CORRECTED SEPTEMBER 13, 2007; SEE LAST PAGE

Psychological Review
2007, Vol. 114, No. 3, 806–829

0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.3.806

Resolving the IQ Paradox: Heterosis as a Cause of the Flynn Effect and Other Trends

Michael A. Mingroni
Newark, Delaware

IQ test scores have risen steadily across the industrialized world ever since such tests were first widely administered, a phenomenon known as the *Flynn effect*. Although the effect was documented more than 2 decades ago, there is currently no generally agreed-on explanation for it. The author argues that the phenomenon *heterosis* represents the most likely cause. Heterosis, often referred to as *hybrid vigor,* is a genetic effect that results from matings between members of genetically distinct subpopulations, such as has been occurring in human populations through the breakup of small, relatively isolated communities owing to urbanization and greater population mobility. In Part 1 of the article, empirical findings are listed that are consistent with a heterosis hypothesis but render other hypotheses either implausible or very difficult to test. In Part 2, a formal model of the process of heterosis is presented. The goal of the modeling is to develop a quantitatively rigorous method for estimating the potential contribution of heterosis in the Flynn effect, as well as trends observed in other heritable traits and conditions.

Keywords: heterosis, Flynn effect, IQ, secular trend

One of the most puzzling phenomena in the field of psychological research today is the steady increase that has occurred in IQ test scores, commonly referred to as the *Flynn effect*. The trend toward higher scores has been documented in at least 20 countries, including all of the world's major industrialized nations, although the pace of change has varied somewhat in different countries, at different times, and on different tests (Flynn, 1994). On common test batteries like the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991), gains have typically been on the order of 3 points, or 0.2 standard deviations, per decade. This means that in many countries today a person of average IQ would have been in approximately the top 15% of same-age scorers 50 years ago. Gains on tests such as Raven's Progressive Matrices (Raven, Raven, & Court, 1998), which involve finding patterns in images comprised of geometric shapes, have tended to be larger, as much as 7 points per decade in some cases. In a few countries, particularly in Northern Europe, the trend appears to have stopped, or at least slowed dramatically, with scores possibly even starting to trend downward (Sundet, Barlag, & Torjussen, 2004; Teasdale & Owen, 2005). Although many explanations have been put forward (Neisser, 1998), there is currently no generally agreed-on cause for any part of the trend. In the words of one researcher, the trend is "officially mysterious" (Deary, 2001, p. 112).

In a previous issue of *Psychological Review*, Dickens and Flynn (2001) identified what is perhaps the most puzzling characteristic of the rise in IQ test scores. They referred to this puzzling aspect of the trend as the *IQ paradox*. The so-called paradox stems from the fact that IQ has displayed consistently high heritability over the many decades in which scores have steadily risen. Because the authors dismissed the possibility that the trend could be genetic in origin, the high heritability of IQ would seem to require positing an implausible environmental "factor X" that somehow varies greatly over time to cause IQ gains yet never varies enough at any single point in time to reduce heritability very much. Dickens and Flynn then offered a formal model, which they claimed resolves the paradox without the need to posit either genetic change or an environmental factor X.

In the present article an entirely different, alternative resolution to the IQ paradox is offered. In this approach, the proposed cause of the Flynn effect is the genetic phenomenon *heterosis*, often referred to as *hybrid vigor*. The article is divided into two main parts. In Part 1 the case is made for heterosis as the most plausible cause of the Flynn effect. First, the many alternative hypotheses that have been proposed are listed, and then various empirical findings are presented that are consistent with heterosis but problematic for other explanations. In Part 2 a preliminary formal model of heterosis is presented that begins to paint a picture of what is likely to be occurring at the genetic level to cause IQ gains as a result of heterosis. The goal of the modeling is to begin the process of developing a systematic, quantitatively rigorous method for estimating the potential contribution of heterosis in the IQ trend, as well as trends in other traits that may also be the result of heterosis.

Briefly, heterosis is a genetic effect that will cause population-wide changes in a trait whenever three conditions are met. The first condition is that the population in question must initially have a mating pattern that is less than completely random prior to the occurrence of the trend. Such a deviation from panmixia, or random mating, creates an excess of homozygotes in the population and a deficit of heterozygotes. Second, the population must undergo a demographic change toward a closer approximation to

I would like to thank Yuko Munakata, Harlene Hayne, Linda Gottfredson, Jelte Wicherts, Wendy Johnson, and John Loehlin for their input and suggestions.

Correspondence concerning this article should be addressed to Michael A. Mingroni, 400 Elkton Road, Number 7, Newark, DE 19711. E-mail: mingroni@gmail.com

random-mating conditions. This causes the frequency of homozygotes to decline and that of heterozygotes to increase. Of course, this second condition presupposes that the first condition is already met, as a trend toward more random mating cannot occur in a population already mating randomly. Third, the trait in question must display directional dominance, with more of the genes that influence the trait in one direction being dominant and more of those that influence the trait in the opposite direction being recessive. Given such nonadditive gene action, any increase in the ratio of heterozygotes to homozygotes will cause the distribution of the trait to shift over time in the dominant direction. Heterosis has been mentioned as a potential cause of the IQ trend by a number of researchers over the years (Anderson, 1982; Flynn, 1998; Jensen, 1998; Kane & Oakland, 2000; Mingroni, 2004; see also Dahlberg, 1942, chap. 10). Few would dispute that heterosis could be responsible for at least some part of the trend; what is mainly at issue is whether it could be a major cause.

Part 1: The Case for Heterosis

Assessing the plausibility of heterosis as a cause of the Flynn effect cannot be done in a vacuum. Any scientific hypothesis must be judged relative to all other explanations that have been put forward to account for the phenomenon in question. This first section, therefore, begins with a list of the various hypotheses that have thus far been proposed to explain the IQ trend. After the alternatives have been laid out, a list of empirical findings related to the Flynn effect is presented, findings that are consistent with a heterosis hypothesis but that, taken together, render other hypotheses implausible or, at best, very difficult to test.

Explanations for the Flynn Effect

The following is a list of hypotheses that have been advanced to explain the rise in IQ test scores:

1. *Nutrition:* Improvements in the quantity and/or quality of food consumed by growing children (Lynn, 1998b). I would also include in this category explanations involving increased vitamin or micronutrient intake, as well as the mother's nutrition during the prenatal period.

2. *Medical care:* Improvements in both the quality of, and access to, health care services such as vaccinations, antibiotics, and advanced treatments.

3. *Education:* Improvements in the nature of, and access to, both formal and informal methods of cognitive training. In addition to formal schooling, I would include in this category the advent of things like television and video games, which, though informal in nature, are thought by some to develop and train cognitive abilities in a way similar to formal schooling.

4. *Test-taking attitude (the Brand hypothesis):* Children of the past were generally more risk averse than children of today and therefore less likely to attempt to answer difficult questions (Brand, 1987).

5. *Practice effects:* Children today take more IQ-type tests than children of earlier generations, which could improve scores in at least two ways. First, children who take tests frequently might develop better test-taking strategies. Second, children might know the answers to particular questions, having already seen the same or similar questions on tests they have taken previously.

6. *Selection:* If high-IQ individuals were to have more children and/or have their children at a younger age, then the frequency of IQ-increasing genes would rise over time, causing scores to increase.

7. *Smaller families:* Couples who have fewer children are able to focus more of their limited resources on each child. Therefore, a trend toward smaller families could raise mean IQ because the average child in the population would have more resources, both material and cognitive, devoted to him (Zajonc & Mullally, 1997).

8. *Genomic imprinting:* An epigenetic hypothesis put forward by Storfer (1999) whereby environmental factors such as visual stimulation affect the sperm of males, which in turn lead to changes in the early development of the brain.

9. *The Dickens–Flynn model:* An abstract model with two major features (Dickens & Flynn, 2001). First, the model posits a large number of environmental factors that are correlated with the genes. Changes in these environmental factors are in part responsible for the IQ trend. Second, the model posits a social multiplier effect, whereby the IQ of others in the population affects the IQ of individuals.

The IQ Paradox

Dickens and Flynn (2001) used the term *IQ paradox* primarily to refer to the fact that IQ has displayed high heritability during the many decades that scores have risen, presumably as the result of environmental change. In the present article, the term *IQ paradox* is defined somewhat more broadly to encompass all of the characteristics of the IQ trend that give it a profile that is inconsistent with environmentally driven change. When viewed in this more expansive way, the paradox can be seen to have at least four distinct components, each of which produces its own set of problems for environmental hypotheses.

First, as Dickens and Flynn (2001) pointed out, estimates of the heritability (h^2) of IQ are high, at about .75 in adults (Neisser et al., 1996). Without positing genetic change, this would seem to require positing environmental factors that cause large changes over time yet do not vary enough at any single point in time to reduce heritability estimates very much. Dickens and Flynn referred to such an implausible aspect of the environment as a "factor X." Of note, although Dickens and Flynn carried out their analysis using a value of .75 for h^2, they suggest that assuming values as low as .60 would still necessitate positing implausibly large change in those factors that do create environmental variance within generations (i.e., non–factor Xs). The magnitude of IQ heritability estimates, however, is only the first part of the problem.

In addition to the magnitude of IQ heritability, the fact that estimates appear to have remained stable over time (Jensen, 1998, pp. 322–323) is also a problem for environmental hypotheses. Some hypotheses, such as nutrition, suggest that IQ-depressing environmental factors kept individuals of the past far below their maximum genetic potential for IQ. However, unless these factors depressed everyone's IQ by the same amount, their removal from the IQ environment should have also removed an environmental source of variance, thereby causing heritability estimates to rise over time. Conversely, if the trend is due to something like a practice effect that has artificially raised IQ, this should represent the introduction of a new source of environmental variance that should have caused heritability to decline over time, unless, that is,

everyone today is practicing the same amount. The consistency of heritability estimates would therefore still pose a problem for environmental hypotheses, even if the estimates were lower, because it would suggest that large environmental change has occurred, without either the addition or subtraction of any noticeable source of environmental variance.

The third part of the IQ paradox is the apparent lack of shared family environmental factors that influence IQ, particularly adult IQ (Neisser et al., 1996, p. 85). Studies of adopted individuals have found their IQ in childhood to be somewhat similar to that of their adopted relatives, but the similarity diminishes as they get older, so that there appears to be very little, if any, IQ resemblance among adult adopted relatives (Loehlin, Horn, & Willerman, 1989; Plomin, Fulker, Corley, & DeFries, 1997; Scarr & Weinberg, 1978). Also, studies of monozygotic twins reared apart (MZA) have found their IQs to be nearly as similar as those of monozygotic twins reared together (MZT; Bouchard, Lykken, McGue, Segal, & Tellegen, 1990).

It is important to note that these MZA twin and adoption studies were conducted in populations that either were in the midst of ongoing IQ increases or saw large subsequent gains. As just one example, consider a study of elderly MZA twins in Sweden (Pedersen, Plomin, Nesselroade, & McClearn, 1992). These twins were born circa 1930 and so belonged to cohorts that scored far lower on IQ tests than subsequent cohorts. The MZA twins were correlated .68 on a general intelligence factor, as compared with .70 in a sample of same-age MZT twins. Findings like this suggest that whatever presumed environmental factors were depressing the IQs of individuals in the past could not have varied very much among households, as it does not appear to have made much difference whether monozygotic twins were reared in the same or different households; the IQs of both were depressed to the same enormous extent. Similarly, the presumed environmental cause of the Flynn effect could not have varied much among adoptive households, as any such variance would have made adopted relatives more similar.

The fourth part of the IQ paradox is the apparent lack of any Flynn effect within families (Rodgers, 1999). If the Flynn effect has been due to changes in those aspects of the environment that make siblings different, the so-called nonshared environment, then later born children should be higher in IQ than their earlier born siblings, on average, reflecting the presumed improvements in the environment over time. However, studies of the role of birth order on IQ appear to show no such effect. Between-family studies of birth order, which are methodologically unsound because they sample unrelated children from different families, show an apparent birth order advantage to the earlier born siblings (e.g., Belmont & Marolla, 1973), which is the opposite of what one would expect if the Flynn effect is due to changes in the nonshared environment. Within-family studies of birth order, which sample pairs of siblings from the same family and are therefore methodologically superior, appear to show little or no birth order effect (Rodgers, Cleveland, van den Oord, & Rowe, 2000).

