Birth Order

Reconciling Conflicting Effects

Robert B. Zajonc and Patricia R. Mullally
Stanford University

Secular trends in test scores are accurately predicted by trends in aggregate birth orders. The trend data contradict individual-difference analyses that show birth order as a poor predictor of individual test scores. This article demonstrates why the 2 formulations of the problem—the individually distributed birth order analysis and aggregate-pattern analysis—generate different results. A meaningful interpretation is given by the confluence model, a theory specifying the process whereby the intellectual environment modulates intellectual development. The authors introduce the concept of collective potentiation that specifies collective side effects of birth order. In contrast to genetic theories, the confluence model quantifies the differential environmental contributions to intellectual development of successive siblings, and it offers several well-confirmed derivations that genetic and other developmental theories cannot explain.

When a variable analyzed in one form explains 90% of variance of a given outcome, and in another equally justifiable form, the variable explains only 1%, we have a serious empirical contradiction. Such a contradiction emerges in the study of the role of birth order in intellectual test scores that this article seeks to resolve. In Figure 1, we present three secular trends for test scores that are characterized by a remarkably close correspondence with changes in the aggregate birth order.[1] In the first graph are quantitative Scholastic Assessment Test (SAT) scores, in the second United Kingdom A-level data (taken there by 17-year-olds and similar to SATs in purpose but not in form), and in the third the combined basic skills scores of Iowa elementary and high school children. Aggregate birth orders are calculated quite simply. The number of first children born in a given cohort year is multiplied by one, the number of second children is multiplied by two, third children by three, and so on, for all children born. The sum of these products is then divided by the total number of births in the given year, and a value of the average birth order is obtained. We see from the graphs that a child born in the United States in 1962, who took the SATs 18 years later in 1980, entered at birth a family that, on average, already had 2.05 children (that child's birth order was 3.05), whereas a child born in 1978 (who took the SATs last year) entered a family with only 1.02 siblings (this child's birth order was 2.02). The U.K. birth order figures also changed during this interval from 2.3 in 1962 (who took the A-levels in 1979) to 1.9 for children born in 1978 (who took their A-levels in 1995). Among others, a high aggregate birth order of a cohort means that the proportion of firstborns in that cohort is also relatively high—a factor that is significant, as we explain below, in understanding test score trends.

The correspondence between the test trends and family trends is remarkable. Regression equations showed that the aggregate birth order was by far the major source of variation in test scores, for it alone accounted for as much as 81% of variance in SATs, 86% in the A-levels, and 89% in the Iowa scores.[2]

Editor's note. Lyle E. Bourne, Jr. served as action editor for this article.

Author's note. Robert B. Zajonc and Patricia R. Mullally, Department of Psychology, Stanford University.

We are grateful to Hazel Markus for her valuable comments and to Kim Colwell for her help in the analysis of the data.

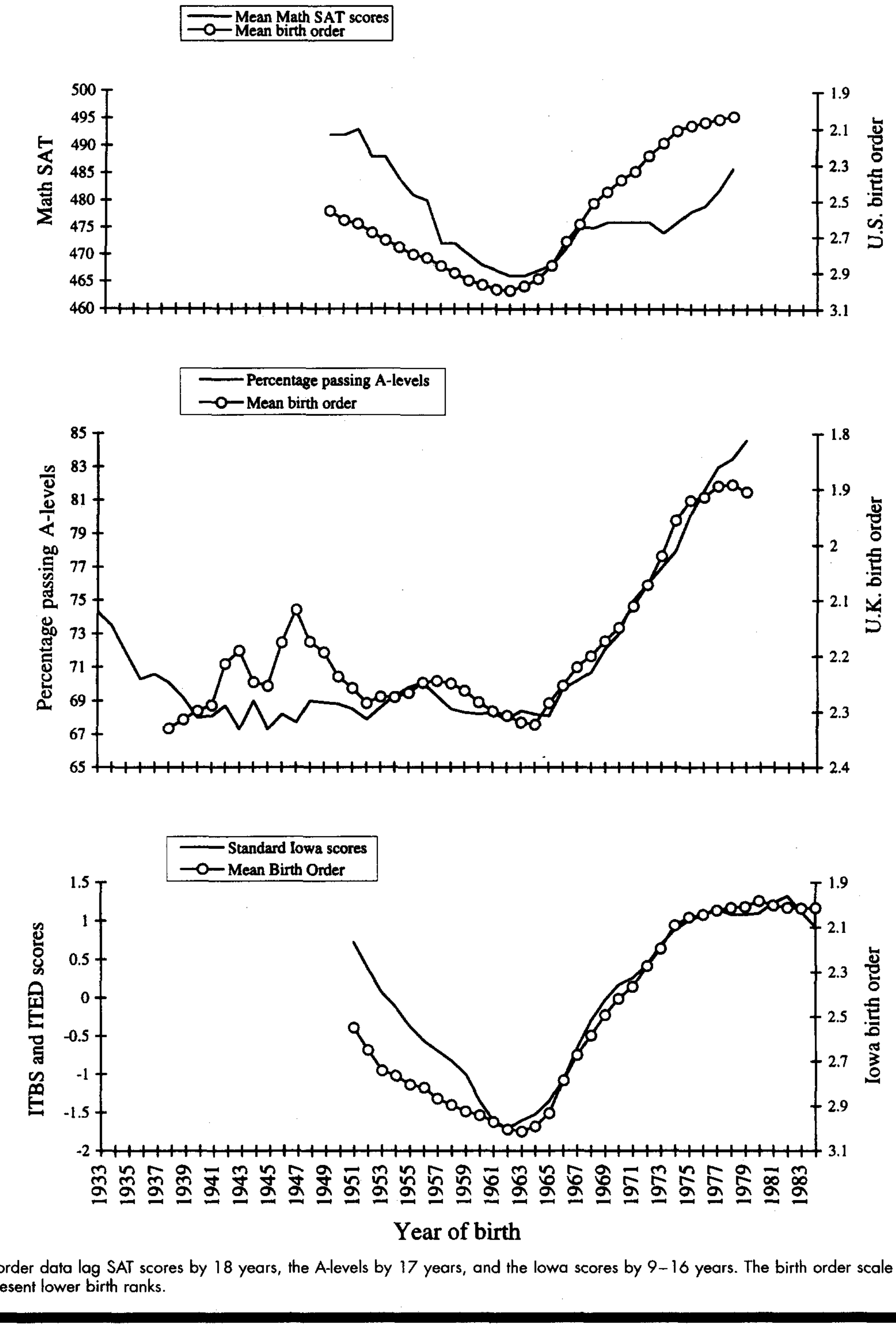
Correspondence concerning this article should be addressed to Robert B. Zajonc, Department of Psychology, Jordan Hall, Stanford University, Stanford, CA 94305. Electronic mail may be sent via Internet to zajonc@psych.stanford.edu.

[1] The SATs come from the official publications of the College Entrance Examination Board and Educational Testing Service. The A-level data are based on a report of the Manchester Centre for Education and Employment Research, kindly supplied by Alan Smithers, School of Education, Brunel University, Twickenham, United Kingdom. The Iowa data were supplied by Jim Gould of the State of Iowa Department of Education and Robert L. Brennan, Director of Iowa Testing Programs, whom we thank for their help. The current data were combined with an earlier data set (Zajonc, 1976), which was collected by W. E. Coffman, the former director of the Iowa Testing Programs. U.S. and Iowa birth orders were calculated from the *Natality Report* of the U.S. Bureau of the Census (1951–1991). The U.K. birth order was taken from the Office of Population Censuses and Surveys (1987), and it includes Wales, whereas the A-levels are only those from the United Kingdom. The Welsh population, however, constitutes only 5% of the combined birth cohort.

[2] All the F ratios of the regression analyses were significant well beyond the $p = .001$ level. The complete results of these regression analyses can be found in Zajonc (1996). The Iowa and U.K. test data were regressed only on aggregate birth order. However, the Quantitative SAT scores were regressed not only on birth order ($\beta = -3.12$, $SE = 0.63$) but on the proportion of men taking the test ($\beta = -0.25$, $SE = 0.22$), the proportion of Whites taking the test ($\beta = 1.52$, $SE = 0.59$), the proportion of the birth cohort taking the test ($\beta = -0.92$, $SE = 0.56$), and parental income ($\beta = 0.03$, $SE = 0.10$). SATs had a partial correlation with birth order equal to $-.79$.

Vol. 52, No. 7, 685–699

Figure 1
U.S. Math Scholastic Assessment Tests (SATs), U.K. A-Levels, and Iowa Basic Skills Scores for Iowa Children and Their Respective Birth Orders



Note. The birth order data lag SAT scores by 18 years, the A-levels by 17 years, and the Iowa scores by 9–16 years. The birth order scale is inverted such that higher values represent lower birth ranks.

