidence exists (e.g., Krantz, 1972; Reese, 1946–1952; Stevens & Galanter, 1957; Teghtsoonian, 1965) for what Stevens and Galanter called “the rather obvious fact that apparent length is very nearly a linear function of physical length” (1957, pp. 378–379). We would follow Stevens and Galanter on this issue, but we have, as mentioned above, no independent evidence.

Finally, cognitive algebra research suffers from the straightforward reliance on analysis of variance (ANOVA) on both theory construction *and* testing. Consider a factorial design with height and width. What questions can we answer if we apply ANOVA to the design? We can discover whether the effect of height, width, both, or the interaction is significant. This already gives us *all* the hypotheses that have been proposed for area judgment and its development by Anderson & Cuneo (1978a), Cuneo (1980), and Wilkening (1979, 1980). These are height-only, width-only, height + width, and multiplying. Even Piaget's theory has been put into the Procrustean bed of the ANOVA language, as we have shown above. The impact of ANOVA on shaping cognitive algebra theory and other cognitive theories has been dealt with in detail elsewhere (Gigerenzer & Murray, 1987). Here, it is sufficient to point out the circle in which traditional cognitive algebra research is caught. Biases both in hypotheses construction and testing can combine: The height + width rule suggested by the ANOVA framework is “validated” by ANOVA through confirming the null hypothesis of no interaction.

This limitation can be resolved. Cognitive algebra theory usually states two or more hypotheses with precise predictions, and there is no reason to declare one

⁴ Children's spontaneous comments during their performance are another source of evidence for the dependency of perceived area on shape. Children overestimating square shapes typically made remarks such as “one is longer, the other is thicker and therefore larger,” “the longer one has less, the square has more,” and “the thicker one is larger.” Throughout all experiments, young children mentioned that they take the shape into account.

of the two hypotheses to be the null hypothesis. It should be recalled that Fisher, who developed null hypothesis testing as it is used in ANOVA, called these tests only a weak argument (see Gigerenzer et al., 1989, chapter 3). In the present study we eliminated this asymmetry by treating *every* hypothesis as a null hypothesis (see also Gigerenzer, Hell, & Blank, 1988), but there are other symmetric testing procedures, such as Fisher's maximum likelihood or Neyman and Pearson's hypothesis testing, that could be used. Here we may see another possible rapprochement between the perceptual constancy and cognitive algebra points of view.

A Three-Step Processing Model

We propose that the development change in paired comparison judgments of area is a three-step process, with considerable overlap of age groups in the first two steps:

Step 1: Centering. The most frequent centering strategy is centering on the longer side. It implies that children pay attention to both sides of a rectangle, but seem not yet to have developed the concept of area. Alternatively, centering could also be seen as a simplistic judgmental strategy that is dependent on the task, not only on the concept of area.

Step 2: Limited Perceptual Constancy. Area perception is influenced by the shape of the rectangles, and the direction of this influence is determined by the covariation between area and shape in the set of rectangles (Equations 16-18).

Step 3: Perceptual Constancy. Area perception matches physical area, independent of shape.

From our experiments, considering all 150 4-9-year-olds, we must conclude that Step 3 is rare in young children, and that Step 1 is not found in normal adults. Limited perceptual constancy is the most frequent transition state for children aged 4-9.

REFERENCES

- Anderson, N.H. (1981). *Foundations of information integration theory*. New York: Academic.
- Anderson, N.H. (1982). *Methods of information integration theory*. New York: Academic.
- Anderson, N.H., & Cuneo, D.O. (1978a). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General*, *107*, 335-378.
- Anderson, N.H., & Cuneo, D.O. (1978b). The height + width rule seems solid: Reply to Bogartz. *Journal of Experimental Psychology: General*, *107*, 388-392.
- Birnbaum, M.H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysics* (pp. 401-485). Hillsdale, NJ: Erlbaum.
- Bogartz, R. (1978). Comment on Anderson and Cuneo's "The height + width rule in children's judgments of quantity." *Journal of Experimental Psychology: General*, *107*, 379-387.