The fact that IQ scores appear not to have risen within families suggests that the primary cause of the Flynn effect is unlikely to be part of the nonshared environment. It would seem to be a family effect. Specifically, it must be due to some factor that becomes fixed within each family at or before the birth of the first child and remains constant, affecting all full biological siblings to the same extent, on average, over the entire child-bearing period of the couple. Although constant within families, the factor has to vary among families over time; it has to cause higher IQ in the children of families started more recently, as compared with older families. The only siblings who appear not to be affected to the same extent, on average, are biologically unrelated children adopted into the home.

Before proceeding, it should be emphasized that the assertion that IQ scores have not risen within families is currently still open to some questioning. First, although birth order studies are relevant to the question, they were not specifically intended to illuminate the nature of Flynn effects. A Flynn-effect-specific study would have to incorporate the age difference between the siblings, as well as the pace of Flynn effects between the siblings' respective birth cohorts, rather than birth order per se. Also, even the methodologically superior within-family studies frequently test siblings at the same time, so that they are necessarily of different ages at time of testing. Each sibling's IQ score must therefore be determined by comparison to an age-appropriate norm, which creates the potential for bias from, among other things, Flynn effects.

To explain how Flynn effects might bias birth order studies, consider a test like the WISC (Wechsler, 1991), which has age-appropriate norms for children between the ages of 6 and 16. If these norms were generated in, say, 1970, then the 6-year-old norm would likely have been derived from children in the 1964 birth cohort, but the 16-year-old norm would have been derived from the 1954 birth cohort. A 3-point Flynn effect between the two cohorts would make the 6-year-old norm 3 points more difficult, on average, because it was generated using a later born, and hence higher performing, cohort. This probably explains why some within-family studies still appear to show an apparent advantage to earlier born siblings, albeit a much smaller advantage than shown by between-family studies (Ernst & Angst, 1983).

Although more research on this question would be helpful, it is important to emphasize that no one has suggested an IQ advantage to later born siblings, which is what would be expected if the Flynn effect was occurring within families. It is also worth mentioning that the potential problems with some of the birth order studies could eventually be sorted out by relying solely on studies in which siblings are tested at the same age. For example, in some countries with compulsory military service, virtually every pair of brothers has been tested at the same age, with the same IQ test often being used for decades. Adult data like this would not have any of the biases that might result from age norming.

With respect to the important question of whether the Flynn effect occurs within families, another possibility must be considered. It is possible that the apparent lack of birth order effect is the result of multiple environmental factors, some favoring later-borns and some favoring earlier-borns, which are canceling each other to produce zero net birth order effect. As just one possibility, greater maternal age might have a negative effect on a child's IQ, which would give an IQ advantage to earlier-borns. But nonshared Flynn effects could be giving an advantage to later-borns, canceling out the maternal age effect. One way to rule out such a possibility would be to compare birth order studies from populations that were experiencing rapid Flynn effects with similar studies from populations in which the Flynn effects had slowed or stopped (e.g., Sundet et al., 2004). If within-family Flynn effects are being canceled by other, equal but opposite, factors, then these other

within-family factors should become apparent in populations experiencing little or no Flynn effect. Another way to test for multiple offsetting factors would be to examine the subtest profile, or even the item profile, of birth order effects. Zero birth order effect on every subtest, or every item, would require multiple factors that happen to have profiles that are exactly equal in magnitude but opposite in direction, which would strain credulity. Further evidence bearing on this question is presented in a later section.

In addition to birth order studies, one could reasonably question to some degree the other empirical findings that create the IQ paradox. For example, one could reasonably argue for a somewhat lower estimate for the heritability of IQ in adulthood than the figure of .75 used by Dickens and Flynn (2001) in their analysis, which was based on the consensus report of an American Psychological Association task force (Neisser et al., 1996). This same report also used the phrase "quite low (zero in some studies)" (Neisser et al., 1996, p. 85) to describe the amount of IQ variance attributable to shared environment (c^2) in adulthood. But here again, one could reasonably find justification for a somewhat higher value. In fact, the same American Psychological Association report notes that studies of children tend to report higher estimates of c^2, and lower values for h^2, than those involving adults. Similarly, there is evidence that shared environment may be more important in families with low socioeconomic status (SES) (Capron & Duyme, 1996; Rowe, Jacobson, & Van den Oord, 1999) and may be underestimated in adoption studies owing to restriction of range of adoptive families (Stoolmiller, 1999). One could also argue that heritability estimates may have changed somewhat over time but not enough to be obvious, or that there have been multiple offsetting factors influencing heritability over time.

Arguments like those above, however, make environmental hypotheses only marginally more plausible. As mentioned above, Dickens and Flynn (2001) suggested that heritability as low as .60 would still require positing implausibly large environmental change. The findings of lower heritability and higher shared environment in children would be significant if the Flynn effect were restricted to children's IQ, but, in fact, adult samples seem to show the largest gains (Flynn, 1998, p. 27). Similarly, it is difficult to imagine that most of the trend, 21 points or more in some countries, could be accounted for solely by gains among the lowest SES segments of the population. Finally, even if a new consensus were to build around markedly different values for the parameters h^2 and c^2 in adulthood and/or future evidence showed a time trend in h^2, this would not, in and of itself, yield a compelling environmental explanation. One would still have to identify and measure the specific environmental factors responsible for the within-generation variance and show how changes in these factors over time could account for the between-generation variance. And so, although a specific compelling environmental explanation for the Flynn effect may one day be forthcoming, that day does not appear to be on the immediate horizon.

Resolving the IQ Paradox

The simplest and most direct way to resolve the IQ paradox in its entirety is to posit genetic change. If the Flynn effect has been primarily genetic in origin, one might still expect consistently high IQ heritability estimates over time. One would not necessarily expect adopted relatives to display much IQ resemblance, or MZA twins to differ much in IQ. Also, one would not expect any within-family Flynn effects, because all siblings, regardless of when they are born, would likely have an equal chance of inheriting their parents' IQ-increasing or IQ-decreasing genes. There are only two proposed genetic processes capable of causing systematic phenotypic change over time: heterosis and selection.

All environmental hypotheses involving postnatal factors are immediately rendered implausible by the IQ paradox. Such factors must have either a shared-family component, a nonshared component, or, as is likely in most cases, both shared and nonshared components. However, MZA twin and adoption studies suggest that the factor, unless it is a prenatal factor, does not have a large shared-family component. And birth order studies suggest that it does not have a large nonshared component. Therefore, the primary cause of the Flynn effect is unlikely to be part of the postnatal environment. We can therefore reasonably reject postnatal nutrition, education, medical care, practice effects, test-taking effort, and family size as major causes of the trend. Dickens and Flynn (2001, p. 348) similarly dismissed such factors because of the high heritability of IQ.

Among the remaining environmental hypotheses that have some reasonable chance to explain the Flynn effect are those involving prenatal environmental factors, which are not shared by adopted siblings but are shared by MZA twins. However, it must be emphasized that in order to explain the apparent lack of birth order effects, these prenatal environmental factors could not have improved within families over time. That is, they must be entirely shared prenatal factors. Consider the case of prenatal nutrition. Mothers in later generations could have had better nutrition than mothers of earlier generations. But for individual mothers, there could not have been any average improvement in nutrition levels from their earlier pregnancies to their later pregnancies, as this would show up in birth order studies. Looked at in another way, any improvements in nutrition occurring in the population could not have helped families in which there was already at least one child. One would need to posit prenatal environmental factors that vary among mothers over time but remain constant over the life of each individual mother. Although one might question their plausibility, the existence of such factors would represent a potential solution to the IQ paradox. It is worth noting that shared prenatal environmental factors are virtually impossible to disentangle from genetic factors.

Another hypothesis that is at least theoretically capable of resolving the paradox is the one involving genomic imprinting (Storfer, 1999). This hypothesis posits environmental changes to the father's sperm, which in turn affect the developing fetus. Because it is technically a prenatal (actually preconception) environmental factor, genomic imprinting offers the same prospect of resolving the IQ paradox as conventional prenatal environmental factors. As with conventional prenatal factors, however, all of the proposed changes to the sperm must occur prior to the birth of the first child in the family. Any additional genomic imprinting effects occurring over the couple's childbearing years would presumably show up as a birth order effect. Although genomic imprinting suffers from other problems, discussed below, and is very difficult to test, it is at least a potential solution to the IQ paradox.

It should be noted, however, that even environmental hypotheses involving shared prenatal factors face the problem of the consistency of heritability estimates over time. Whatever presumed changes may have occurred in the prenatal environment or to the sperm, they cannot have represented the addition, or subtraction, of a large source of environmental variance. For example, if poor prenatal nutrition depressed the IQ of individuals in earlier generations, then any variance in the effect would have reduced heritability estimates. More specifically, it would have increased the sibling correlation of earlier cohorts if it varied at all among mothers at any single point in time. Removing the shared environmental deprivation should have removed a factor whose variance made siblings more similar, and so should have been accompanied by a decline in the sibling correlation, which does not appear to have occurred (Jensen, 1998, pp. 322–323).

This brings us to the Dickens–Flynn model (Dickens & Flynn, 2001), which specifically claims to resolve at least one major part of the IQ paradox. When examined closely, however, the model falls far short of anything that might be considered a resolution to the problem. In fact, it raises as many puzzling questions as it claims to answer. Rather than clarifying the situation, it only serves to demonstrate just how difficult it is going to be to ever explain the Flynn effect as the result of environmental change. Briefly, the Dickens–Flynn model attempts to resolve the IQ paradox primarily in two ways. First, the authors posit a large number of environmental factors that are correlated with the genes. Some triggering mechanism shifts the mean effect of these environmental factors, thereby causing IQ gains. The second major feature of the model is a social multiplier effect, in which the mean IQ of the population affects the IQ of individuals.

The first problem with the Dickens–Flynn model is that positing environmental factors that are correlated with the genes and calling the variance they induce environmental variance is a highly questionable practice that has never been used before to analyze the variance of any trait in any species. Until Dickens and Flynn, it was generally thought that gene–environment correlations would cause the influence of environment to be overestimated, not underestimated (Rutter & Silberg, 2002, p. 471). In effect, the procedure allows the authors to shift an arbitrarily chosen amount of h^2 from the genetic side of the ledger to the environmental side. Although this shifted variance is now called environmental variance, it still retains all of the manifest characteristics of genetic variance, as described by long-established quantitative genetic theory. In the Dickens–Flynn model, high heritability is no longer a problem because the heritability now captures more environmental than genetic variance. As already discussed, positing genetic change immediately resolves the IQ paradox; thus, it is not surprising that positing change in "environmental effects induced by differences in the genotype" (Dickens & Flynn, 2001, p. 346) also resolves the problem. Unfortunately, even if one accepts the existence of these pseudogenetic environmental factors, the fact that they must be so highly correlated with the genes makes it impossible to disentangle their effects from those of actual genes. We must accept that they exist, distinct from the genes, on faith, as they cannot be measured empirically.

The second problem with the Dickens–Flynn model is that it still fails to confront the difficulty posed by the findings from adoption and MZA twin studies, which suggest that the cause of the Flynn effect cannot have a large shared component. Although Dickens and Flynn (2001) do not specifically partition the environment into shared and nonshared components, they seem to admit that the presumed environmental cause of the Flynn effect cannot have a large shared component (see their footnotes 8 and 4). However, even if we accept the existence of environmental factors that are highly correlated with the genes, something in the nature of the relationship between these factors and the genes must still change over time to cause IQ gains. In their terminology, there must be "triggers" that set off these changes. However, admitting that the cause of the Flynn effect cannot have a large shared component implies that the changes cannot be triggered in some households before they are triggered in others. Otherwise it would cause shared environmental effects to appear on MZA twin and adoption studies. Triggering that occurs with such a level of uniformity and pervasiveness across the affected populations strains credulity as much as any factor X. This point has also been discussed by Loehlin (2002, p. 758).