Robert B. Zajonc

Cross-sectional studies also show a strong relationship between aggregate test scores and birth order (e.g., Belmont & Marolla, 1973; Breland, 1974; Davis, Cahan, & Bashi, 1977; Institut National d'Études Démographiques, 1973; Zajonc, 1976, 1983; Zajonc & Bargh, 1980b; Zajonc, Markus, & Markus, 1979). In these studies, temporal trends were not examined. Rather, the researchers compared average test scores of large numbers of individuals whose birth orders were known. The comparisons of average test scores classified according to the birth order of the test takers also yielded systematic differences, with as much as 96% variance in aggregate test scores accounted for by birth order (Zajonc & Bargh, 1980b). Yet similar studies in which individual test scores were examined for their correlation with the individual's particular birth order failed to account for more than 3% of variance in intellectual scores (Brackbill & Nichols, 1982; Ernst & Angst, 1983; Galbraith, 1982; Grotevant, Scarr, & Weinberg, 1977; Hauser & Sewell, 1983; Kessler, 1991; Olneck & Bills, 1979; Retherford & Sewell, 1991; Schooler, 1972; Steelman & Mercy, 1980; Svanum & Bringle, 1980; Zajonc & Bargh, 1980a; but see Zajonc, Markus, Berbaum, Bargh, & Moreland, 1991). In fact, in the Belmont and Marolla findings, aggregate birth order accounted for nearly 80% of all variance, but when analyzed individual-by-individual, birth order had only a negligible effect in the very same data set (Ernst & Angst, 1983).[3]

How can we reconcile these conflicting findings? Are the aggregate results "true" and the individually distributed analyses "false?" Or is it the other way around? The reader will see that both results are true, but they answer different questions. There are two ways of formulating the problem of birth order. Studies that report negative birth order results formulate the problem in terms of individual differences. These studies seek to explain all variance in test scores, and their research strategy is to examine as many factors as possible, birth order included, that influence individual test scores and compare their relative contributions to the total variance explained. And it is indeed the case that in comparison with such factors as parental education or socioeconomic status (SES), birth order alone is found in these individually distributed effects studies to contribute very little variance.

The problem of birth order can also be formulated by asking a different question: Other things being equal, to what extent are variations in birth order alone associated with variations in intellectual scores?[4] In other words, the focus is not on explaining all individually distributed variation in intellectual scores, but the focus is on birth order in particular and not on its role in comparison with other factors that have been identified as important sources of variation in test scores. In this article, we ask the derivative question, namely whether secular trends in aggregate test scores have any association with trends in birth order, an association that is clearly suggested by the data in Figure 1, and how this association can be explicated.

How is it possible for the same variable to account for almost all the variance in the criterion when analyzed in the aggregate (i.e., when averages of birth order are correlated with averages of test scores) and for a negligible amount of variance when correlations are computed between birth ranks of particular individuals and their particular scores? Belmont and Marolla (1973), in their study of 386,114 Dutch recruits, reported birth order differences, independent of family size, significant at $p < 10^{-13}$. The aggregate data yielded 77% of the variance explained by birth order alone (Zajonc & Bargh, 1980b). However, the same data set had a raw β, estimated by Marjoribanks and Walberg (1975), of only $-.078$, which translates into an increment in test scores of less than one tenth of a standard deviation per unit of birth order. Which of the two sets of figures quantify the birth order effect?

There may not be a meaningful answer to this question, because, as the reader shall see, both forms of analysis (distributed and aggregate) have their own interpretations and implications. But the distinction has profound empirical, methodological, and theoretical consequences because contradictory conclusions, often not acknowledged as such, have been drawn from both sources of analysis. There are several phenomena in which aggre-

[3] Differences in IQ associated with birth order are not trivial. For example, Belmont and Marolla (1973), in a study with nearly 400,000 18-year-old male participants, found a difference of about one fourth of a standard deviation in Raven scores (equivalent to four IQ points) between firstborns and fifthborns in families of five, and Breland (1974), in a study of nearly 800,000 National Merit candidates, showed, for the same comparison, a difference in IQ of nearly eight points (i.e., more than half of a standard deviation; p. 1013). As explained in the text, in the aggregate these differences are not small at all and have serious collective consequences.

[4] Note that we do not regard birth order or family configuration to be the main or even partial causes of intellectual development. Rather, they are viewed here as quantifiable factors that show consistent correlations with test scores.

Patricia R. Mullally

gate and individually distributed analyses yield different outcomes, and quite often similar controversies arise. For example, Mischel (1968) alarmed psychologists by stating that personality measures do not predict behavior, not to any respectable degree of validity. This conclusion was based on a review of individually distributed analyses in which a host of other factors that also determine behavior were not measured or controlled. However, Epstein (1986, 1990) showed that when outcomes are aggregated, a substantial relationship between personality and behavior is found. Smoking, too, shows strong consequences for lung cancer when analyzed in the aggregate but shows little when individually distributed chances of lung cancer are calculated for smokers and nonsmokers. Accordingly, those who wish to deny the effects of smoking cite individually distributed data, and those who are opposed cite aggregate data. Ernst and Angst (1983), critics of birth order effects, took the very low β weight, estimated for individual data as the "true" evidence of the absence of birth order differences, and ignored systematic aggregate birth order differences reported by Belmont and Marolla (1973) and Breland (1974).

The statistical effects of aggregation can be estimated by an adaptation of the Spearman–Brown prophecy (Zajonc, 1962). The typical use for the Spearman–Brown formula allows us to estimate an increment in test reliability when we increase the number of test items, provided all items have the same correlation with the criterion as the original set of items. However, the formula can be adapted for the distinct form of aggregation that is involved in birth order data. An example will be helpful. In one of the earliest experiments on group judgments, Gordon (1924) had 200 participants rank weights for their magnitude. Each participant's ranking of weights was compared with the true order, and Gordon obtained an average correlation of .41. But he then randomly constructed artificial groups of 5, 10, 20, and 50 individual judgments, with each group having its own average. When these average judgments were correlated with the true order, the correlations were .68, .79, .86, and .94 for averages of 5, 10, 20, and 50 randomly selected grouped judgments, respectively. Eysenck (1939) raised reliability of judgments from .47 to .98 simply by aggregating individual judgments into groups of 5, 10, 20, and 50. Note that the analyses that at one time explained 0.17% and at another 96% of the variance were performed on the very same data points, except for grouping or aggregating of the data. The first represents an average correlation, and the second represents the correlation of averages.[5] Thus, when birth order effects are evaluated on an individual-by-individual basis, the obtained relationship is essentially an average correlation. However, when researchers average, for any given trend, yearly birth orders and yearly test scores and calculate a correlation coefficient between them, the relationship obtained is the correlation of averages.[6] It is important to note in this respect, however, that although aggregate analysis will generally yield higher correlation coefficients, the power of the independent variable to make predictions for single individuals remains unchanged.

---

[5] The following approximation of the Spearman–Brown prophecy formula (Zajonc, 1962) estimates the correlation of averages, r_{xy}, from the average correlation, r_{xy}, where n is the number of observations in each unit or category of the aggregate.

$$r_{xy}^2 = \frac{nr_{xy}^2}{1 + (n - 1)r_{xy}^2}.$$

The approximation assumes that $r_{xy}^2 = r_{yy'}$, the average intercorrelation among participants. It follows from the approximation that given one entry per cell, as is the case in averaging individual correlations, $n = 1$, and given $r_{xy} = .1$, we obtain $r_{xy}^2 = .01$, the amount of variance typically accounted for an individual-difference approach to birth order that uses distributed analysis. However, if the number of observations in each unit of the aggregate (i.e., in each cohort) is increased thousandfold (i.e., $n = 1{,}000$) and if r_{xy} remains equal to .1, we obtain a correlation of averages, r_{xy}, equal to .95.

[6] Note that when the prophecy formula is applied to estimate increments in test reliability by adding test items, it is a forgone conclusion that the reliability will increase. Such an increase in correlation between birth order and test scores will also be obtained when we aggregate within the entire unchanged n of observations, as, for example, in the distributed and the aggregated scores of the Belmont–Marolla (1973) data. But not all methods of aggregating will increase the reliability or correlation coefficient. In fact, the increment in a correlation coefficient deriving from the above form of aggregation can occur only if there is an underlying and pervasive relationship in the total n between the two variables. That not all forms of aggregation must necessarily result in increased correlation coefficients is shown by the following example in the SAT data. Within any cohort, individual SATs and the individual's parental income are correlated between .30 and .45. When we aggregate, say within the 1995 cohort, and examine the relationship between SAT scores and parental income, the correlation rises to .99. The same, of course, is obtained for the aggregated Belmont–Marolla data. But when we aggregate income differently, for example, over the 20-year secular trend (1975–1995), the relationship between the aggregate SATs and aggregate parental income drops from .35 to −.19! This means, of course, that the association between SATs and birth order in the three secular trends shown in Figure 1 is not simply or only a matter of aggregating individual scores. It means that there exists an underlying and pervasive relationship between the two variables.

Even if the aggregate regression coefficient for birth order and SATs were over .90, the birth order of a single individual taken at random would still be a poor predictor of his or her SAT score. This point was clearly demonstrated by Ross and Nisbett (1991, pp. 102–108) in their cogent analysis of the Mischel–Epstein controversy. Of course, aggregate trends in birth order remain excellent predictors of aggregate trends in SATs.

Why Should Birth Order Influence Test Trends?