- Bredenkamp, J. (1984). Weitere kritische Anmerkungen zur Anwendung der Informationsintegrationstheorie in der Entwicklungspsychologie [Further critical comments on information integration theory in developmental psychology]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 16, 134–145.
- Brunswik, E. (1934a). *Wahrnehmung und Gegenstandswelt* [Perception and the world of objects]. Leipzig: Deuticke.
- Brunswik, E. (1934b). Flächeninhalt und Volumen als Gegenstände der Wahrnehmung [Area and volume as perceptual objects]. In O. Klemm (Ed.), *Bericht über den XIII. Kongress der Deutschen Gesellschaft für Psychologie* (pp. 120–123). Jena: Fischer.
- Bühler, K. (1913). *Die Gestaltwahrnehmungen* [Gestalt perceptions]. Stuttgart: Spemann.
- Butzin, C.A., & Anderson, N.H. (1973). Functional measurement of children's judgments. *Child Development*, 44, 529–537.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic.
- Cuneo, D.O. (1980). A general strategy for quantity judgments: The height + width rule. *Child Development*, 51, 299–301.
- Cuneo, D.O. (1982). Children's judgments of numerical quantity: A new view of early quantification. *Cognitive Psychology*, 14, 13–44.
- Dixon, W.J., & Massey, F.J., Jr. (1969). *Introduction to statistical analysis* (3rd ed.). New York: McGraw-Hill.
- Felfoldy, G.L. (1974). Repetition effects in choice reaction time to multidimensional stimuli. *Perception & Psychophysics*, 15, 453–459.
- Garner, W.R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Gigerenzer, G. (1983a). Über die Anwendung der Informations-Integrations-Theorie auf entwicklungspsychologische Problemstellungen: Eine Kritik [Applying information integration theory to developmental issues: A critique]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 15, 101–120.
- Gigerenzer, G. (1983b). Informationsintegration bei Kindern. Eine Erwiderung auf Wilkening [Children's information integration. A reply to Wilkening]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 15, 216–221.
- Gigerenzer, G. (1984). Lässt sich die Flächenwahrnehmung als "kognitive Algebra" beschreiben? [Can we model area perception by "cognitive algebra"?] *Psychologische Beiträge*, 26, 113–119.
- Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 513–525.
- Gigerenzer, G., & Murray, D.J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L.J., Beatty, J., & Krüger, L. (1989). *The empire of chance. How probability changed science and everyday life*. Cambridge, England: Cambridge University Press.
- Graesser, C.C., & Anderson, N.H. (1974). Cognitive algebra of the equation: Gift size = generosity \times income. *Journal of Experimental Psychology*, 103, 692–699.
- Granit, A.R. (1921). A study on the perception of form. *British Journal of Psychology*, 12, 223–247.
- Hempstead, L. (1900). The perception of visual form. *American Journal of Psychology*, 12, 185–192.
- Krantz, D.H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 9, 168–199.
- Kun, A., Parsons, J.E., & Ruble, D.N. (1974). Development of integration processes using ability and effort information to predict outcome. *Developmental Psychology*, 10, 721–732.
- Leeser, O. (1916). Über Linien- und Flächenvergleiche [On judgments of lines and areas]. *Zeitschrift für Psychologie*, 74, 1–127.