The idea of a social multiplier is also problematic for several reasons. First, it is not really specific enough to even qualify as a true causal explanation. In their model, the mean IQ of the population (P in their terminology) is an independent variable that influences the dependent variable, the IQ of the individual. However, the mean IQ of the population is, by definition, composed of the IQs of all of the individuals in the population, so that their proposed cause is identical to the effect they are trying to explain. The social multiplier is just another way of stating something we already know, namely, that mean IQ increases seem to be associated with the IQ increases of a large number of individuals. The same logic could just as easily "explain" a decline in IQ or stable population IQ, as Dickens and Flynn (2001) admit. What the social multiplier does not begin to tell us is why scores have gone up and not down; why increases occurred faster in certain countries, at certain times, and on certain tests; or where scores will go in the future.

Another difficulty for the social multiplier is that it would seem to be an entirely nonshared environmental factor, and so it runs into a problem from birth order studies, which suggest no within-family Flynn effects. In a population in which IQ is rising over time, siblings born at different times should be affected differently by the social multiplier, which in the model is tied directly to P, the mean IQ of the population. More specifically, later born siblings, who are born into a higher IQ population (higher P), should have higher IQ, on average, than their earlier born siblings. The apparent lack of IQ advantage to later born siblings, however, suggests that this does not occur. Dickens and Flynn (2001) never specifically address the question of whether the Flynn effect has occurred within or between families, or its implications for the plausibility of their model.

One last problem with the social multiplier is that the mean IQ of the population, by definition, never varies across the population at any single point in time but has changed over time. Therefore, the social multiplier, which is tied to the mean IQ of the population, qualifies as a factor X, with all of the inherent problems associated with the plausibility of such factors. For example, as just mentioned, the social multiplier would appear to have no shared family component. That is, it does not vary among families at any single point in time because the mean IQ is exactly the same for everyone in the population. Of interest, in response to a criticism by Rowe and Rodgers (2002), Dickens and Flynn (2002,

p. 769) admitted that contact with one's family should constitute a portion of the social multiplier effect, but their analysis showed that this component is very small. In other words, cognitive interactions with one's family do not have substantially more effect on the individual than interactions with complete strangers. Therefore, in accepting the concept of the social multiplier, we are asked to believe that an individual's IQ is influenced by the mean IQ of millions of people, the vast majority of whom he will never meet; however, the adult IQ of an adopted individual appears to be largely unaffected by the mean IQ of his adopted family, with whom he has spent time nearly every day of his life. It is difficult to imagine exactly how there could be a population social multiplier but no family or community social multiplier.

The idea of a population social multiplier has other troubling implications. For example, many citizens of high-IQ developed countries are currently living and raising families in low-IQ developing countries. This is potentially a very good thing, insofar as such families are likely to increase international and intercultural understanding among nations. However, if we accept the idea of an IQ social multiplier, we must ask whether the parents in these families are exposing their children to an IQ-depressing environment by raising them amid a population with low average IQ. It is doubtful that many today believe that there would be such an effect, but it would seem to be an inevitable consequence of the social multiplier.

Parallel Trends in Multiple Traits and Conditions

IQ is far from the only heritable human trait to have undergone large changes over time in multiple countries across the world. Storfer (1999) has cited parallel trends in myopia and brain size in arguing for genomic imprinting as the likely cause of the Flynn effect. Mingroni (2004) has cited trends in height, growth rate, myopia, asthma, autism, attention-deficit/hyperactivity disorder, and head dimensions, arguing that all represent the possible effects of heterosis along with the IQ trend. More recently, Kehle, Bray, Theodore, Zhou, and McCoach (2004) have added the increase in children diagnosed with emotional disturbance or social maladjustment to the list of possible heterosis effects. Additionally, Comings (1996) has gathered evidence of parallel trends in the incidence of several psychological disorders, including anxiety, depression, attention-deficit/hyperactivity disorder, and autism, and argued for selection as the likely cause (see also Rutter & Smith, 1995). As with IQ, there is currently no compelling explanation for any of the trends listed above.

The occurrence of parallel trends in many other heritable traits and conditions is entirely consistent with a heterosis hypothesis for the IQ trend. In fact, if heterosis has been a significant cause of the IQ trend, it would be surprising if no other trends were observed. This is because heterosis simultaneously affects all traits that display directional dominance. If heterosis is causing the Flynn effect, the only other traits that one would expect to be entirely unaffected would be those in which there is no directional dominance whatsoever, such as traits in which all of the gene action is perfectly additive or in which the dominant genes influencing the trait in one direction are exactly offset by dominant genes influencing the trait in the opposite direction. Although the model of heterosis presented in Part 2 of this article deals primarily with IQ, it will be shown that it can readily be adapted to the other traits and conditions listed above and possibly more that may become evident through future investigation.

Some of the traits listed above display heritability as high as, or higher than, IQ and so share at least part of the IQ paradox. Also, there is no clear evidence that the heritability of any of the traits has changed over time (but see Sibbald, 1997, for a possible exception in the case of asthma). In the case of at least one of the other trends, height, all four features of the IQ paradox have been documented, thus creating a "height paradox" identical to that seen in IQ. Even one of the earliest MZA twin studies of height, conducted in a U.S. population that was much shorter than today's population, showed the reunited twins to be remarkably similar in height (Newman, Freeman, & Holzinger, 1937). This would suggest that whatever presumed environmental factor was stunting the population must have been so uniform in its effects that it did not matter whether monozygotic twins were raised in the same or different homes; they were still stunted to virtually the same enormous extent. Also, there is no evidence of any birth order effect in height (Ernst & Angst, 1983).

The situation with heterosis as a potential cause of the height trend is remarkably similar to the situation with IQ. Many studies conducted over the years support the proposition that the genes that influence height display directional dominance (e.g., Billy, 1980; Damon, 1965; Ferak, Lichardnova, & Borjinova, 1968; Hulse, 1964; Schreider, 1969; Schull & Neel, 1965; Shapiro, 1936; Wolanski, Jarosz, & Pyzuk, 1970). Not surprisingly, therefore, given the demographic changes that have occurred, a number of researchers over the years have mentioned heterosis as a potential cause of the height trend (Billy, 1980; Damon, 1965; Ferak et al., 1968; Hulse, 1964; Mueller, 1986; Schreider, 1969; Tanner, 1990; Van Wieringen, 1986; Wolanski et al., 1970). In fact, the heterosis hypothesis was first put forward as the most likely explanation for the height trend over 60 years ago by one of the leading figures of 20th-century population genetics, Gunnar Dahlberg (1942). Dahlberg argued simply that the very high heritability of height rendered environmental hypotheses implausible and that the only reasonable genetic cause was heterosis, which he referred to as the isolate effect. He also suggested intelligence might be increasing as well. As with IQ, no effort has yet been made to measure the potential contribution of heterosis to the height trend, at least not in any quantitatively rigorous way.

One major difference between research into the height and IQ trends, however, is that among researchers investigating the height trend, there is not the same general recognition that the high heritability of the trait poses theoretical problems for environmental hypotheses as there is among IQ researchers. A good example of this can be seen in the following quote from a leader in the study of human growth, J. M. Tanner:

> Frequently, when one confronts nonbiologist audiences with the proposition that height is a proxy for economic conditions, one gets the comment "But surely height is inherited!" What has to be explained is that the variation between the heights of *individuals* within a subpopulation is indeed largely dependent on differences in their genetic endowment; but the variation between the means of groups of individuals (at least within an ethnically homogenous population) reflects the cumulative nutritional, hygienic, disease, and stress experience of each of the groups. In the language of analysis of variance, most of the within-group variation is due to heredity, and most of the between-group variation is due to childhood environment. (Tanner, 1994, p. 1)

This is precisely what Dickens and Flynn refer to as the IQ paradox, with height substituted for IQ. Unlike Dickens and Flynn, however, Tanner does not recognize the situation with respect to height as a paradox. The plausibility of factors that somehow vary a lot over time and between populations yet hardly vary at all within populations at any single point in time is never questioned. This issue is discussed in more detail in a later section that deals with explaining environmental changes.

Under the same reasoning that was used for the IQ paradox above, the height paradox renders most environmental hypotheses for the trend implausible. The only exceptions are those hypotheses involving prenatal factors that do not change within families over time—one of which, genomic imprinting, is specifically dismissed as a potential cause of the height trend by its leading advocate (Storfer, 1999, pp. 195–196). Moreover, something like a Dickens–Flynn model adapted to the height paradox would probably be even less plausible than it is in the context of IQ. In addition to the higher heritability of height, it is difficult to imagine why there would be a large gene–environment correlation in height, or a social multiplier.

The occurrence of multiple simultaneous trends also poses a great problem for selection. This is because, as plant and animal breeders know, selecting for even one trait at a time is a difficult enough proposition, but selecting for multiple traits simultaneously is far more difficult. For example, in order for selection to explain the Flynn effect, individuals with higher IQ would need to have a higher birth rate. Similarly, in order for selection to cause the height trend, taller people would need to have a higher birth rate. However, to explain simultaneous trends in both traits, those with higher birth rates would have to be both taller and higher in IQ. To explain just a few of the trends that have been observed, only those who are relatively tall and high in IQ and fast maturing and asthmatic and myopic would need to have more children. Because the traits in question are not highly correlated, very few individuals in the population would meet all of the required criteria. Moreover, in the case of IQ there is specific evidence of negative, not positive, selective pressure (Lynn, 1998a). Also, it is difficult to imagine very much selective pressure in favor of conditions like asthma, myopia, or autism. Therefore, while it is possible that selective pressures could be acting on the traits, in parallel with whatever is causing the changes over time, it is highly unlikely that selection could be the primary cause of all, or even a few, of the trends.

Measurement Invariance

In recent years, multigroup confirmatory factor analysis (MGCFA) has been applied by some researchers in an effort to better understand the nature of IQ differences observed among different groups (see Lubke, Dolan, Kelderman, & Mellenbergh, 2003, for a general description). Briefly, in this type of analysis MGCFA is used to test the proposition that observed group IQ differences are due to differences in the latent factors thought to underlie the test, such as verbal ability, spatial ability, or a general intelligence factor *g* (e.g., Carroll, 1993). When the results of the analysis are consistent with group differences solely in the latent factors, the test is said to be *measurement invariant* with respect to the groups studied. The failure to observe measurement invariance suggests that the differences are due to other factors, instead of or in addition to, differences in the latent factors.

Because heterosis would almost certainly be expected to affect the latent factors, one would initially expect measurement invariance to be observed between cohorts if heterosis were the sole cause of the Flynn effect. Wicherts et al. (2004) analyzed five data sets in which IQ test data were available for two different cohorts and found that measurement invariance was untenable in all five data sets. This finding would appear to be inconsistent with heterosis as the sole cause of the rise in IQ. It is important to note, however, that the researchers also found that several of the data sets displayed partial measurement invariance. That is, when some of the subtests were effectively taken out of the analysis, measurement invariance was found to be tenable for the remaining subtests. This suggests that the trends observed on at least some subtests are likely due to changes in the latent factors.

The failure to observe complete measurement invariance between cohorts does not allow one to preclude the possibility that heterosis is a partial or even major cause of the Flynn effect. Therefore, findings like that of Wicherts et al. (2004) provide little justification for abandoning other available opportunities to test a heterosis hypothesis, such as conclusively determining whether the Flynn effect has occurred within families or testing for intergenerational genetic changes in the frequency of heterozygotes. It must also be mentioned that unlike IQ, observed changes in traits like height, age at menarche, or myopia almost certainly could not be explained by artifacts resulting from the testing instruments.