Researchers often find trends that, quite accidentally, show close correlations with other unexpected factors. For these accidental correlations, subsequent studies eventually discover that they are either chance events or that both factors are actually influenced by a third one. But the parallels between the trends of birth orders and test scores are not fortuitous, given that over the 30 years there were over 3 million observations per birth order data point and over 1 million per SAT data point (and given that test scores consistently track birth order trends for three quite different populations and for different tests). More important, however, these trend data are entirely in agreement with a theory—the confluence model—published 20 years ago and are in fact predicted by the model (Zajonc, 1976). We discuss here the implications of the confluence model for understanding the test trends observed in Figure 1.

The basic idea of the confluence model is that intellectual growth of children can be enhanced or hindered by the immediate family circumstances, which by virtue of their family position are different even for children of the same family. This is evident by comparing some siblings for their linguistic experiences—experiences that have a profound influence on their later test scores (Huttenlocher, Haight, Bryk, & Seltzer, 1991). Compare, for example, the linguistic experiences of a firstborn son before the birth of his younger sibling with those of a girl who already has several younger brothers and sisters. During his presibling period, this firstborn interacts mainly with adults. He is exposed to a fairly sophisticated language, a rich vocabulary, and a variety of life's domains. He lives in an adult world. Just being surrounded by adults gives him auditory access to a pool of words, many of which he will be asked about in his SATs. However, the later-born girl who has several younger siblings hears a less diverse and more limited language, has less interactive access to her parents, and witnesses less often the process and product of abstract thought. She hears the language of toddlers because she lives in a toddler's world. The pool of words in which she is immersed is restricted, more primitive, contains very few low-frequency items, and does not afford the formation of complex sentences. In fact, Hart and Risley (1995) were able to quantify the pool of words in which children are immersed. They report that children of parents on welfare hear about 600 words per hour, children of working-class parents hear 1,200 words, and children of professional parents hear as many as 2,100.

The confluence model quantifies the way intellectual growth is influenced by the intellectual environment in the home, focusing on family configuration. But how can we quantify intellectual environment? Obviously, such a quantification requires a simplification, as all models do—a simplification that in return allows for precise formulation that produces testable and reliable estimates of the empirical data. Thus, we simply assume that the more intellectually mature that the people are who interact with the growing child, the more mature are the verbal, analytical, and conceptual experiences of the child. These experiences, we assume, will reveal themselves in the test scores of the child. Test scores may or may not measure what is popularly thought of as "intelligence," and it is true that the concept of intelligence itself needs better understanding (Neisser et al., 1996). But test scores represent a stark reality. Whatever they measure, one's life course can be significantly changed by what scores one manages to obtain.

A few simplified examples illustrate the major concepts that can be found in greater detail in prior publications (Zajonc, 1976, 1983; Zajonc & Bargh, 1980b; Zajonc et al., 1979). Consider the intellectual environment of the firstborn: Typically, the family environment consists of two mature adults and the newborn.[7] If we assign numerical values to this situation, say in mental-age units, such as those used in the Stanford Binet, we might be able to quantify in the abstract the quality of the family environment as it changes during the course of the child's development. We can assign, for instance, a value of 30 to each of the parents and 0 to the newborn child, take an average $(30 + 30 + 0)/3 = 20$, and consider this value as representing some approximation to the relative level of intellectual resources accessible to the newborn child, assuming all other things are constant.

Say, after four years, a new child is born into the above family. The second born's intellectual environment at birth is now $(30 + 30 + 4 + 0)/4 = 16$. And say that, after a lapse of three more years, there is a third offspring. The average value is reduced further: $(30 + 30 + 7 + 3 + 0)/5 = 14$. Thus, each successive sibling is born into a more "diluted" intellectual environment.

Note a few patterns:[8]

1. The individual whose score we are interested in predicting is also included in calculating the environment. In most analyses of environmental effects, the individual and the environment are treated as distinct entities. But here, because we are dealing with effects that occur over time, the person is treated as part of his or her own dynamically changing environment that influences him or her. The individual is part of that environment because,

[7] Actually, family patterns are rapidly changing and the proportion of single-parent households has grown substantially, especially in Europe.

[8] The confluence model does not use averages, as in the simplified examples, but root mean squares. Also, individual levels for a given age are calculated by assuming a growth function, $f(t) = 1 - e^{-k^2t^2}$. See Zajonc and Bargh (1980b) for a complete discussion of the formal aspects of the model.

like other family members, he or she influences it also and, as a result, is the target of a subsequent environmental influence in which he or she had some part. If A influences B at Time 1, and B influences A at a later Time 2, then A was influenced by A's own environment that A previously influenced. It is the case that parents have an influence over their children. But it works the other way too. Just observe the language of parents who have a new baby or listen to the speech of kindergarten teachers. For unspecified reasons, Ernst and Angst (1983), however, considered this feature of the model to be its defect. Yet, if this theoretically compelling assumption is not made, empirical data cannot be fitted by the model.

2. Gaps between successive births are also important: For the last born, a gap of six years is better than a gap of two years, because he or she has an older sibling who knows more of what he or she needs to know.

3. Twins have a poorer environment because both have an immature sibling at birth: $(30 + 30 + 0 + 0)/4 = 15$. This is in accordance with data showing that the IQ of twins is about five points less than singly born children and that of triplets is seven points less (Record, McKeown, & Edwards, 1970).

4. Children in one-parent families are also at a disadvantage because there is only one adult to contribute to the average level of intellectual environment (Carlsmith, 1964): $(30 + 0)/2 = 15$. Conversely, a larger number of adults (e.g., uncles, grandparents, child-care persons) provide an environment that is richer in intellectual resources.

5. Parental absolute scores remain unchanged at 30, assuming that by that age they have reached the typical asymptote. However, should the model be applied to populations for whom it can be assumed, for whatever reason, higher or lower adult intellectual levels, parental values can be adjusted accordingly (see, e.g., Zajonc & Bargh, 1980b), a procedure that elevates or reduces the absolute level of the predicted scores.

These illustrations specify the environment only at birth. One should not extrapolate from these illustrations, however, that the relative quality of the environment is constant over the entire growth period, such that it is always richer for the firstborn than for later borns. A subtle nontrivial inference follows from the confluence model, for the model predicts important changes, indeed a reversal of birth order effects that take place soon after the second child is born. In fact, in only a few years, the environment of the second born surpasses the first. This unexpected feature of growth patterns was revealed not by data but by the mathematics of the confluence model, and it turned out subsequently to be empirically true. This is so because two factors interact over time. The environment of the firstborn, although superior at birth, when measured at age four—$(30 + 30 + 4 + 0)/4 = 16$—is, in fact, inferior to that of the second born when also measured at age four, which is $(30 + 30 + 8 + 4)/4 = 18$.

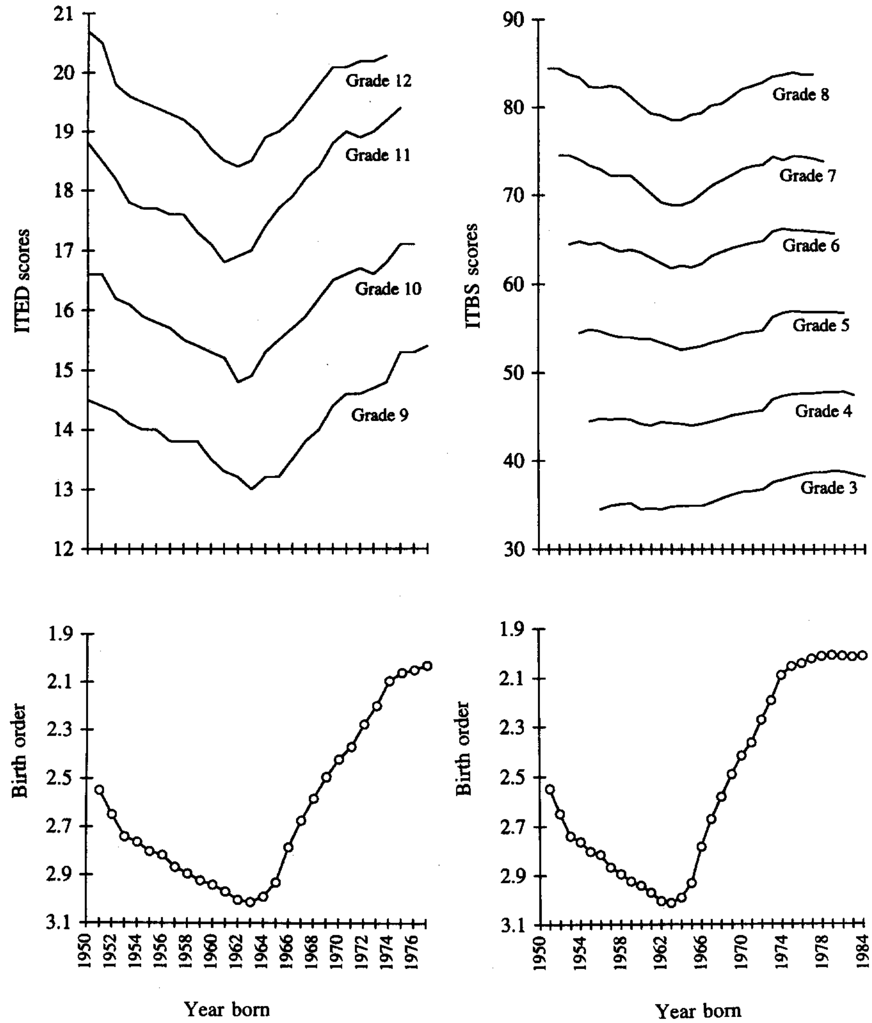
At the same time, however, according to the confluence model, another factor comes into play, a factor that has a positive developmental influence and that can more than compensate for the poorer environment of the firstborn in the early years of his or her life. There is a basic difference between the first- and last-born child, affording the former an opportunity for growth enhancement that is denied to last and only children. The firstborn in a family of two or more often acts as a surrogate parent. His or her younger siblings ask questions about the meanings of words, they need help with various tasks (e.g., how to hold a bat), and they appeal to their older sibling to explain how things work or why they work in a certain way. In short, the firstborn child is in a way a "tutor" to the younger siblings, which has been shown to enhance mental growth and academic achievement (Bargh & Schul, 1980; Wagner, 1982; Whitman, 1988). However, since a one-year-old seldom requires an intellectual tutorial, older siblings do not begin to benefit from their teaching function until the younger sibling can begin to ask questions—perhaps after age 2—and can enter into an interaction in which some cognitive processes are involved. Eventually, the contribution of the teaching function overcomes the deficit of the firstborn's early environment, and it does so at about age 11 (± 2 years).