- Lenk, E. (1926). Über die optische Auffassung geometrisch-regelmässiger Gestalten [On the optical representation of regular geometrical shapes]. *Neue Psychologische Studien*, 1, 573–612.
- Leon, M. (1982). Extent, multiplying, and proportionality rules in children's judgments of area. *Journal of Experimental Child Psychology*, 33, 124–141.
- Lohaus, A. (1986). Zum Einstaz von Antwortskalen in der Datenerhebung bei Kindern und Jugendlichen [On the use of response scales in children and adolescents]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 18, 214–224.
- Lohaus, A., & Trautner, H.M. (1986). Informationsintegration bei Kindern: Eine empirische Analyse auf der Basis von Paarvergleichen [Information integration in children: An empirical analysis using paired comparisons]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 18, 97–113.
- Miller, P.M. (1982). Children's and adults' integration of information about noise and interest level in their judgments about learning. *Journal of Experimental Child Psychology*, 33, 536–546.
- Monahan, J.S., & Lockhead, G.R. (1977). Identification of integral stimuli. *Journal of Experimental Psychology: General*, 106, 94–110.
- Parducci, A. (1974). Contextual effects: A range-frequency analysis. In E.C. Carterette & M.P. Friedman (Eds.), *Handbook of perception* (Vol. 2, pp. 127–141). New York: Academic.
- Peters, W. (1933). Versuche über den Einfluss der Form auf die Wahrnehmung der Flächengrösse [Experiments on the impact of shape on the perception of area]. *Zeitschrift für Psychologie*, 129, 323–337.
- Piaget, J. (1969). *The mechanisms of perception*. London: Routledge & Kegan Paul.
- Poulton, E.C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, 86, 777–803.
- Rausch, E. (1949). Variabilität und Konstanz als phänomenologische Kategorien [Variability and constancy as phenomenological categories]. *Psychologische Forschung*, 23, 69–114.
- Rausch, E. (1952). *Struktur und Metrik figural-optischer Wahrnehmung* [Structure and metric of figural-optical perception]. Frankfurt/Main: Kramer.
- Rausch, E. (1964). Einzelgegenständlichkeit als phänomenale Eigenschaft [Singularity as a phenomenological property]. *Psychologische Forschung*, 28, 33–45.
- Reese, E.P. (Ed.). (1946–1952). *Summary report from the psychophysical research unit* (Tech. Rep. No. SDC 131–1–5). South Hadley, MA: Mt. Holyoke College, U.S. Navy Special Devices Center.
- Richter, H.R. (1988). Untersuchungen zur Informationsverarbeitung bei der Wahrnehmung von Flächen [Information integration studies of area perception]. In W. Marx (Ed.), *Verbales Gedächtnis und Informationsverarbeitung* (pp. 110–132). Göttingen, W. Germany: Hogrefe.
- Richter, H.R., & Gigerenzer, G. (1984). Erste Ergebnisse einer Längsschnittstudie zur Entwicklung der Flächenwahrnehmung [A longitudinal study of the development of area perception: First results]. *Psychologische Beiträge*, 26, 654–658.
- Rubin, E. (1921). *Visuell wahrgenommene Figuren* [Visually perceived figures]. Berlin: Universitas.
- Sander, F., & Volkelt, H. (1962). *Ganzheitspsychologie* [Holistic psychology]. Munich: Beck.
- Schönemann, P.H., Cafferty, T., & Rotton, J. (1973). A note on additive functional measurement. *Psychological Review*, 80, 85–87.
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309–316.
- Selinka, R. (1939). Der Übergang von der ganzheitlichen zur analytischen Auffassung im Kindesalter [The transition from holistic to analytic representation in childhood]. *Zeitschrift für Pädagogische Psychologie und Jugendkunde*, 40, 256–278.
- Shepp, B.E. (1978). From perceived similarity to dimensional structure: A new hypothesis about perceptual development. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 135–167). Hillsdale, NJ: Erlbaum.
- Shepp, B.E., Burns, B., & McDonough, D. (1980). The relation of stimulus structure to perceptual

- and cognitive development: Further tests of a separability hypothesis. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 113–145). Hillsdale, NJ: Erlbaum.
- Siegler, R.S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8, 481–520.
- Siegler, R.S. (1978). The origins of scientific reasoning. In R.S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 109–149). Hillsdale, NJ: Erlbaum.
- Silverman, J.W., & Paskewitz, S. L. (1988). Developmental and individual differences in children's area judgment rules. *Journal of Experimental Child Psychology*, 46, 74–87.
- Singh, R., Sidana, U.R., & Saluja, S.K. (1978). Playgroup attractiveness studied with information integration theory. *Journal of Experimental Child Psychology*, 25, 429–436.
- Singh, R., Sidana, U.R., & Srivastava, P. (1978). Averaging processes in children's judgments of happiness. *Journal of Social Psychology*, 104, 123–132.
- Stevens, S.S. (1971). Issues in psychophysical measurement. *Psychological Review*, 78, 426–450.
- Stevens, S.S., & Galanter, E.H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 54, 377–411.
- Surber, C.F. (1984). Issues in using quantitative rating scales in developmental research. *Psychological Bulletin*, 95, 226–246.
- Teghtsoonian, M. (1965). The judgment of size. *American Journal of Psychology*, 78, 392–402.
- Tighe, T.J., & Shepp, B.E. (1983). *Perception, cognition, and development: Interactional analyses*. Hillsdale, NJ: Erlbaum.
- Verge, C.G., & Bogartz, R.S. (1978). A functional measurement analysis of the development of dimensional coordination in children. *Journal of Experimental Child Psychology*, 25, 337–353.
- Weintraub, D.J. (1971). Rectangle discriminability: Perceptual relativity and the law of Prägnanz. *Journal of Experimental Child Psychology*, 88, 1–11.
- Werner, H. (1933). *Einführung in die Entwicklungspsychologie* [Introduction to developmental psychology] (2nd ed.). Leipzig: Barth.
- Wilkening, F. (1979). Combining of stimulus dimensions in children's and adults' judgments of area: An information integration analysis. *Developmental Psychology*, 15, 25–33.
- Wilkening, F. (1980). Development of dimensional integration in children's perceptual judgment: Experiments with area, volume, and velocity. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 47–69). Hillsdale, NJ: Erlbaum.
- Wilkening, F. (1983). Entwicklung der Informationsintegration: Eine Antwort auf Gigerenzer's Kritik [Development of information integration: A reply to Gigerenzer's critique]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 15, 207–215.