Efforts like those of Wicherts et al., which try to understand the nature of the Flynn effect, can only complement efforts like those discussed in this article, which primarily try to get at the cause of the Flynn effect. For example, Wicherts (personal communication, May 15, 2006) cited the case of a specific vocabulary test item, *terminate*, which became much easier over time relative to other items, causing measurement invariance to be less tenable between cohorts. The likely reason for this was that a popular movie, *The Terminator*, came out between the times when the two cohorts took the test. Because exposure to popular movie titles represents an aspect of the environment that should have a large nonshared component, one would expect that gains caused by this type of effect should show up within families. Although it might be difficult to find a data set suitable for the purpose, it would be interesting to try to identify specific test items that display Flynn effects within families. Such changes cannot be due to genetic factors like heterosis, and so a heterosis hypothesis would initially predict that measurement invariance should become more tenable after removal of items that display within-family trends. One could also look for items in which the heritability markedly increases or decreases over time. In the particular case cited above, one would also expect a breakdown in the heritability of the test item, as evidenced, for example, by a change in the probability of an individual answering correctly given his or her parents' responses.

It should also be possible to test a heterosis hypothesis directly through the use of MGCFA. Specifically, one could compare the differences observed between inbred and noninbred groups with those observed between cohorts. Similarly, genetic testing of subjects may be able to provide a measure of genetic heterozygosity that could be incorporated into the analysis as a background variable to estimate the potential of heterosis to explain the IQ

trend (see Lubke et al., 2003, for an example using SES as the background variable).

Catch-Up Development

Recall that the IQ paradox poses problems for all environmental hypotheses except those involving prenatal factors, specifically shared prenatal factors. In addition, Flynn effects have been observed in very young subjects (Hanson, Smith, & Hume, 1985; Tasbihsazan, Nettelbeck, & Kirby, 1997), which likewise implicates factors operating at least very early in life, if not before birth. Findings like this make studies of the effects of prenatal environment particularly relevant to the question of whether the Flynn effect is likely environmental in origin. In this section two studies of the effects of prenatal deprivation are discussed, as well as studies of the long-term effects of certain postnatal deprivations. The results of these studies suggest that the human body has a fairly robust capacity to bounce back from obvious deprivations with no apparent long-term effects on either height or IQ. This poses a real problem for environmental hypotheses for the trends, particularly because the studies were done in populations that were in the midst of ongoing increases in both traits.

During World War II certain parts of the Netherlands experienced severe food shortages. The children born in these regions were found to have lower birth weight and smaller head circumference at birth than children born in nonfamine regions, presumably as a result of the famine conditions. When the boys born during the famine were later tested at age 18 for military conscription, no differences were found between those from famine and those from nonfamine regions with respect to either height or IQ (Stein, Susser, Saenger, & Marolla, 1972). Of note, subsequent Dutch cohorts, born after World War II, continued to witness at least 15 more years of continuous increases in both traits (Flynn, 1987; Van Wieringen, 1986). Therefore, if the subsequent postwar trends in height and IQ were environmental, one must believe that virtually all children born near the end of the war, those from famine as well as those from nonfamine regions, were significantly stunted in their growth and depressed in IQ owing to a poor environment; however, the children born in the famine regions were not more stunted or more depressed in IQ than children born in nonfamine regions. In other words, there must have been pervasive stunting and IQ depression, but an obvious prenatal deprivation like a famine did not seem to worsen it, at least with respect to the adult phenotype.

Twinning represents another type of prenatal deprivation, inasmuch as twins must share both nutrition and space. Not surprisingly, twins have been found to be shorter and lower in IQ than nontwins at early ages. However, the differences gradually dissipate during childhood, so that twins appear to completely catch up to nontwins by about the age of 7 (Wilson, 1986). Here again, it is important to emphasize that this finding of catch-up growth in twins came from a population that was in the midst of ongoing phenotypic changes. As with the Dutch famine study, to believe that the population-wide changes over time have been environmental, we must believe that virtually all children of the past suffered from a height-stunting and IQ-depressing environment but that an obvious deprivation like twinning had no additional long-term negative effect.

Relatively minor environmental insults also seem to have short-term negative effects on growth. For example, it is common for children suffering from infection or other obvious illness to stop growing temporarily. However, once the infection is conquered, the child experiences a period of unusually rapid growth, referred to as catch-up growth, which seems to entirely compensate for the period during which growth was halted (Tanner, 1990, pp. 165–171). Importantly, the phenomenon of catch-up growth has been observed in populations that subsequently went on to witness large increases in height. In fact, according to Tanner (1981, p. 72) catch-up growth was first documented in 18th-century Prussia, in a population that was much shorter than its modern descendents. Therefore, unless the height trend has been genetic, one must believe that pervasive environmental factors kept virtually all individuals of the past well below their presumed maximum genetic potential but that obvious insults like infection and identifiable diseases did not worsen the stunting, as children showed no long-term effects from these.

If the trends in height and IQ have been genetic, then findings like those above can be explained relatively easily. Genetic factors place an upper limit on human size and cognitive ability. Obvious deprivations, like famine, twinning, or infection, can temporarily deflect a child's phenotype below his or her genetic potential, but the normal environmental conditions that generally prevailed throughout the industrialized world during the 20th century were sufficient to allow most individuals to attain a phenotype reasonably close to their maximum genetic potential and to allow children who had suffered environmental insults to largely recover from them.

Within-Family Height–IQ Studies

In trying to answer the question of whether the Flynn effect and the height trend have occurred within families, the best available evidence, from birth order studies, suggests that they have not. Further evidence supporting this contention comes from studies of the association between height and IQ within families. If both the height and IQ trends were occurring within families, later born siblings would be both taller and higher in IQ than earlier born children. This would create a positive within-family correlation between the two traits. Several studies, however, have failed to find such a correlation (Jensen & Sinha, 1993).

The lack of within-family correlation between height and IQ is easily explained by a heterosis hypothesis. One need only posit separate genes influencing the two traits—that is, no pleiotropy. In this case, a sibling who inherits a disproportionately large share of the parents' height-increasing genes would be no more or less likely to inherit more IQ-increasing genes, and so there would be no within-family correlation. Admittedly, the magnitude of the expected correlation if both trends were occurring within families would likely be small, and so further research is probably warranted before a definitive conclusion is made on the issue. Nevertheless, on the basis of both the within-family correlation studies and birth order studies, it would appear rather unlikely that the trends in height and IQ have occurred within families.

Environmental Changes

There is no question that numerous aspects of the environmental conditions of life have changed during those times, and in those

places, that IQ gains have occurred. More children go to school for longer periods. Great advances have taken place in the field of medicine. The average person now has more disposable income. Diets have changed dramatically. Jobs have become more complex; indeed, the whole pace of life seems faster today than it was in earlier times. It is difficult to believe that there is no connection between these changes and the fact that IQ test scores have been rising. And it is intuitively appealing to assume that these environmental changes must be causing the changes in traits like height and IQ. This intuitive appeal might explain why investigators into trends in traits like height have been willing to largely overlook the theoretical problems for environmental hypotheses posed by the high heritability of the traits.

The problem with assuming that the environmental changes are causing the phenotypic changes, however, is that the environmental changes could just as easily be the effect, rather than the cause, of rising intelligence in the affected populations. If intelligence has really increased, one might easily expect all of the environmental changes that have been observed. A population experiencing increases in intelligence might expect to see things like rising education levels and rising income levels as a result of having a brighter, more productive workforce. Wealthier populations could afford more expensive food, and so diets would change. Brighter workers could handle more complex jobs. Separating cause from effect in this case would seem to represent an irresolvable chicken-and-egg conundrum. However, an explanation in which heterosis causes intelligence gains, which in turn lead to environmental changes, is preferable to environmental explanations for two reasons.

The first reason to prefer a heterosis hypothesis is that, unlike environmental hypotheses, it does not rely primarily on correlations of averages, or "ecological" correlations, which are the least reliable in scientific investigation. One of the best examples of this type of ecological association is that observed between height and income. There is an entire subdiscipline of history, known as *anthropometric* history, that tries to assess the quality of the environment of past generations by examining their physical characteristics, such as height. Studies in this field have identified a very close relationship between mean height changes and changes in per capita income. The following equation, from Drukker and Van Meerten (1995), provides a good example of the kind of work done in the field of anthropometric history.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{18} X_{18} + e. \quad (1)$$

In the above equation, Y is the mean adult height of a particular birth cohort. The X terms are values of the per capita income for the country in the first 18 years of the birth cohort's lives. The β coefficients and the final e term are determined empirically so that the mean height predicted by the equation best fits actual height data over the time period in question, usually several decades. Analyses using equations like this have demonstrated the capacity to produce excellent fitting curves, spanning many decades of height and income data in many populations (Komlos, 1994, 1995; Steckel & Floud, 1997). In fact, the association between height and income has proven so robust that equations like that above have even been used to estimate likely per capita income for periods in which good historical income data are missing but good height data are available (Brinkman, Drukker, & Slot, 1988). The fact that average height and per capita income have tended to rise in tandem is so well established that any explanation for the height trend must address it; that it could be a coincidence is difficult to imagine.

A heterosis hypothesis is not only capable of explaining relationships like that between height and income but able to do so in a way that overcomes a major problem with the prevailing view that the changes in height are caused by the changes in income. The problem with such an explanation is that the presumed causal relationship between height and income holds only at the ecological level; that is, the average height of the population is highly correlated over time with its average, or per capita, income. However, at the individual level this relationship breaks down, at least in genetically informed studies. As Tanner stated in the quote presented earlier, genes, not income, seem to be the primary source of variance in height among individuals at any single point in time. The monozygotic twin that is raised in the more affluent home appears to be no taller than the one reared in the less affluent home. And so, just as with the social multiplier proposed by Dickens and Flynn (2001), we are left to ask how the proposed cause of the height trend, namely income, affects the heights of millions of people on a population-wide scale whereas differences in family income seem to have no effect at the individual level. I would also add that, as with the social multiplier, a seemingly inevitable consequence of Equation 1 is that moving one's family from a high-income nation to a low-income nation will cause one's children to be shorter as a result.

A heterosis hypothesis offers a much more satisfying explanation for findings like the correlation between average height and per capita income. In this explanation, heterosis causes the trends in both height and IQ, which have largely paralleled each other (e.g., Sundet et al., 2004). The increases in IQ reflect real increases in the intelligence of the populations. Income rises as a result of the workforce being brighter and more productive. The advantage of such an explanation is that the proposed cause appears to have effects at the individual level. For example, whereas there is little evidence that high income causes higher IQ or greater height in genetically informed studies, there is abundant evidence that high IQ leads to greater productivity and higher income (Gottfredson, 1997; Hunter & Hunter, 1984). In fact, even within the same family, the higher IQ sibling is more likely to work in a professional job (Murray, 1997).

The second reason why a heterosis hypothesis is preferable to alternative environmental hypotheses is that it offers the prospect of providing a deeper explanation of the phenomena in question, one that better gets at the root causes of things. By contrast, environmental hypotheses for the various trends tend to leave basic questions unanswered. To use per capita income as an example again, even if one accepts that increases in per capita income are responsible for the height trend, one is still left with the question of why per capita income has risen and not fallen or why it has risen in some countries but not others, or the question of when it will end. Given enough money, could we increase the heights of populations indefinitely?

A heterosis hypothesis offers a much more specific explanation for things than any of the alternatives. The various trends are all caused by the same genetic phenomenon. Demographic changes cause an increase in heterozygotes, which in turn causes phenotypic changes in traits that display directional dominance. At the ecological level, one should be able to explain past trends by examining genetic differences between the older generation and

the current generation in the frequency of heterozygotes. At a more fine grained level, families that experience the largest intergenerational increases in heterozygosity should account for a larger share of the intergenerational phenotypic changes. One should also be able to predict where, and to what extent, the trends will continue into the future. For example, if children born this year are more heterozygous, on average, than children born last year, then in 20 years one should see an annual increase in all of the traits between the two cohorts when they are adults. The trends should stop when the frequency of heterozygotes levels off, such as would occur when mating becomes completely random. By extension, one should also be able to predict which countries will likely experience increases in per capita income in the future.