All children are last born for some period of time, and during that period of growth they do not benefit from the teaching function. The only child never has the opportunity to teach; he or she remains the last born at all times. And when measured past age 11 ± 2, that child's scores are generally lower than those of children in families of two.[9] It is a subtlety of the theory and of the empirical data that, when overlooked, may easily lead to discounting the birth order effect, for it turns out that test scores increase with birth order but only when measured past age 11 ± 2. Because some studies found positive test scores to increase with birth order, whereas others found them to decrease, the entire field of birth order research was declared by some (Ernst & Angst, 1983; Schooler, 1972) to be chaotic and inconsequential.[10] But, mirabile dictum, when age of testing is taken into account, very meaningful patterns emerge. A French survey (Tabah & Sutter, 1954) of national intellectual levels reported data of siblings in intact two-child families that showed that the IQ of the second born surpasses that of the firstborn for ages 6 and 7, is about the same for the ages 8 and 9, and falls below that of the firstborn afterwards. This pattern of data is replicated in 50 studies

[9] The preponderance of one-child families in China offers a very interesting population for an intensive study of family configuration effects on test scores.

[10] The effects of intervals between successive births are also age dependent, and they depend on birth rank as well. For example, when tested at age seven, the intellectual environment of the firstborn who has a sibling five years younger is not as rich as the firstborn who has a sibling only two years younger: $30 + 30 + 7 + 2 < 30 + 30 + 7 + 5$. The longer the interval is, the less the benefit. The reverse is true for the later-born child. When tested at seven, the second born with an interval of five years is better off than the second born with an interval of two years: $30 + 30 + 12 + 7 > 30 + 30 + 9 + 7$. The longer the interval is, the greater the benefit.

Figure 2
Iowa Test of Basic Skills (ITBS) and Iowa Test of Educational Development (ITED) Composite Standard Scores and Birth Order of Iowa Children in Grades 3–12



that were analyzed more than a decade ago (Zajonc, 1983; Zajonc et al., 1979). We show below another example of the age dependence of birth order effects (see Figure 2). Perhaps inattention to this subtlety of birth order effects is one reason why the critics concluded that there are no birth order effects at all. Ernst and Angst (1983, pp. 43–69) ignored age dependence of birth order effects and cited as counterevidence to the model many studies of very young children (e.g., Svanum & Bringle, 1980) that failed to show birth order effects—studies that according to the confluence model should not show positive birth order effects in the first place.

The interaction of the two factors—the modulating effects of the "teaching function" and of the intellectual environment—explains many seemingly anomalous phenomena. Whereas the firstborn's situation is "diluted" by new siblings, he or she can eventually benefit from their presence by acting as a tutor to them. Because the only child can seldom assume a teaching function and thus benefit from it, his or her scores are generally lower than those of two-children families. This explains why the only child does not obtain the highest test scores and why, therefore, the family-size effect is often not monotone and is at a maximum for a two-child family (Zajonc, 1983; Zajonc et al., 1979). When measured past age 11, the only child has lower scores than one from a two- and sometimes even three-child family. In an analysis of six large data sets for the effects of aggregate birth order, populations tested at ages younger than 11 did not show the typical decline of scores with aggregate birth order (Zajonc & Bargh, 1980b). Only those tested at ages 17 and 18 featured the decline of scores with lower aggregate birth orders.[11] When the Iowa scores plotted above are broken down by grades, a clear birth order effect is obtained only for children in grades higher than fifth, that is, older than 11 years (Figure 2). No other theory seeking to explicate birth order effects acknowledges the age dependence of these effects let alone offers a conceptual analysis of these patterns.

Alternatives to the Confluence Model

There were some attempts to attribute birth order and family-size effects to a resource depletion process (Becker & Tomes, 1976; Blake, 1989; Gottfried, 1984; Lindert, 1977; Murnane, Maynard, & Ohls, 1981; Taubman & Behrman, 1986). The depletion theory posits that there is a fixed or nearly fixed inventory of resources (i.e., financial, affective, intellectual), resources that, when divided among an increasing number of siblings, leave a decreasing portion for each, which, researchers claim, explains why children from larger families have low test scores. Because there is less to draw on for each successive child, the depletion theory also claims to explain why test scores decrease with birth order.

The resource depletion theory encounters at least three serious contradictions. According to the resource depletion theory, only children should score the highest on intellectual performance tests. They do not (e.g., Breland, 1974; Zajonc et al., 1979). Second, the positive relationship between intellectual scores and birth order should be independent of age of testing. It is not (Tabah & Sutter, 1954; Zajonc et al., 1979). Third, according to the depletion model, the decline of scores with family size should be modulated by income. Children from families with higher SES levels should not show a drop in test scores with increasing family size because their resources are large enough to support more offspring, and later-born children in high SES families, therefore, should not suffer substantial resource disadvantages. Yet, in contradiction to depletion theory, the relationship between family size and intellectual scores is the same for all socioeconomic levels (e.g., Breland, 1974; Zajonc, 1976). Of course, the absolute level of scores rises with SES or parental education. However, because the same developmental advantages and disadvantages among siblings, relative to each other, are constant, the patterns of differences in test scores among distinct birth orders also remain the same across the various SES levels, and differences in test scores associated with family size have the same patterns as well. Only the intercepts of these two functions change with SES.

The Confluence Model Cannot Be Tested on Individuals

We have no practical means at present to test specific derivations from the confluence model on individual data points. During the course of the child's growth period, major changes occur when new siblings are born, or when some of them or any member of the household leaves the home. Some older children act intensively and extensively as tutors. Others participate marginally. Some parents relate to successive children in the same way, and others are more inconsistent. In some families, the standard of living may rise with the age of children, and in others it may decline. One child may have inspiring teachers or friends, whereas another lives in a dull neighborhood. Some parents love to read to their children, whereas others prefer to watch television. And a variety of traumatic events can have an impact on the development of the child. To take account of all these and similar conditions that may in some way influence test scores is virtually impossible in a sample of individual cases.

Moreover, a test of the confluence model, on the basis of individually distributed birth order effects, is simply impossible because the complexity of such a test goes well beyond even the most generous research means at researchers' disposal. Consider the task. Suppose we start modestly with no more than five siblings. If we wish to consider gender, we begin with 62 sibship categories (Male [M], Female [F], MM, MF, FM; FF, MMM, MMF, etc.). But the confluence model specifies birth intervals to be important also. Excluding multiple births, which could be studied separately, we may wish simply to examine differences only between short and long birth intervals. We now have 1,778 sibling categories and 8,258 participant categories. Thus, if we find a population that can generate 10 participants per cell, we need 100,000 participants, and probably many more, because some sibling categories will be very difficult to locate. We also know that the age of testing is important. The longitudinal approach would require repeated measures on these thousands of participants over a substantial period of time. And we have not begun to control for the host of other psychological, social, experiential, and economic factors that may contribute to the test scores, nor have we committed ourselves to observe the interactions between siblings and parents or among the siblings and record the tutorial help offered and received. Perhaps, there is only one test of the confluence model that approaches these requirements. It is a study of two-child families measuring both children's IQs at different ages (Tabah & Sutter, 1954). Consistent with the confluence model, this study clearly demonstrated the age dependence of birth order effects. Inferences consistent with the confluence model are therefore best examined on large samples, using aggregate analyses. Given a large sample and a classification of participants in which factors confounding birth order effects are allowed to be absorbed by the averaging process, strong aggregate effects of birth order are clearly and systematically revealed. That is the case when test trends are analyzed and when large samples are grouped according to family positions. Given that aggregate pat-

[11] The partial correlation coefficients of aggregate test scores with aggregate birth order, when family size was held constant, for participants over 11 years of age were −.88 and −.90, whereas those for younger participants were both positive, .73 and .05. The respective standardized βs were −.53 and −.82 for the older participants and .27 and .02 for the younger participants (Zajonc & Bargh, 1980b).

terns of birth order explain almost all the variance in test score trends means that, at the limit, birth order, even though obscured by extraneous noise, must be a meaningful factor at the individual level as well.[12]

In this sense, the problem of testing the confluence model is the same problem that is present in statistical mechanics. The motions of millions of individual particles cannot be recorded and measured simultaneously, and even though the theory of these aggregate motions derives from laws of motion of individual bodies, the outcomes are in aggregate form. The same is true of the confluence model. Even though it has been written for the individual case, it can thus far only have been examined on aggregate data.