Although the above discussion has focused on the example of per capita income, the same arguments will likely apply to just about any environmental hypothesis for any trend, such as one in which changes in education are the presumed cause of the IQ trend (e.g., Barber, 2005; Williams, 1998) or nutrition is the cause of either the height or the IQ trend. Regardless of the specific environmental cause proposed, the high heritability of the traits means that the proposed causal relationship will probably not persist beyond the ecological level. And in the case of just about any environmental hypothesis, one will still be left with basic questions unanswered, such as why education levels have risen and not fallen or why they have risen more in some countries and at some times, or when the trends will stop, if ever.

Questions Related to Evolution

One of the basic precepts of modern evolutionary theory is that there is a reasonably close connection between the phenotype and genotype. For example, according to Darwinian evolutionary theory, morphological changes observed in the fossil record are assumed to primarily reflect genetic change. Of course, the genotype–phenotype relation need not be perfect; even monozygotic twins differ to some extent, sometimes considerably, but it is assumed that there is a reasonable limit to such environmental plasticity. However, to accept that the various changes that have occurred in human populations are environmental, one must believe that there has long been a very large disconnect between genotype and phenotype in our own species. For example, Dutch men today are 16 cm taller than their ancestors of the 1850s (Van Wieringen, 1986); this represents a gain in excess of two standard deviations. Gains of similar relative magnitude have been observed in traits such as growth rate and measured intelligence. The magnitude of the trends raises at least two puzzling questions with respect to our understanding of evolutionary theory.

The first question one must ask is that if the phenotypic changes observed in our own species are environmental in origin, then how many speciation events in the history of life on earth have been driven by similar environmental change? An analysis of fossil records by Alroy (1998) gives some perspective to the magnitude of the changes that have occurred in our own species. Alroy was testing *Cope's rule*, which states that new species are generally larger than the ancestral species from which they evolved. In analyzing a large number of speciation events among North American mammals, Alroy found that the new species were, on average, 9% larger in body mass than their putative ancestral species. Today Dutch men are about 9% taller than their ancestors of just six generations ago. Though height is not exactly the same as body mass, the fact remains that the changes that have occurred in body size in certain human populations are large enough that they might reasonably be viewed as a speciation event if one had only the human fossil record to rely on. If our species possesses so much environmental plasticity for body size, then how many other species have similar plasticity, and how many of the morphological changes observed in the fossil record have been environmental, not genetic, in origin?

A second question raised by environmental hypotheses for the trends over time is how and why did human populations acquire and maintain genes capable of enabling them to attain phenotypes so different from their realized phenotypes? For example, why did a population with an average male height of 156 cm acquire, and maintain, genes capable of making them 172 cm without ever coming close to realizing such large body size? What possible use could such genes have served that earlier population; should there not have been a selective pressure in favor of height-reducing genes? Yet if one accepts that the trend in height has been environmental, one must accept that this occurred not just in one population but in multiple populations across the world. The same question could be raised in connection with other trends, such as those seen in the growth rate and measured intelligence, and is even more puzzling in the cases of traits with no apparent advantage to fitness, such as asthma or autism.

A heterosis hypothesis not only preserves the genotype–phenotype connection that underpins modern evolutionary theory but also offers the possibility of providing new insights into the process of evolution. For example, heterosis represents an ideal mechanism for speciation. As is shown in the next section, heterosis is theoretically capable of causing large genetically based phenotypic changes in multiple polygenic traits. Because the process can rely entirely on the redistribution of existing genes, the pace of change relies solely on the pace of demographic change, which, in theory, could occur in just one generation. A population that has undergone large rapid changes in multiple traits might very well find itself better adapted to some new ecological niche. At this point, an entirely new set of selective pressures could then preserve and extend the favorable changes, while winnowing out unfavorable changes.

Heterosis as a mechanism for speciation could help explain several findings from the field of evolutionary biology, findings that have otherwise puzzled researchers. For example, it could explain the aforementioned Cope's rule. If body size displays positive directional dominance in most species, then speciation events caused by heterosis would more likely involve an increase, rather than a decrease, in body size. Also, because it is capable of causing rapid change, heterosis as a mechanism for speciation could explain the general lack of intermediate forms in the fossil record (Gould & Eldredge, 1977).

Last, a heterosis hypothesis is consistent with the neutral theory of genetic mutations (Kimura, 1983), which states that most mutations appear to be neither advantageous nor disadvantageous with respect to fitness. As will be seen in Part 2 of this article, a heterosis hypothesis posits dominant genes influencing traits in one direction and recessive genes influencing the traits in the opposite direction. In such a framework, the phenotype, and hence fitness, is determined by the balance of dominant to recessive alleles in the population. No individual allele confers either fitness

or lack of fitness; too many dominants can be just as detrimental to fitness as too many recessives, as an excess of either will cause the phenotype to deviate from optimal values. Because individual alleles do not necessarily increase or decrease fitness, individual mutations would not be expected to have much influence on fitness, either positive or negative.

The Nature of Genetic Dominance

Decades ago, Fisher (1928, 1931) proposed an explanation for the occurrence of genetic dominance, including directional dominance. He suggested that a prolonged period of selective pressure on a trait in one direction would bring about positive directional dominance in that trait. On the basis of the results of inbreeding studies, discussed in more detail below, IQ appears to display positive directional dominance. That is, more of the genes that increase IQ are likely dominant, whereas more of those that lead to lower IQ are likely recessive. Therefore, Fisher's theory would suggest there was a prolonged selective pressure in favor of the high-IQ phenotype in the recent evolutionary history of our species. If, as Fisher proposed, directional dominance is frequently caused by directional selective pressure, it would pose a problem for heterosis as a cause of some, but not all, of the observed trends. Although one can imagine selective pressures in favor of height, IQ, or growth rate, it is difficult to believe that there could have been prolonged selection in favor of the genes that predispose individuals to conditions like asthma, myopia, or autism.

However, from its inception, Fisher's theory on the evolution of dominance has been questioned. One alternative, advanced in Fisher's time by Wright (1934) and supported in slightly different form by subsequent researchers (Kacser & Burns, 1981), argues that genetic dominance is an inevitable consequence of the pathway by which genes ultimately influence the phenotype. In this so-called physiological explanation, the genes influence the phenotype by coding for various proteins, or gene products, such as enzymes, hormones, immunoglobins, histones, or neurotransmitters. Because the biochemical reactions that eventually influence the phenotype are complex, inheriting two copies of a dominant allele instead of one copy at any given locus rarely alters the ultimate effect of the genes on the phenotype; other limiting factors usually come into play.

If the physiological explanation for genetic dominance is correct, several of the trends that have occurred make a lot of sense in the context of a heterosis hypothesis. The physiological explanation suggests that directional dominance is likely intrinsic to the production by the genes of all proteins, which would mean increases in the levels of all proteins if mating patterns become more random. At least some of the trends that have been observed appear to be consistent with increases in various gene products. For example, the increases in height and growth rate are consistent with a rise in growth hormones. The rise in asthma is consistent with an increase in the body's production of the immunoglobin IgE, which is produced in response to allergens. The development of myopia is consistent with an increase in the production of the neurotransmitter acetylcholine, given that atropine, which inhibits the production of acetylcholine, has been shown to halt the development of myopia (Shih et al., 2001). The rise in autism might be explained by a rise in serotonin, which has been shown to be elevated in autistic individuals (Cook & Leventhal, 1996). The rise in IQ might be explained by increases in brain growth factors, hormones like testosterone, or any of a number of neurotransmitters. Fortunately, protein levels can be measured, so that if an across-the-board increase in all proteins is in fact occurring, it should be possible to detect it. Although measuring such increases would not prove that heterosis is responsible, it would be one more piece of the puzzle that fits neatly within the context of a heterosis hypothesis.

Part 2: IQ Changes Over Time—A Basic Heterosis Model

In addition to being the most plausible explanation for the Flynn effect, a heterosis hypothesis is also by far the most testable. At the individual and family level, one can determine whether the IQ trends have occurred within families or instead between families, as predicted by a heterosis hypothesis. One can look for associations between the various traits that have changed and genetic heterozygosity. One could even look for associations between genetic heterozygosity and the levels of specific proteins.

At the ecological level, the hypothesis predicts that an increase in genetic heterozygosity should be observed in populations that have experienced IQ gains. Such a trend has recently been observed in Iceland, whose population has been the subject of extensive genetic testing (Helgason, Yngvadottir, Hrafnkelsson, Gulcher, & Stefansson, 2005). Researchers there were testing the assumption of random mating in the population because any deviation from random mating would influence the statistical analysis of gene association studies. They found that mating was not random; instead, the frequencies of the genes studied differed significantly among the various regions of the country. Of importance for the present discussion, the regional genetic differences were most pronounced among the oldest cohorts and lowest among the most recent cohorts. That is, there had been a trend toward more random mating across the country throughout the 20th century, with an accompanying increase in the frequency of heterozygotes. The researchers attributed the finding to demographic changes like urbanization and suggested that the situation in Iceland was probably typical of populations in other industrialized countries.

As discussed in the previous section, a trend toward more random mating, with its consequent increase in heterozygotes, will tend to cause population-wide changes in all traits that display directional dominance. On the basis of the results of inbreeding studies, discussed in more detail below, IQ does appear to display positive directional dominance. Therefore, if the finding of genetic change in Iceland is widely replicated in other populations, it would become not only possible but probable that some portion of the Flynn effect has been due to heterosis. It is interesting to note that although heterosis has been mentioned as a possible partial cause of the Flynn effect by many researchers, no effort has yet been made to estimate its potential contribution, at least not in any quantitatively rigorous way.

In the remainder of this article, stochastic modeling is used in an effort to begin the process of developing a quantitatively rigorous method for estimating the potential contribution of heterosis in the IQ trend, as well as the trends seen in other traits. Although the exercises presented below use genetic data from a real population, the demographic changes they posit are hypothetical. The present effort must therefore be viewed as only a hypothetical example of

how to carry out this type of analysis. Despite its partly speculative nature, however, the analysis includes the basic elements that will likely be required of any serious effort to better understand the phenomenon of heterosis. Although it contains a number of simplifications, a preliminary analysis of the type described below, using real data, should be sufficient to obtain a reasonable estimate of the potential contribution of heterosis. On the basis of such an estimate, one could then decide whether to proceed with more complex models or to abandon heterosis as unlikely to ever explain much of the trend.

Structural Genetic Assumptions

The magnitude of the effect that a genetic process like heterosis has on a trait will depend on the characteristics of the genes that influence that trait. However, the specific genes that influence IQ have yet to be identified. In a situation like this where the potential causal factors cannot be observed directly, the only available alternative is to try to develop a model of the underlying factors, in this case the genes. Genetic models with different characteristics will predict different results, which can then be compared with observed findings. The idea is to find a model that best fits the most findings and then estimate the likely effect of heterosis on the assumption that the resulting model accurately reflects the genetic structure of IQ.

In developing a model of the genes that influence IQ, the only prerequisite that has been imposed here is that the genes must display at least some directional dominance. This is essential not only to explain Flynn effects as a result of heterosis but also to explain the occurrence of inbreeding depression in IQ. In all other respects the goal has been to keep the presumed genetic structure as simple as possible in this initial effort.

The modeling effort begins by assuming that IQ is influenced by the genes at L loci. It is further assumed that each locus has an equal effect on IQ, all loci are autosomal, and there are no epistatic interactions among the genes at different loci. It is also assumed that the genes are the only factors that influence IQ. That is, factors like measurement error and environmental influences are being ignored for the time being.

Recall that there must be dominant genes that tend to increase IQ and recessive genes that reduce IQ. For simplicity, then, it is assumed in the model that there are just two alleles at each IQ locus, one dominant allele and one recessive allele. With just two alleles at each locus, there are only three possible genotypes that an individual can inherit at a given locus. These genotypes are listed in Table 1, along with their presumed effect on measured intelligence.

In the framework shown in the table, D is the dominance level; it represents the effect on intelligence of inheriting a heterozygous genotype at a particular locus. A value for D of .5, exactly midway between 0 and 1.0 (the effects of the corresponding homozygous genotypes), would represent perfectly additive gene action. D must be at least somewhat greater than .5 to cause mean IQ increases as a result of heterosis, as well as to cause inbreeding depression. Although theoretically possible, it will be assumed here that D cannot exceed 1.0, which represents gene action in which dominance is complete.