Massive Gains in IQ: The Flynn Effect

In contrast to the SATs, A-levels, and the Iowa scores time series, all of which had clear periods of decline, Flynn (1987) offered data showing that aggregate test scores have been steadily rising. These data, which came to be known as the Flynn effect, have received a variety of interpretations (Neisser et al., 1996). In general, however, the IQ averages are viewed as reflecting positive and desirable changes at the level of individual scores. That is, taken individually, "people are becoming smarter" (Azar, 1996, p. 20), and they are becoming smarter at the rate of three to seven IQ points per decade.

There are three features of the Flynn (1987) data to be noted. For the most part, the rising trend was interpolated from a few data points, recorded in different countries, separated by varied intervals, and distributed over several decades. Many of the data points also came from populations of varied age. The U.S. data, for example, combined populations between 2 and 75 years of age. Yet the age of testing, as we pointed out above, is a crucial factor in test scores. It is likely, therefore, that the nonmonotonicity of the SATs, the A-levels, and the Iowa tests found in Figure 1 is probably concealed in Flynn's interpolated figures. Flynn (in press) proposed that SATs and IQ measure different proficiencies, but that would not explain why there are declines on tests that are highly related to the IQ measures (e.g., SATs) and why the minima on all the curves happen to occur precisely for 1962 $\pm$ 2 birth cohorts, which are the cohorts of the lowest aggregate birth order. Thus, instead of the three-fifths standard deviation increase that would be expected if there was in fact a three-point-per-decade rise, Quantitative SATs have yet to equal the high they reached 30 years ago. A-levels were also in a decline for the cohorts born between 1943 and 1953, from 74% to 67% passing the test (Figure 1).

There are two ways an average of a frequency distribution can change. When we say that the price of the average house has risen in 1996, this might mean that a comparable house bought in the previous year was less expensive. Or it might simply mean that the proportion of better houses sold in 1996 was greater than in 1995. The same holds for averages of test scores. This average can change over time when each or most of the scores of comparable individuals changed. That is, test takers may be smarter. Or it can change when the distribution of the scores is now different without anyone being more intelligent than students were previously. Of these two quite obvious possibilities, it is the first that is taken as the official or, at least, the attractive interpretation of the Flynn effect. Yet it is more likely that, at least in large part, average scores increased by virtue of changes in the proportion of earlier-born children in the cohort, a hypothesis left unexplored thus far.

Consider the following example that uses realistic figures. In 1962, Iowa recorded 14,964 first and 14,464 second births out of a total of 60,990. Thus, early borns represented 48.25% of all these births. One decade later, there were fewer total births in Iowa, namely 47,234. Of those, 16,892 were first and 12,746 were second births. Thus, the percentage of early births in Iowa in 1972 was considerably larger: 62.75%. It has been shown (Zajonc, 1976) that the proportion of SAT scores over 500 and the proportion of firstborns in the population are highly correlated. For the sake of argument, assume that, on the average, the overall aggregate Iowa IQ for the cohort born in 1962 was 100. Assume also, consistent with a variety of birth order data, that early borns scored, on the average, 105 on the IQ test. This must mean that the IQ of the later-born children was 95.3. What would the Iowa aggregate IQ be in 1972, given the changes in the birth order composition of the cohort but no changes in the test scores of early borns versus later borns? It can be computed as follows: $62.75(105) + 37.2(95.3) = 101.4$ (an increment of 1.4 only by virtue of recomposition of the Iowa cohort birth orders). The early borns were just as "smart" in 1972 as they were in 1962. Nor did the later borns change scores. There were just more early borns in 1972. Therefore, more often than not, the rise and decline in scores are better understood by examining changes in fertility patterns.

Collective Potentiation of Birth Order Effects: A Bonus?

Note that the estimated figure of a 1.4 increase in IQ that was due to changes in the Iowa cohort composition alone (see above) is less than the three-point (and certainly less than the seven-point) increase per decade that was predicted by Flynn (1987). The fall and rise of SATs with birth order also exceeds the amount we would expect from static cross-sectional aggregate birth order effects (Zajonc & Bargh, 1980b). For example, the drop in Quantitative SATs seen in Figure 1 between birth cohorts in 1950 and in 1962 was one fourth of a standard deviation, whereas the increase in birth order figures was only on the order of half of a birth rank. But if we consult the cross-sectional analysis such as first offered by Belmont and Marolla (1973), a difference of half of a birth rank

[12] We are careful to say that they explain *trends,* or differences in aggregate scores of children in different positions in the family configuration, not absolute levels of scores.

corresponds to a difference of only one tenth of a standard deviation in intellectual test scores. Hence, the actual SAT drop for the above 12-year period was 2.5 times greater than we would expect from the corresponding birth order changes alone.

We might speculate, however, that the hypothetical 1.4-point Iowa rise that was due to recomposition of the cohort's birth order as well as the large decline in SATs between 1950 and 1962 birth cohorts could be accompanied by what might be a *collective potentiation* of birth order effects, that is, an increment in intellectual scores that is greater than one could derive from the average birth order changes alone. Consider the following argument. A year's generation of newborns with 20% of first births transfers this proportion to the classrooms and to the entire peer environment. If firstborns constitute 40% of the cohort, the typical classrooms will have a greater proportion of more advanced students who set higher standards and provide challenges to others, resulting in higher overall performance not just of firstborns but of later borns as well. Thus, a cohort's average test scores may rise first because the cohort comprises a larger proportion of early borns who themselves score higher on intellectual tests and second because their presence in the cohort recalibrates standards for all members of the cohort who now will develop faster and reach higher levels of proficiency.

Collective potentiation of individually distributed effects is a common phenomenon found in many domains, such as epidemiology, economics, and political behavior. It is manifest in protests, riots, revolutions, and other forms of mass movements (LeBon, 1895; Moscovici, 1985). In marketing, we see it in fashions and fads. The effect of popularity of a consumer item has an effect on individual preferences, but its combined aggregate effects, whereby popularity augments preferences and preferences increase popularity, are a clear bonus (Cooper, 1973). The dynamics of collective potentiation are clearly apparent in the fluctuations of the stock market. People tend to buy and sell a stock not only because they have some information—information of doubtful validity—that it will go up or down but simply because other people, who are equally ill informed, are buying and selling that stock. In many cases, collective potentiation assumes significant, regional, national, and even international ramifications, leading to the emergence of institutional consequences in the form of organizations, parties, sects, dieting clubs, and doomsday groups.

The collective potentiation of birth order works in a similar fashion. Because the quality of our education depends to a great extent on our peers, a given year's offspring with a higher proportion of early borns has a "bonus": They will benefit from the higher standards of excellence throughout their schooling and advance more rapidly in the more challenging intellectual environment. However, not only students would be affected by a greater proportion of firstborns. Teachers, too, would advance their students more rapidly because they would be confronted with a more favorable distribution of grades than would be the case in a generation of students with few firstborns. The benefits of these collective programs carried out by a community of learners are extensively discussed by Bruner (1996) and by Brown and Campione (1990). Under these circumstances, it is not difficult to imagine that when birth order rises, the average national or state test scores would also rise by a quantity higher than estimated from individual birth order changes alone. The effects would rise not only because there are more firstborns in the cohort but because all members of the cohort would benefit from collective potentiation provided by the school–peer environment. Here, then, is a case when, given a large proportion of firstborns in the classroom or in the peer group, the whole becomes greater than the sum of its parts.[13] At the same time, the collective potentiation of birth order effects is one of the important reasons why aggregate data are so much more robust than individually distributed birth order effects.

Creative products, be they in the realm of science, technology, sport, art, literature, music, or drama, are not randomly distributed over the globe or over the history of civilization. They emerge in a few highly concentrated times and locations. All but 1 of the 26 National Hockey League head coaches come from Canada. The finest string instruments come from Cremona. Brilliant theatrical productions are concentrated in London's West End and New York's Broadway. The best software comes from Silicon Valley. In these locations and in these times, a few highly talented individuals set high standards for others, invented new techniques, and propagated new approaches. They inspire others who build on examples set for them and, in turn, improve on previous products that challenge the original contributors to seek to achieve even higher standards.

In general, many group differences, especially those in test scores, might well derive from collective potentiation effects. A large proportion of low or high scores in a cohort will have potentiating effects on the entire cohort regardless of their origins. Thus, for example, if stereotype threat (Steele, 1997) depresses scores of individuals by an amount x, then the cohort will suffer by an amount $x + \epsilon$, where ϵ is the collective potentiation of individual effects.