Table 1
Framework for the Genetic Structure of IQ

Genotype	Effect on intelligence
Dominant-dominant	+1
Dominant-recessive	+D
Recessive-recessive	0

Note. D = dominance level.

Given the assumption that the genes are the only influence on IQ, the above framework makes it possible to calculate an intelligence level for any individual based solely on that person's complete genotype, that is, the genotypes he or she has inherited at all L loci. This intelligence level is given by Equation 2. In the equation, L_{DD} is the number of loci at which the individual has inherited two dominant alleles, L_{DR} is the number of loci at which the individual has inherited one dominant and one recessive allele, and L_{RR} is the number of loci at which the individual has inherited two recessive alleles. L_{DD}, L_{DR}, and L_{RR} will therefore sum to L, the total number of loci. Each L term is multiplied by its corresponding effect on intelligence, +1.0, +D, or 0, from Table 1.

$$\text{Intelligence} = (1)L_{DD} + (D)L_{DR} + (0)L_{RR}. \quad (2)$$

The term *intelligence* used in Equation 2 and Table 1 is intended in the context of this article to mean the abilities measured by IQ tests. More accurately, it can be thought of as a measure of the individual's genetic predisposition to develop IQ-type cognitive abilities. In more complex models this intelligence level could be deflected by things like environmental factors and measurement error. It should be noted that the intelligence level calculated for an individual using Equation 2 has no dimensions. Because intelligence is not measured on any absolute scale, an individual's intelligence level can be quantified only relative to others. It is also important to emphasize that the intelligence level calculated using Equation 2 will not be in an IQ metric but is more akin to a raw IQ test score. In theory, it can range from 0 (all double recessive loci) to L (all double dominant loci), assuming again that D cannot exceed 1.0. To put the value calculated in Equation 2 into an IQ metric, intelligence levels would first have to be calculated for all individuals in the population. According to the distribution of these levels, each individual's score would then have to be normed, so as to give the distribution a mean of 100 and standard deviation of 15.

Population Genetic Assumptions

Recall that in order for heterosis to cause changes in a trait there must be a deviation from random mating in the population and a demographic trend toward more random mating. A good example of how mating can deviate from random in a population comes from a study conducted in the Parma Valley region of Italy in the late 1950s by Cavalli-Sforza, Moroni, and Zei (2004). The researchers studied the frequencies of the alleles at three blood-group loci in 74 different villages in the valley. What they found was that the population was not genetically homogenous; instead, allele frequencies varied, sometimes markedly, among the villages. For example, frequencies of the M allele, at the MN locus, ranged from a low of .353 in one village to more than double that, .753, in the village where the allele was most common. It is this kind of nonrandom mating that gives the population the latent potential for future phenotypic changes as a result of heterosis.

It is a simple matter to calculate the excess of homozygotes and the deficit of heterozygotes at the MN locus brought about by the deviation from random mating in the Parma Valley. There are two alleles at the MN locus, the M allele and the N allele. With only two alleles, there are only three possible genotypes that an individual can inherit at this locus: M-M, N-N, or M-N. Because both alleles are dominant, it is easy to distinguish those with each genotype, on the basis of the presence of either the M or the N protein, or both, in the blood. The Italian researchers genotyped 2,815 individuals and counted 882, 628, and 1,305 individuals with the M-M, N-N, and M-N genotypes, respectively. Therefore, out of the 5,630 alleles examined, .545 of them ([2 × 882 + 1,305]/5,630) were M and .455 were N. Therefore, if mating were random in the population, one would expect 29.7% of the population to be M-M ($.545^2$), 20.7% to be N-N ($.455^2$), and 49.6% to be M-N (2 × .545 × .455), according to the Hardy–Weinberg equation. By contrast, the actual percentages in the population were 31.3%, 22.3%, and 46.4% for the M-M, N-N, and M-N genotypes, respectively. Therefore, the deficit between the expected and observed frequency of heterozygotes (M-N) was about 3.2% of the population; the excesses of M-M and N-N individuals were each 1.6% of the population.

Simulating Demographic Change

The deviation from random mating in the Parma Valley represents the potential for a change in the ratio of heterozygotes to homozygotes if mating were to become more random, which in turn would affect all traits that display directional dominance. In this section an effort is made to assess the likely genetic effects of a hypothetical demographic change in the population. Specifically, we will suppose that the 74 villages studied eventually amalgamate, or coalesce, into a single, randomly mating population. In the model, therefore, two hypothetical generations of the population are posited: an earlier generation, in which mating is less than random, and a later generation, in which mating is completely random. The real alleles, M and N, will be used in some of the simulations as a kind of template for the hypothetical dominant and recessive alleles that influence IQ. This does not mean that we are assuming that the M-N locus is involved in IQ but simply that the hypothetical genes that influence IQ deviate from the expectations of random mating in the earlier generation to a similar extent as these known alleles did back in the 1950s.

Recall that we are positing one dominant and one recessive allele at each of L loci. For simplicity, the model will assume that the ratio of dominant to recessive alleles in the overall population is the same at all loci that influence IQ. The parameter R will be used to denote the frequency of the recessive allele at each IQ locus. The frequencies of the dominant alleles at all loci will therefore all be $1 - R$. Setting the parameter R immediately determines the probabilities of inheriting each of the three possible genotypes at a given locus in the final, randomly mating generation. According to the Hardy–Weinberg equation, the probability of inheriting two recessives in the final generation will be R^2, the probability of inheriting two dominants will be $(1 - R)^2$, and the probability of being heterozygous will be $2R(1 - R)$.

The earlier, nonrandomly mating generation must have fewer heterozygotes and more homozygotes than the eventual randomly mating generation. The parameter ΔHe will be used to denote the deficit of heterozygotes in the initial generation. Therefore, the probability of being heterozygous in the initial generation will be $2R(1 - R) - \Delta He$. This deficit of heterozygotes must be offset by excesses of the two homozygous genotypes. Because each heterozygote is composed of one dominant and one recessive allele, the excess of double dominants must be the same as the excess of double recessives. Therefore, the probability of inheriting two recessive alleles in the earlier generation will be $R^2 + (1/2)\Delta He$, and the probability of inheriting two dominants will be $(1 - R)^2 + (1/2)\Delta He$.

If one were to use the M allele in the Parma Valley as the template for the recessive allele at the typical IQ locus, one would use model parameters of .545 for R and .032 for ΔHe (3.2% of the population). Similarly, if N is the template for the typical recessive allele, one would use an R value of .455 and a ΔHe value of .032. Several of the models presented below use values of .03 for ΔHe and .5 for R, which gives our hypothetical dominant and recessive alleles characteristics similar to the alleles at the MN locus. Because actual IQ genes have yet to be identified, one could reasonably choose alleles at any locus as a template for IQ genes.

Stochastic Modeling of Heterosis

Setting the model parameters R and ΔHe determines the probabilities of inheriting a given genotype in both the earlier and later generations. Again, we are assuming the probabilities are the same at all loci. With these probabilities, it is a simple matter to stochastically generate genotypes at all L loci for a large number of hypothetical individuals in each generation. Once complete genotypes are generated and after the model parameter D is specified, an intelligence level can then be calculated for each individual using Equation 2 above. The final step is to calculate the mean and standard deviation of the distribution of these intelligence levels in both the earlier and the later generations. By comparing the distributions in the two generations, one can estimate the expected effect of the demographic change on IQ.

Table 2 lists the results of the simulation of demographic change for models with different combinations of the four major parameters. These simulations, as well as others discussed in later sections, were carried out using computer programs I wrote in the Visual Basic language that is built into Microsoft Excel; they are available to interested readers upon request. The first four columns of Table 2 list the major model parameters: L, the number of loci; D, the dominance level; R, the frequency of the recessive allele at each locus; and ΔHe, the deficit of heterozygotes in the earlier, non-randomly-mating generation. The next four columns list the four primary results: the means and standard deviations of the distributions of intelligence in the earlier generation (μ_1, σ_1) and the later generation (μ_2, σ_2). The last three columns list additional results derived from the four primary results. The first derived result is the Flynn effect, defined as $(\mu_2 - \mu_1)/\sigma_1$. The values in parentheses are simply the Flynn effect, $(\mu_2 - \mu_1)/\sigma_1$, multiplied by 15 so as to be put into an IQ metric. The second derived result is a measure of the extent to which the standard deviation of IQ would be expected to change as a result of the demographic change, defined as $(\sigma_2 - \sigma_1)/\sigma_1$. So, for example, the first model listed in Table 2 ($L = 100$, $D = .6$, $R = .5$, $\Delta He = .03$) predicts a Flynn effect of 0.08 standard deviations, or 1.2 IQ points, and a 3% decline in the standard deviation of IQ, if mating were to

Table 2
Results of the Simulation of the Effect of Demographic Changes on Intelligence

L	D	R	ΔHe	μ_1	σ_1	μ_2	σ_2	$\frac{(\mu_2 - \mu_1)}{\sigma_1}$	$\frac{(\sigma_2 - \sigma_1)}{\sigma_1}$	$\frac{\sigma_2}{\mu_2}$
100	**.6**	.5	.03	54.7	3.67	55.0	3.57	0.08 (1.2)	−.03	.07
100	**.8**	.5	.03	64.1	3.94	65.0	3.84	0.23 (3.4)	−.02	.06
100	**1.0**	.5	.03	73.5	4.41	75.0	4.33	0.34 (5.1)	−.02	.06
100	.8	.5	**.02**	64.4	3.90	65.0	3.84	0.15 (2.3)	−.02	.06
100	.8	.5	**.03**	64.1	3.94	65.0	3.84	0.23 (3.4)	−.02	.06
100	.8	.5	**.04**	63.8	3.97	65.0	3.84	0.30 (4.5)	−.03	.06
100	.8	**.4**	.03	73.5	3.50	74.4	3.37	0.26 (3.9)	−.04	.05
100	.8	**.5**	.03	64.1	3.94	65.0	3.84	0.23 (3.4)	−.02	.06
100	.8	**.6**	.03	53.5	4.20	54.4	4.14	0.21 (3.2)	−.02	.08
50	.8	.5	.03	32.0	2.78	32.5	2.72	0.16 (2.4)	−.02	**.09**
75	.8	.5	.03	48.1	3.41	48.8	3.33	0.20 (3.0)	−.02	**.07**
100	.8	.5	.03	64.1	3.94	65.0	3.84	0.23 (3.4)	−.02	**.06**

Note. Boldface indicates the parameter values that were varied while other parameters were held constant. L = number of loci; D = dominance level; R = frequency of recessive alleles; ΔHe = deficit of heterozygotes prior to demographic trend; μ_1 = mean intelligence level of population prior to demographic trend; σ_1 = standard deviation of intelligence prior to demographic trend; μ_2 = mean intelligence level of population after demographic trend; σ_2 = standard deviation of intelligence after demographic trend; $(\mu_2 - \mu_1)/\sigma_1$ = Flynn effect in standard deviation units (values in parentheses show Flynn effect in IQ points); $(\sigma_2 - \sigma_1)/\sigma_1$ = measure of the change over time in standard deviation; σ_2/μ_2 = ratio of standard deviation to mean after demographic trend.

become completely random in the population. The third derived result is the ratio of the standard deviation to the mean in the later generation, σ_2/μ_2; the reason for including this last result will become clear below.

To help illustrate how the model behaves, each of the four major model parameters is varied in turn while the other three are held constant. The first two results from Table 2 are not surprising. From the first three rows, one can see that positing more directional dominance (greater D) leads to greater Flynn effects, all other things being equal. This is not unexpected given that directional dominance is an essential component of the process of heterosis. Similarly, the second three rows show that positing a greater deviation from random mating in the initial generation (greater ΔHe) also leads to greater Flynn effects, all other things being equal. As with directional dominance, an initial deviation from random mating is an essential component of heterosis, and so positing a greater initial deviation naturally leads to larger Flynn effects when mating eventually becomes random.

The effect of varying R, the frequency of the recessive alleles, is somewhat more interesting. In general, reducing R has two effects. First, it results in larger Flynn effects, all other things being equal. Second, it results in larger declines in the standard deviation of IQ. This latter finding is particularly interesting in light of studies suggesting that the increases in mean IQ have been accompanied by declines in IQ variance (Colom, Lluis-Font, & Andres-Pueyo, 2005; Sundet et al., 2004; Teasdale & Owen, 2000; but see also the exchange between Dickens & Flynn, 2002, and Rowe & Rodgers, 2002). A finding of declining IQ variance over time could be explained within the context of the heterosis hypothesis by positing a genetic structure of IQ in which the recessive alleles are relatively rare (lower R) and the dominant alleles are relatively common. It should be emphasized that models with different parameters could just as easily explain stable IQ variance or even an increase in variance, although this last outcome would likely require positing dominant alleles that are extremely rare. It is also worth noting here that the trend in height has sometimes been accompanied by an increase in height variance (Sundet et al., 2004; Van Wieringen, 1986).

Perhaps the most interesting finding from Table 2 is the fact that increasing the number of loci, L, leads to larger Flynn effects, all other things being equal. It is interesting to note that Dickens and Flynn (2001) posited a similar effect. The only difference is that in their model the increased Flynn effects are caused by positing more environmental factors that are correlated with the genes, as opposed to more actual genes. In both their model and the present model, however, what causes the effect is the law of large numbers, which states that the variance will decline relative to the mean whenever an outcome is determined by the sum of more mutually independent factors. One can see from the last three rows of the final column of Table 2 that as L increases, the ratio σ/μ declines, so that the same percentage increase in the mean represents a relatively larger gain over time when measured relative to the standard deviation.

The fact that positing more loci or positing recessives that are rarer leads to larger expected Flynn effects means that there is no theoretical upper limit to the potential effect of heterosis. As long as there is at least some directional dominance, even a very small trend toward more random mating could cause indefinitely large phenotypic changes, at least in theory, provided there is a large enough number of loci, or recessives that are rare enough, influencing the trait in question. As will be seen below, however, the need to comport with other relevant findings, such as inbreeding study results, will effectively constrain the potential contribution of heterosis in the trends over time.

It is important to point out that the process posited above did not involve the systematic addition or subtraction of any genes from the overall gene pool of the population. There are the same numbers of dominant and recessive alleles in the later generation as in the earlier generation. The same alleles have merely been redistributed in a more random manner over a wider area. As mentioned earlier, because heterosis uses existing genes, the rate at which the phenotypic changes can occur is limited only by the

pace of the demographic changes, which, in theory, could occur in just one generation.

Simulating Inbreeding

As mentioned several times above, any compelling explanation for the Flynn effect must make sense not only at the ecological level but also at the more fine grained level of the individual or the family. In the context of the heterosis hypothesis this means, among other things, that one should observe the effects of differences in genetic heterozygosity among individuals within the same generation. Moreover, the magnitude of the within-generation effect should be consistent with the magnitude of the effect between generations.

At present, the best evidence of the effect of differences in heterozygosity on IQ comes from consanguinity, or inbreeding, studies. In these studies, the offspring of parents who are of some known relation, like cousins, are compared in IQ with the offspring of parents who are of no known relation. Differences in IQ between the inbred and noninbred children are presumed to result from the greater genetic homozygosity in the inbred group, caused by their parents' likely genetic similarity, combined with directional dominance in the genes that influence IQ. For example, a large study of children in postwar Japan found the offspring of first cousins to be approximately 3 points (0.2 standard deviations) lower in IQ than children of unrelated parents, after controlling for an array of potentially confounding environmental factors (Schull & Neel, 1965; see also Jensen, 1983, for a review of other studies). Additionally, the study found no significant difference in the variance of IQ between the inbred children and the noninbred children. The authors reported similar inbreeding depression for height.

The effect of inbreeding on IQ will depend on its genetic structure, and any model of the genes that influence IQ must predict inbreeding depression that is consistent with observed findings. In this section, models like those used above in the simulation of demographic change are used to estimate the likely magnitude of the expected inbreeding depression for various models, each with a different set of major parameters.

To simplify the modeling of inbreeding, it will first be assumed that we are dealing with a single population in which mating is mostly random, except for the consanguineous unions themselves. This assumption would be appropriate, for example, if one were conducting the inbreeding study within just one of the 74 villages studied in the Parma Valley, rather than taking subjects from a number of different villages. This simplification limits the number of required model parameters to only three: L, the number of loci; D, the dominance level; and R, the frequency of the recessive allele at each locus. The next step is to imagine two groups within the population: the offspring of randomly mating parents and the offspring of parents who are first cousins. The former group will be referred to as the *outbred group* and the latter will be called the *inbred group*.

Estimating the distribution of intelligence in the outbred group is done in precisely the same way as the final randomly mating population in the previous heterosis simulations. Because the parents of the outbred group are assumed to mate at random, the probability that an individual in the outbred group will inherit two recessive alleles at any locus is simply R^2, the probability of inheriting two dominants is $(1 - R)^2$, and the probability of being heterozygous is $2R(1 - R)$. On the basis of these probabilities, genotypes at all loci for a large number of individuals in the outbred group are stochastically generated. On the basis of their genotypes, an intelligence level is calculated for each individual using Equation 2. The mean and standard deviation of the distribution of intelligence in the outbred group are then calculated. These are listed in Table 3 as μ_{out} and σ_{out}. Notice that for each set of model parameters, the distribution of intelligence in the outbred group (μ_{out}, σ_{out}) is the same as the final randomly mating generation (μ_2, σ_2) of the previous section; this is because in both cases mating is assumed to be random.

It should be noted that in a more precise model, the outbred group technically should contain a slight excess of heterozygotes and deficit of homozygotes, in order to offset the excess of homozygotes and deficit of heterozygotes in the inbred group. However, this effect should not be very large as long as the cousin matings make up a relatively small percentage of all unions. For simplicity, the effect will be ignored in the current modeling effort.

As mentioned above, the inbred group will contain more homozygotes and fewer heterozygotes, on average, than the outbred group. According to genetic theory, the offspring of first cousins

Table 3
Results of the Simulation of First Cousin Mating

L	D	R	μ_{out}	σ_{out}	μ_{in}	σ_{in}	$\frac{(\mu_{in} - \mu_{out})}{\sigma_{out}}$	$\frac{(\sigma_{in} - \sigma_{out})}{\sigma_{out}}$
100	**.6**	.5	55.0	3.57	54.7	3.68	−0.09 (−1.3)	.03
100	**.8**	.5	65.0	3.84	64.1	3.94	−0.24 (−3.7)	.03
100	**.0**	.5	75.0	4.33	73.4	4.42	−0.36 (−5.4)	.02
100	.8	**.4**	74.4	3.37	73.5	3.51	−0.27 (−4.0)	.04
100	.8	**.5**	65.0	3.84	64.1	3.94	−0.24 (−3.7)	.03
100	.8	**.6**	54.4	4.13	53.5	4.20	−0.22 (−3.2)	.02
50	.8	.5	32.5	2.71	32.0	2.78	−0.17 (−2.6)	.03
75	.8	.5	48.8	3.33	48.0	3.41	−0.21 (−3.2)	.03
100	.8	.5	65.0	3.84	64.1	3.94	−0.24 (−3.7)	.03

Note. Boldface indicates the parameter values that were varied while other parameters were held constant. L = number of loci; D = dominance level; R = frequency of recessive alleles; μ_{out} = mean intelligence level of outbred group; σ_{out} = standard deviation of intelligence of outbred group; μ_{in} = mean intelligence level of inbred group; σ_{in} = standard deviation of intelligence in inbred group; $(\mu_{in} - \mu_{out})/\sigma_{out}$ = inbreeding depression in standard deviation units (values in parentheses show inbreeding depression in IQ metric); $(\sigma_{in} - \sigma_{out})/\sigma_{out}$ = measure of the effect of inbreeding on the standard deviation.

will be homozygous by descent at 1/16 of all loci, on average, owing to their parents' common ancestry (see, e.g., Lynch & Walsh, 1998, p. 142). To simulate this, the modeling process starts with the genotypes already generated above for those in the outbred group. Every locus in each outbred individual is then assigned a 1-in-16 probability of being homozygous by descent. Those loci stochastically chosen to be homozygous by descent are then "forced" to become homozygous by making the second allele the same as the first, if the two alleles are not already the same. After the genotypes of those in the outbred group are altered in this way, they now reflect the likely genotypes of those in the inbred group. It is then a simple matter to recalculate intelligence levels according to these newly altered genotypes and calculate the mean and standard deviation of the resulting levels. These are listed in Table 3 as μ_{in} and σ_{in}, the mean and standard deviation of the inbred group.

Table 3 lists the results of the simulation of first-cousin mating for models with the same sets of parameters (except ΔHe) as were used in Table 2 for the simulation of demographic change. The first three columns list the major model parameters: *L*, the number of loci; *D*, the dominance level; and *R*, the frequency of the recessive allele at each locus. The next four columns list the primary simulation results: the expected means and standard deviations of the outbred group (μ_{out}, σ_{out}) and the inbred group (μ_{in}, σ_{in}). The last two columns list two additional results derived from the primary results. The first of these is the inbreeding depression, defined as $(\mu_{in} - \mu_{out})/\sigma_{out}$. As before, the values in parentheses are the inbreeding depression multiplied by 15 so as to be in an IQ metric. The second derived result is a measure of the expected difference in IQ variance between the inbred and outbred groups, $(\sigma_{in} - \sigma_{out})/\sigma_{out}$. So, for example, the first model in Table 3 ($L = 100$, $D = .6$, $R = .5$) would predict that the distribution of IQ in the offspring of first cousins will be 1.3 points (0.09 standard deviations) lower and have a standard deviation that is 3% greater than the outbred group.

The results in Table 3 of the simulation of inbreeding are essentially a mirror image of the simulation of demographic changes from Table 2. This is not surprising because inbreeding and heterosis are both manifestations of the same process, albeit working in different directions. Whereas increasing the dominance level *D* led to greater Flynn effects, it also leads to greater inbreeding depression, all other things being equal. Reducing *R* also leads to greater inbreeding depression. In addition, reducing *R* causes greater IQ variance in the inbred group relative to the outbred group. Last, increasing *L* leads to larger inbreeding depression.

With only three major model parameters, it is possible to display the results of the simulation of inbreeding for models with a wide array of different parameters. Figure 1 displays the expected effects of first-cousin mating for different models of the genes that influence IQ. In each of the three graphs, the dominance level *D* is fixed at one of three different values: 1.0, .8, or .6. The generally horizontal lines represent lines of equal *R* values, and the generally vertical lines are lines of equal *L*. The results along the horizontal axis are the inbreeding depression values, $(\mu_{in} - \mu_{out})/\sigma_{out}$ in standard deviation units. The results along the vertical axis give the relative change in the standard deviation resulting from inbreeding $(\sigma_{out} - \sigma_{in})/\sigma_{out}$. The graphs in Figure 1 can be used in two different ways. One can go into the graphs with model parameters *L*, *D*, and *R* and find the expected effects of inbreeding on both the mean of IQ (horizontal axis) and the standard deviation of IQ (vertical axis). Or one can go into the graphs with the results of an inbreeding study and find different sets of model parameters that are consistent with those results, each for a different value of *D*.