How Long Have IQ Scores Been Rising?

There is no doubt that the average scores have been rising somewhat. Certainly, a larger percentage of our population reaches a higher level of education today. The standard of living has increased enormously since 1918. And the SES of the average person has risen in most parts of the world, affording educational attainment to ever-

[13] The collective potentiation of birth order effects could be readily estimated by comparing the scores of children in a given sibling size and a given birth rank who come from cohorts with a small proportion of firstborns with one with a high proportion of firstborns. Cohorts with a high frequency of firstborns should generally show higher scores for comparable sibling size and birth rank. In fact, the clearest results should be obtained for only children.

increasing segments of the population. The metacognitive environment that greets the 21st century is no doubt richer.

Together with these changes, however, there also have been natality changes that have had substantial impact on family configuration. In most countries, the birth rate has declined substantially during this century—by 50% in most Western countries. The Flynn effect has been illustrated by a linear increase of IQ from 1919 until 1990 (Azar, 1996; Horgan, 1996). Most people tested in 1919 were born about 1900, and those tested in 1990 were born in 1971. It is interesting, therefore, that over the 1900–1971 interval, birth rates dropped in Belgium from 28.9 to 14.4, in Denmark from 29.7 to 15.2, in France from 21.3 to 17.1, in Italy from 33.0 to 16.8, in the Netherlands from 31.6 to 17.2, in the United Kingdom from 28.7 to 15.9, and in the United States from 30.7 to 17.2. And birth rates are correlated with the average orders of births. Because we saw a strong relationship between birth order and test scores in both the rise and the decline of these scores, there is no doubt that birth order itself contributed to the secular trend in intelligence. This assertion does not contradict the contribution of other factors, such as parental education or SES. But the three-point-per-decade increase cannot be documented as a continuous aggregate trend into the past, not only because we are still waiting for SATs to return to their 1969 record levels, but because we would have to suspect that our ancestors of only one century ago were morons.

SATs and A-Levels Into the 21st Century

On the basis of the close association between birth order and test scores, some predictions of future test trends can be made because most of those who will be taking their SATs and A-levels in 17 or 18 years have been born, and we know their birth orders. On the basis of these data, it appears that if no other factors intervene, and if scores are not again recentered or restandardized, SATs will remain quite stable for the next two decades, whereas A-levels should begin to drop next year, will continue dropping slightly until the year 2000, and then will level off. This prediction, of course, holds only if more rigorous standards are not imposed on the exams in the meantime. But this prediction is made with some confidence, because taking account of birth order trends alone, an accurate prediction was made 20 years ago for SATs, forecasting a decline that would continue until 1980 followed by a rise thereafter (Zajonc, 1976). The confirmation of this prediction over 40 years is seen in Figure 1.

Family Environment and Test Scores

Given that test scores have tracked aggregate birth order yearly over several decades, in more than one country, on more than one test, and in populations of different ages, we must conclude, contrary to some recent views (Baumrind, 1993; Neisser et al., 1996; Rowe, 1994), that family environment factors are far from trivial in the analysis of test scores and test score trends. The recent work (Plomin & Daniels, 1987) that showed a correlation among siblings to be as low as that among random individuals seemed to support the view of the negligible effects that families have on their offspring. The work on differences among siblings, however, has not taken account of variations due to birth order. It is entirely possible that siblings are quite different from one another (Plomin & Daniels, 1987) because each grows up in a different environment that changes as new siblings are born and mature (Richardson, 1936).

Alternative Explanations of the Link Between Test Trends and Family Trends

Only 30 years ago, SATs went into free fall, as did A-levels 40 years ago. The decline was immediately met with a host of unsupported conjectures. Parents were not spending enough quality time with their offspring; TV was rotting their brains; schools were neglected by local governments; and crime, drugs, smoking, rock music, and, yes, eroding standards were often listed as serious contributing causes of the decline. But when SATs began to rise in the 1980s, President Reagan did not hesitate to take full credit. Each year, if the results of SAT tests or A-levels are down, school officials, governors, presidents, and the entire educational system get blamed for the decline. It is, therefore, not surprising that when the test scores are up the same officials do not hesitate to take credit. Last year's SATs reached a 20-year high. Last year, also, the highest ever proportion of students in the United Kingdom passed A-levels. Predictably, the two countries reacted in their distinctive ways. The United Kingdom recoiled in shock that "standards were eroding" (Charter, 1996), whereas Americans were self-congratulatory. Experts affirmed with satisfaction that "more rigorous courses . . . started to pay off" (Arenson, 1996).

Could we view the changes in the national SATs, the Iowa scores, and the A-levels as mediated by changes in standard of living, SES, expenditures per pupil, and so on?[14] None of these factors, even considering appropriate time lags, show either a minimum for the 1962 cohort, a decline before, or a rise after that year. It is obvious that one could not attribute the three aggregate test trends shown in Figure 1 to a changing SES of these populations. It suffices to consider the following example. In 1965, 1980, and 1995 (which correspond to the 1947, 1962, and 1977 birth cohorts), the average national Quantitative SAT scores were 493, 466, and 482, respectively. According to the Educational Testing Service, a 16-point difference in SATs corresponds to a difference of about $10,000 in 1995 constant dollars of parental income. This would mean that the entire population of SAT takers

[14] The correlation between intellectual test scores and school expenditures per pupil is equal to .33, but there are glaring outliers. For example, North Dakota has the highest SATs in the nation (515 Verbal and 592 Quantitative), but it spends only $4,636 per pupil. New York spends more than twice as much ($9,300), but it ranks 42nd in SATs (419 Verbal and 473 Quantitative). See also footnote 6.

would had to have dropped in their family income by \$19,000 between 1965 and 1980 and then risen again by \$11,000 between 1980 and 1995. In fact, there was no low point for parental income of the national SAT takers in 1962. The correlation between aggregate Quantitative SATs and parental income was in fact −.19.

It is interesting in this respect that, at the individually distributed level, in contrast to birth order effects, SES explains a great deal of variance but is incapable of explaining secular test score trends. In the Broman, Nichols, and Kennedy (1975) study, of all factors examined, the socioeconomic index, of which income was the most important component, contributed the largest proportion of variance to test scores. Most correlations between SES and test scores on individually distributed data fall between .25 and .50 (Jensen, 1969). Yet SES does not figure as a predictor of secular test score trends (see footnote 6). Variations in income are conspicuous and large at the individually distributed level and in comparing economically different regions, say counties of the United States. However, an average income of a country or a state does not change dramatically from year to year. And it is this lesser variability in SES, much restricted in its range, that cannot feature SES as a significant factor in SAT trends.

Verbal SATs

Verbal SAT trends are somewhat different from Quantitative SATs. Even though, following the 1980 minimum, the Verbal SATs have a flatter curve than their Quantitative counterparts, they still show a partial correlation with birth order that is equal to −.79. The regression analysis reveals that factors other than birth order had additional and stronger influence on the Verbal SAT time series than was the case with Quantitative SATs. The standardized β coefficient for the proportion of White students taking SATs is considerably larger for Verbal SATs than for Quantitative SATs. The explanation of this difference might lie in the dramatic changes that occurred in the demographic composition of the SAT cohorts, especially in those born after 1967 (i.e., taking their SATs after 1985). In 1975, for example, the SAT takers comprised 7.9% African Americans, 1.4% Hispanic Americans, and 1.4% Asian Americans. In 1995, these percentages were 9.7, 7.4, and 7.6, respectively, and were increasing at very different rates.

Important in this respect is the fact that the Asian American population, which has increased quite substantially in the percentage of SAT participants, has typically higher quantitative scores than the White population but has lower verbal scores. For example, in 1995, Asian Americans scored 538 on the quantitative section, whereas Whites scored 498. But Asian Americans had only 418 on the Verbal SAT, whereas Whites scored 448. Both differences are equivalent to about one third of a standard deviation. In contrast, the African American and the Hispanic American groups showed differences between quantitative and verbal scores that were similar to those of the White students.

If verbal skills are more language dependent than quantitative skills, then quantitative scores would vary less with ethnic background than verbal scores. A quadratic equation is the same in Chinese, Spanish, and English. But a substantial proportion of students from some ethnic groups do not speak English as their first language. They are, therefore, at a disadvantage in terms of lexicon, idioms, local knowledge, associations, analogies, acquaintance with proverbs, popular similes, and all those other particular verbal resources that are required for quick answers to the Verbal SATs. Other changes in the composition of test takers included a drop among male test takers from 50% to 46% and an increase in the percentage of students in their cohorts taking the tests, from 23% to 32%. It is interesting, however, in this respect that the Verbal SAT scores of Whites were 500 in 1941 and dropped to 454 fifty years later, whereas their Quantitative SATs remained the same (Hayes et al., 1996). Clearly, the change in Verbal SATs cannot be explained by the demographic changes in the SAT population alone, and other factors, thus far not identified, must contribute to these variations in trends.

Another recent explanation of the disparity between Quantitative and Verbal trends focuses on schoolbook simplification (Hayes, Wolfer, & Wolfe, 1996). Hayes et al. examined samples of 800 elementary school readers for their lexical difficulty (Zakaluk & Samuels, 1988). Comparing readers published between 1919 and 1945 with those published between 1946 and 1962 and those published between 1963 and 1991, they found that there was a substantial decline in lexical difficulty. Hayes et al. attributed the pattern of verbal scores to lexical simplification of the reading materials. This hypothesis, of course, cannot apply to the Quantitative SATs, to A-levels, or to the Iowa data, all of which increased after 1980. Nor do they agree with the differential birth order effect in the Iowa scores when age of testing is taken into account (Figure 2). More important, the data for elementary and high school reading tests in the state of New York showed the same rise as the other time series for the cohort born past 1962 (Zajonc, 1976).

Caution

We must not draw premature conclusions about family planning on the basis of the confluence model and the data that it describes. Score differences between siblings and among families of different sizes are indeed small. But what are *small* differences? The 1% of variance explained by birth order found by Grotevant et al. (1977) was considered of trivial significance. Marjoribanks and Walberg's (1975) analysis of the Belmont and Marolla (1973) study that generated a β of −.078 for birth order is viewed by some as sufficient to disprove the confluence model (Ernst & Angst, 1983).

When the problem is formulated to describe all individually distributed effects on intellectual scores—that is, the investigation seeks to identify all the sources of variations that produce or are associated with individual differences in intellectual scores—birth order will be found as a variable contributing relatively little variance. But, outside of parental education and SES, most other variables have only a negligible influence over test scores. Broman et al. (1975), examining 26,760 four-year-olds, were able to account for no more than 30% of variance altogether even though they entered 27 variables in their regression analysis that included not only mother's education but such factors as nystagmus, Apgar score, and neonatal hematocrit. Not surprising, because their participants were only four years old, they did not show consistent birth order effects.

But consider an example from epidemiology. An extensive study of 355,000 men screened in the early 1970s for cholesterol level (U.S. General Accounting Office, 1996) revealed that a difference in deaths from coronary heart disease (CHD) between those with a normal cholesterol level (i.e., 198 mg/dl) and those in the danger zone (226 mg/dl) was only .3%, with the former group having a risk of 4 in 1,000 deaths and the latter group 7 in 1,000. The corresponding survival figures for the two categories of risk were 99.6% and 99.3%, respectively, a difference that strikes us as negligible. And, accordingly, the relationship between cholesterol level and mortality six years after screening had a β of .00006! Yet major policy decisions are carried out with enormous sums of money invested in research, growing pharmaceutical profits, and patients' expenses, all justified by these risk figures. Note in this respect that in comparison with the U.S. General Accounting Office's raw β, the Marjoribanks and Walberg (1975) estimate of $-.078$ for birth order is more than 1,000 times greater than the comparable figure for cholesterol's role in CHD deaths. Of course, each β must be interpreted within its own context. However, it cannot be argued, on the strength of a few selected calculations that the influence of birth order is so weak as to be disregarded. Just as cholesterol is regarded as a risk factor, birth order can also be interpreted as a risk factor, a risk that may deny college entrance or a successful career. Hence, birth order trends characterized by fluctuations as large as 30 SAT points may have more significant consequences than it would at first appear. The implications of the above are especially important when researchers consider that the individually distributed analyses and aggregate-pattern analyses yield such distinctly different results: the first quite weak and the latter quite substantial but both correct. The argument implies that family configuration factors can be invoked in large-scale and long-term policy planning but are not very useful in understanding why Joe or Jill has low scores on the SAT. Yet if we know that Joe is a laterborn child and comes from a large family, whereas Jill is firstborn in a family of two, we could specify some differences in their risks of academic failure or success. And, although very small, these risks are more substantial than those associated with cholesterol.

Many factors, such as parental support, enthusiasm for one's teachers, effort, or, on the negative side, a neighborhood environment that derogates academic values, all contribute to a particular student's SATs. These factors are held constant when large aggregate data sets are analyzed. Moreover, test scores are not everything. Birth order and family size may contribute quite differently to psychological well-being and economic attainments than to test scores. For example, children from larger families might be more affiliative, more affectionate, good leaders, less prone to depression, or otherwise healthier. We simply have no data for these outcomes. We do know, however, from the work of Sulloway (1996) that when firstborn and later-born scientists are compared, the former are more likely to represent conservative positions and to cling to established theories whereas later borns are the revolutionaries. Darwin (the fifth of six children) and Copernicus (the last of three children) are classic examples.

Test Score Differences and Their Collective Consequences

The differences associated with birth order when compared at the individually distributed level appear small indeed, as small as 1.5 IQ points per unit of birth order. But they are quite substantial when we inspect the secular trends. For instance, over a period of 12 years, between 1968 and 1980 (i.e., for the 1950 and 1962 birth cohorts), the Quantitative SATs dropped from 494 to 466 (see Figure 1). That amounts to one fourth of a standard deviation. The collective consequences of such differences are considerable. A community of 1,000 individuals with a mean IQ of 100 has 22 individuals who score below 70, that is, individuals who have to be cared for by the community, and 22 individuals who score above 130, that is, individuals who are most likely to contribute to the community's prosperity and progress. A one-fourth standard deviation shift in IQ, say a drop to a mean of 96, increases the number of those who have to be cared for to 41 and reduces the number of exceptionally gifted to 12. The changes in the intellectual resources of this community are enormous, even though a difference in IQ score of 4 points for a particular individual is within the range of measurement error.

Promise

The resolved paradox of analyses that explain little variance when using individually distributed data versus aggregate analyses that explain almost all variance does not apply uniquely to the birth order puzzle alone. It can be readily transferred to a variety of social psychological problems. Most commonly, scholars working at the individual level, particularly in the area of personality and developmental psychology, tend to ignore collective and cohort effects, and those working at the collective level tend to ignore the individual processes. This has certainly been the case with birth order effects. The separation of these approaches is not surprising given that some fields (e.g., anthropology, demography, economics, sociology, and political science) specialize in phenomena that take place mainly at the aggregate level, whereas psychology is preoccupied with individual phenomena. One branch of psychology is ideally suited to bridge the individual–collective chasm: social psychology. Ventures of this type, however, have been extremely rare, even in social psychology. The emergence of interdisciplinary initiatives between aggregate-focus and distributive-focus fields, such as economic psychology, cognitive anthropology, cultural psychology, and so on, is a promising reaction that will view phenomena from both an aggregate and a distributive perspective.

Although the SATs might be stable and may not drop, they are at miserably low levels. They were in the 490s for the cohort born in the mid-1940s. For the cohort born in 1962, the SATs were at their all-time low. Today they are at about 460 for the combined verbal and quantitative sections of the test. We cannot trust the Flynn effect to bail us out, given that extrapolation from the observed trends proves to be dangerous. Yet the scores can be higher. We have more and better means at our disposal now than at any other time in history.

Unlike the resource depletion model, which might advise reducing family size or increasing resources per family, the confluence model has some direct practical implications. We can change the scholastic experience of our children. One of the factors contributing to intellectual development is the tutorial function of the older children. Being a teacher or a tutor can substantially augment and accelerate mental growth. There are data showing that the teacher gains more than the learner in the process of teaching (Bargh & Schul, 1980; Kagan, 1992; McMahon & Goatley, 1995; Wagner, 1982; Whitman, 1988). Peer teaching can be easily implemented in our elementary and middle schools. For example, half of a class could be instructed about decimals. The other half could learn about the lowest common denominator. Then, after each has acquired some minimal understanding, they would teach each other. Impressive improvements in a variety of skills and domains have been achieved through such reciprocal teaching procedures (Bruce & Chan, 1991; Kelly, Moore, & Tuck, 1994; Moore, 1988). The key to the improvement is that the students are imbedded in a *metacognitive environment*—an environment in which items of knowledge are distributed over the community of learners (Bruner, 1996) who know who might know what and how well (Brown & Campione, 1990). The sharing of such knowledge and skills takes effect more rapidly than in a classroom conducted by a single teacher who is the sole source of knowledge and expertise.

The trends and changes in scores that we have analyzed here are a clear reflection of family patterns, confirming the important role of environmental factors and, especially, family and peer influences, which have been repeatedly downsized and outsourced in favor of genetic hypotheses. These diverse data support derivations from the confluence model—data that cannot be explained by alternative theories. Contrary to the prevailing doubts in its explanatory value, birth order is being rapidly reinstated as a salient factor in psychology.

REFERENCES

Arenson, K. W. (1996, August 23). Students continue to improve, College Board says. *The New York Times*, p. A1.

Azar, B. (1996, June). People are becoming smarter—Why? *The APA Monitor*, p. 20.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefit of teaching. *Journal of Educational Psychology, 72*, 593–604.

Baumrind, D. (1993). The average expectable environment is not good enough: A response to Scarr. *Child Development, 64*, 1299–1317.

Becker, G. S., & Tomes, N. (1976). Child endowments and the quantity and quality of children. *Journal of Political Economy, 84*, S143–S162.

Belmont, L., & Marolla, F. A. (1973, December 14). Birth order, family size, and intelligence. *Science, 182*, 1096–1101.

Blake, J. (1989). *Family size and achievement.* Berkeley: University of California Press.

Brackbill, Y., & Nichols, P. L. (1982). A test of the confluence model of intellectual development. *Developmental Psychology, 18*, 192–198.

Breland, H. M. (1974). Birth order, family configuration, and verbal achievement. *Child Development, 36*, 1011–1019.

Broman, S. H., Nichols, P. L., & Kennedy, W. A. (1975). *Preschool IQ: Prenatal and early developmental correlates.* Hillsdale, NJ: Erlbaum.

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking: Or a context by any other name. *Contributions to Human Development, 21*, 108–126.

Bruce, M. E., & Chan, L. K. (1991). Reciprocal teaching and transenvironmental programming: A program to facilitate the reading comprehension of students with reading difficulties. *Remedial and Special Education, 12*, 44–54.

Bruner, J. S. (1996). *The culture of education.* Cambridge, MA: Harvard University Press.

Carlsmith, L. (1964). Effect of early father absence on scholastic aptitude. *Harvard Educational Review, 34*, 3–21.

Charter, D. (1996, August 5). A-level record sparks standards row. *The London Times*, p. 1.

Cooper, L. G. (1973). A multivariate investigation of preferences. *Multivariate Behavioral Research, 8*, 253–272.

Davis, D., Cahan, S., & Bashi, J. (1977). Birth order and intellectual development: The confluence model in the light of cross-cultural evidence. *Science, 196*, 1470–1472.

Epstein, S. (1986). Does aggregation produce spuriously high estimates of behavior consistency? *Journal of Personality and Social Psychology, 50*, 1199–1210.

Epstein, S. (1990). Comment on the effects of aggregation across and within occasions on consistency, specificity, and reliability. *Methodika, 4*, 95–100.

Ernst, C., & Angst, J. (1983). *Birth order: Its influence on personality.* Berlin, Germany: Springer-Verlag.

Eysenck, H. J. (1939). The validity of judgments as a function of the number of judges. *Journal of Experimental Psychology, 25*, 650–654.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Flynn, J. R. (in press). Intelligence on the rise? Secular changes in IQ and related measures. In U. Neisser (Ed.), *The rising curve.* Washington, DC: American Psychological Association.

Galbraith, R. C. (1982). Sibling spacing and intellectual development: A closer look at the confluence model. *Developmental Psychology, 18*, 181–191.

Gordon, K. H. (1924). Group judgments in the field of lifted weights. *Journal of Experimental Psychology, 3*, 398–400.

Gottfried, A. W. (Ed.). (1984). *Home environment and early cognitive development: Longitudinal research.* New York: Academic Press.

Grotevant, H. D., Scarr, S., & Weinberg, R. A. (1977). Intellectual development in family constellations with adopted and natural children: A test of Zajonc and Markus model. *Child Development, 48*, 1699–1703.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: P. H. Brookes.

Hauser, R. M., & Sewell, W. H. (1983). Birth order and education attain-

ment in full sibships (Working Paper No. 83-31). Madison: University of Wisconsin, Center for Demography and Ecology.
Hayes, D. P., Wolfer, L. T., & Wolfe, M. F. (1996). Schoolbook simplification and its relation to the decline in SAT—Verbal scores. *American Educational Research Journal, 33,* 489–508.
Horgan, J. (1996, November). Get smart, take a test: A long-term rise in IQ scores baffles intelligence experts. *Scientific American,* pp. 12–24.
Huttenlocher, J., Haight, W., Bryk, A., & Seltzer, M. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236–248.
Institut National d'Études Démographiques. (1973). *Enquête nationale sur le niveau intellectuel des enfants d'âge scolaire* [National survey of the intellectual level of school-age children]. Paris: Author.
Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1–123.
Kagan, S. (1992). *Cooperative learning.* San Juan Capistrano, CA: Resources for Teacher.
Kelly, M., Moore, D. W., & Tuck, B. F. (1994). Reciprocal teaching in a regular primary school classroom. *Journal of Educational Research, 88,* 53–61.
Kessler, D. (1991). Birth order, family size, and achievement: Family structure and wage determination. *Journal of Labor Economics, 9,* 413–426.
LeBon, G. (1895). *Psychologie des foules* [Psychology of crowds]. Paris: Olean.
Lindert, P. H. (1977). Sibling position and achievement. *Journal of Human Resources, 12,* 198–219.
Marjoribanks, K., & Walberg, H. J. (1975). Birth order, family size, social class, and intelligence. *Social Biology, 22,* 261–268.
McMahon, S. I., & Goatley, V. J. (1995). Fifth graders helping peers discuss texts in student-led groups. *Journal of Educational Research, 89,* 23–34.
Mischel, W. (1968). *Personality and assessment.* New York: Wiley.
Moore, P. J. (1988). Reciprocal teaching and reading comprehension: A review. *Journal of Research in Reading, 11,* 3–14.
Moscovici, S. (1985). *L'âge des foules.* Paris: Maison des Sciences de l'Homme.
Murnane, R. J., Maynard, R. A., & Ohls, J. C. (1981). Home resources and children's achievement. *Review of Economics and Statistics, 63,* 369–377.
Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.
Office of Population Censuses and Surveys. (1987). *Birth statistics: Historical series of statistics from registrations of births in England and Wales, 1837–1983.* London: Her Majesty's Stationery Office.
Olneck, M. R., & Bills, D. B. (1979). Family configuration and achievement: Effects of birth order and family size in a sample of brothers. *Social Psychology Quarterly, 2,* 135–148.
Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences, 10,* 1–60.
Record, R. G., McKeown, T., & Edwards, J. H. (1970). The relation of measured intelligence to birth order and maternal age. *Annals of Human Genetics, 34,* 61–69.
Retherford, R. D., & Sewell, W. H. (1991). Birth order and intelligence: Further tests of the confluence model. *American Sociological Review, 56,* 141–158.
Richardson, S. K. (1936). The correlation of intelligence quotients of siblings at the same chronological age levels. *Journal of Juvenile Research, 20,* 186–197.
Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* New York: McGraw-Hill.
Rowe, D. C. (1994). *The limits of family influence: Genes, experience, and behavior.* New York: Guilford Press.
Schooler, C. (1972). Birth order effects: Not here not now! *Psychological Bulletin, 78,* 161–175.
Steele, C. M. (1997). A threat in the air: How stereotypes shape the intellectual identity and performance. *American Psychologist, 52,* 613–629.
Steelman, L. C., & Mercy, J. A. (1980). Unconfounding the confluence model: A test of sibship size and birth order effects on intelligence. *American Sociological Review, 45,* 571–582.
Sulloway, F. (1996). *Born to rebel.* Westminster, MD: Random House.
Svanum, S., & Bringle, R. G. (1980). Evaluation of confluence model variables on IQ and achievement test scores in a sample of 6- to 11-year-old children. *Journal of Educational Psychology, 72,* 427–436.
Tabah, L., & Sutter, J. (1954). Le niveau intellectuel des enfants d'une même decrease famille [Intellectual level of children of the same family]. In *Le niveau intellectuel des enfants d'âge scolaire.* Paris: Presses Universitaires de France.
Taubman, P., & Behrman, J. R. (1986). Effects of number and position of siblings on child and adult outcomes. *Social Biology, 33,* 22–34.
U.S. Bureau of the Census. (1951–1991). *Natality report.* Washington, DC: U.S. Government Printing Office.
U.S. General Accounting Office. (1996). *Cholesterol treatment: A review of clinical trials evidence.* Washington, DC: U.S. Government Printing Office.
Wagner, L. (1982). *Peer teaching.* Westport, CT: Greenwood Press.
Whitman, N. A. (1988). *Peer teaching.* College Station, TX: Association for the Study of Higher Education.
Zajonc, R. B. (1962). A note on group judgments and group size. *Human Relations, 15,* 177–180.
Zajonc, R. B. (1976, April 16). Family configuration and intelligence. *Science, 192,* 227–236.
Zajonc, R. B. (1983). Validating the confluence model. *Psychological Bulletin, 93,* 457–480.
Zajonc, R. B. (1996, June). *Intellectual environment and intellectual performance.* Paper presented at the convention of the American Psychological Society, San Francisco.
Zajonc, R. B., & Bargh, J. (1980a). Birth order, family size, and decline of SAT scores. *American Psychologist, 35,* 662–668.
Zajonc, R. B., & Bargh, J. (1980b). The confluence model: Parameter estimation in six divergent data sets on family factors and intelligence. *Intelligence, 4,* 349–361.
Zajonc, R. B., Markus, G. B., Berbaum, M. L., Bargh, J. A., & Moreland, R. L. (1991). One justified criticism plus three flawed analyses equals two unwarranted conclusions: A reply to Retherford and Sewell. *American Sociological Review, 56,* 159–165.
Zajonc, R. B., Markus, H., & Markus, G. B. (1979). The birth order puzzle. *Journal of Personality and Social Psychology, 37,* 1325–1341.
Zakaluk, B., & Samuels, S. J. (1988). *Readability: Its past, present and future.* Newark, DE: International Reading Association.