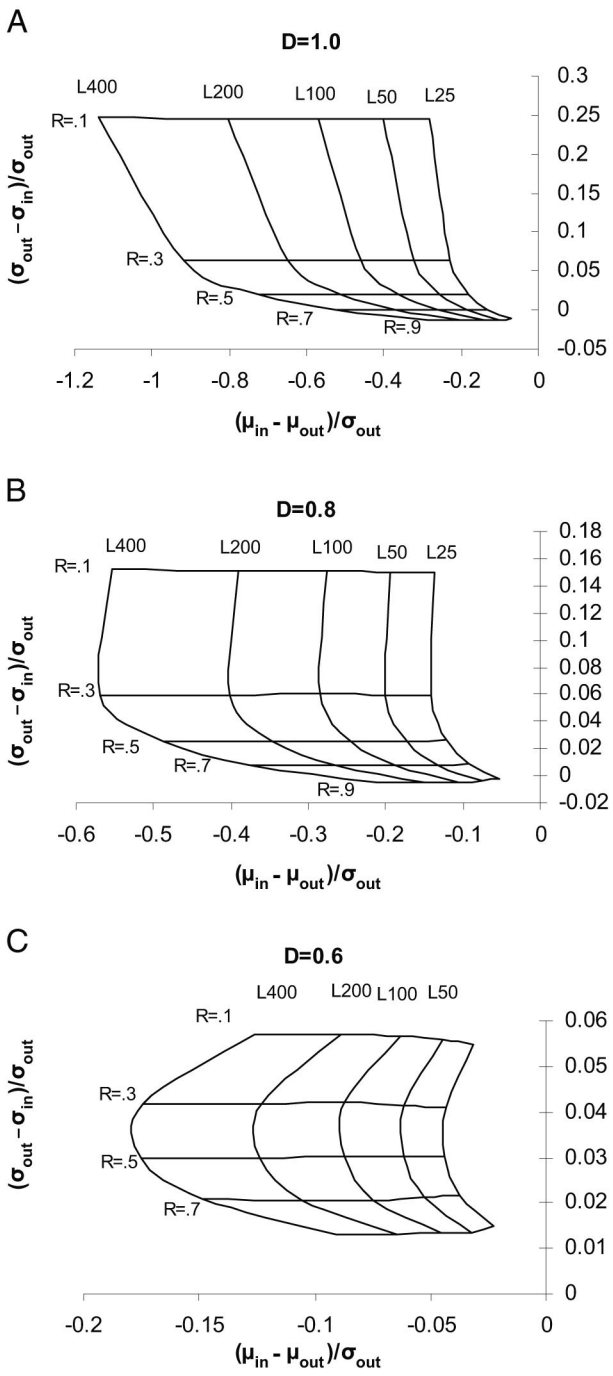

Figure 1. Inbreeding simulation results, with *D* fixed at 1.0 (A), .8 (B), and .6 (C). *D* = dominance level; *L* = number of loci; *R* = frequency of recessive alleles; μ_{in} = mean intelligence level of inbred group; σ_{in} = standard deviation of intelligence in inbred group; μ_{out} = mean intelligence level of outbred group; σ_{out} = standard deviation of intelligence of outbred group; $(\mu_{in} - \mu_{out})/\sigma_{out}$ = inbreeding depression in standard deviation units; $(\sigma_{out} - \sigma_{in})/\sigma_{out}$ = measure of the effect of inbreeding on the standard deviation.

Combining Demographic and Inbreeding Simulations

As was seen in the simulations of demographic change above, devising genetic models capable of accounting for very large Flynn effects is relatively easy, as long as the models are not constrained by any other requirements; for example, one can simply posit a very large number of loci influencing IQ. However, because the effects of demographic change tend to mirror the effects of inbreeding, genetic models that predict very large Flynn effects as the result of heterosis also tend to predict very large inbreeding depression. Therefore, the need to comport with observed inbreeding study results greatly limits the universe of plausible genetic models and thereby limits the potential contribution of heterosis in the IQ trend.

One can imagine how this balancing act between Flynn effects and inbreeding results might play out in the population of the Parma Valley. Let us suppose that researchers were to return to the Parma Valley today and find that mating was now random among the 74 villages that were studied back in the 1950s, just as was assumed in the simulations of demographic change above. In devising a model of IQ genes, one could reasonably justify using model parameters of .03 for ΔHe and .5 for R, as these resemble the characteristics of real alleles such as those at the MN locus. This leaves only the parameters D (the dominance level) and L (the number of loci) to be specified.

The main constraint on the values that D and L can take on is the need to comport with inbreeding study results. Recall that in the Japanese study cited earlier, the offspring of first cousins were approximately 3 IQ points (0.2 standard deviations) lower in IQ than those whose parents were of no known relation. Table 4 lists the results of inbreeding simulations for five genetic models specifically chosen to comport with the Japanese study results. In all five models, the value of R, the recessive allele frequency, was set to .5 so as to resemble alleles at the MN locus. The values for the dominance level, D, are allowed to vary from .6 to 1.0. Last, the value of L in each model was determined through an iterative process, so as to predict an inbreeding depression of 3 IQ points in the offspring of first cousins. Note that as the dominance level D increases, one needs fewer loci to predict the same amount of inbreeding depression. Of note, the expected increase in the standard deviation of the inbred group is less than 5% in all of the models and so comports reasonably well with the finding of no significant difference in IQ variance between inbred and outbred children in the Japanese study. All five genetic models in Table 4, therefore, can be said to comport reasonably well with the Japanese inbreeding study.

The next step is to take the five models from Table 4 and use each of them in the simulation of demographic change in the Parma Valley; one need only add the additional model parameter ΔHe, which is set at .03 on the basis of the alleles at the MN locus. The results of these simulations are shown in Table 5. Note that for all five models the expected changes in IQ over time are similar. All five simulations predict an increase in IQ of about 3 points (0.2 standard deviations) and a slight decline, of less than 4%, in the standard deviation of IQ. These results suggest that if one were to return to the Parma Valley today and find that mating had become random, one could reasonably account for 3 points of IQ gains since the 1950s as the result of heterosis. In doing so, one would be positing genetic IQ models that comport with inbreeding study findings, and one would only be positing changes in the distribution of the hypothetical IQ genes that are similar to changes observed in actual genes. Looked at in another way, the above analysis suggests that in the late 1950s, one could say that the 74 villages could have reasonably possessed the latent potential for about 3 points of future IQ gains as a result of heterosis.

Several factors make it difficult to extrapolate the simulated results above to estimate the actual potential contribution of heterosis in the Flynn effect. First, the population of the Parma Valley constituted less than 1% of the entire population of Italy in the 1950s. The most distant villages in the area studied were only about 70 km apart, but certainly many people have migrated greater distances during the 20th century. Because all of Italy likely contained greater genetic variability than 74 neighboring villages in a single valley, the latent potential of the nation as a whole was likely greater than 3 points. Also, the alleles at the MN locus were chosen because there happened to be data collected for them. A survey of alleles at all loci might identify alleles that show relatively greater geographic differentiation. For example, alleles that are very rare may show greater relative differences among regions than allele with frequencies near .5, like the M and N alleles. Last, the above analysis is very sensitive to the value one uses for the observed inbreeding depression. A value of 3 points was used here, but another study by Agrawal, Sinha, and Jensen (1984) reported an inbreeding depression of approximately 6 points, which would effectively double the potential contribution of heterosis in the Flynn effect if it were used in the above analysis (see also Mingroni, 2004, for a discussion of some of the uncer-

Table 4
Models Designed to Agree With Japanese Inbreeding Study

L	D	R	μ_{out}	σ_{out}	μ_{in}	σ_{in}	$\frac{(\mu_{in} - \mu_{out})}{\sigma_{out}}$	$\frac{(\sigma_{in} - \sigma_{out})}{\sigma_{out}}$
520	.6	.5	286.0	8.15	284.4	8.39	−0.20 (−3.0)	.03
140	.7	.5	84.0	4.35	83.1	4.47	−0.20 (−3.0)	.03
68	.8	.5	44.2	3.17	43.6	3.25	−0.20 (−3.0)	.03
42	.9	.5	29.4	2.63	28.9	2.69	−0.20 (−3.0)	.02
31	1.0	.5	23.2	2.41	22.8	2.46	−0.20 (−3.0)	.02

Note. L = number of loci; D = dominance level; R = frequency of recessive alleles; μ_{out} = mean intelligence level of outbred group; σ_{out} = standard deviation of intelligence of outbred group; μ_{in} = mean intelligence level of inbred group; σ_{in} = standard deviation of intelligence in inbred group; $(\mu_{in} - \mu_{out})/\sigma_{out}$ = inbreeding depression in standard deviation units (values in parentheses show inbreeding depression in IQ metric); $(\sigma_{in} - \sigma_{out})/\sigma_{out}$ = measure of the effect of inbreeding on the standard deviation.

Table 5
IQ Changes Expected in Parma Valley

L	D	R	ΔHe	μ_1	σ_1	μ_2	σ_2	$\frac{(\mu_2 - \mu_1)}{\sigma_1}$	$\frac{(\sigma_2 - \sigma_1)}{\sigma_1}$
520	.6	.5	.03	284.4	8.38	286.0	8.15	0.19 (2.8)	−.03
140	.7	.5	.03	83.2	4.47	84.0	4.35	0.19 (2.8)	−.03
68	.8	.5	.03	43.6	3.25	44.2	3.17	0.19 (2.8)	−.02
42	.9	.5	.03	27.5	2.63	28.0	2.60	0.19 (2.8)	−.02
31	1.0	.5	.03	22.0	2.43	22.6	2.34	0.19 (2.8)	−.02

Note. L = number of loci; D = dominance level; R = frequency of recessive alleles; ΔHe = deficit of heterozygotes prior to demographic trend; μ_1 = mean intelligence level of population prior to demographic trend; σ_1 = standard deviation of intelligence prior to demographic trend; μ_2 = mean intelligence level of population after demographic trend; σ_2 = standard deviation of intelligence after demographic trend; $(\mu_2 - \mu_1)/\sigma_1$ = Flynn effect in standard deviation units (values in parentheses show Flynn effect in IQ points); $(\sigma_2 - \sigma_1)/\sigma_1$ = measure of the change over time in standard deviation.

tainties inherent in inbreeding studies and ideas for improving them). For these and other reasons, it would be difficult to extrapolate the above findings to estimate the likely contribution of heterosis in any actual observed IQ gain; genetic data covering the entire nation would likely be required.

Simulating Families

The genetic structure of IQ will influence the extent to which family members resemble each other in the trait. Any proposed model of the genes that influence IQ must therefore comport with the results of familial studies of IQ. Bouchard and McGue (1981) analyzed a large number of such studies and calculated weighted averages for the correlations observed among family members of different degrees of kinship. They arrived at weighted averages of .42, .47, and .50 for the parent–child, sibling, and midparent–child IQ correlations, respectively. In this section, models of the genes that influence IQ presented above are used to estimate the expected correlations among family members for models with different parameters.

As with the analysis of inbreeding, the present analysis assumes that we are dealing with a single randomly mating population, so there is no need for the parameter ΔHe, only L, D, and R. To estimate the three family correlations in question, we need to simulate a large number of nuclear families, each composed of two parents and two children. Genotypes for the parents in each family are stochastically generated in the same way as individuals in the randomly mating generation in the heterosis simulations and the outbred group in the inbreeding simulations. The parents' probability of inheriting two recessive alleles at a given locus is R^2, the probability of inheriting two dominants is $(1 - R)^2$, and the probability of inheriting one of each type of allele is $2R(1 - R)$. Because we are assuming that all loci influencing IQ are autosomal, it does not matter which parent is designated as the father and which as the mother, nor does it matter whether the children are boys or girls.

Each child's genotype at a given locus is determined by the parental genotypes already generated and the laws of Mendelian inheritance. At each locus, one of the father's alleles is chosen at random and passed on to the child, as is one of the mother's alleles. This allele selection process is done for the two children independently; that is, the inheritance of a particular allele by the first child does not influence which parental allele will be passed on to the second child. Once genotypes are generated for all four family members in a large number of families, an intelligence level is calculated for each individual using Equation 2. On the basis of these levels, the three correlations, parent–child, sibling, and midparent–child, are calculated.

Figure 2 shows the results of the simulation of families for models with different parameters. It turns out that varying L, the number of loci, has a negligible effect on the family correlations, and so this variable has been left out of the results. Each of the three graphs depicts one of the family correlations. Each curve in the graphs shows the relationship between D, the dominance level, and the particular family correlation r, for a given value of R, the recessive frequency. In general, increasing the dominance level, D, tends to reduce the expected family correlations. Also, positing recessive alleles that are relatively rare (smaller R) tends to reduce the expected family correlations.

In presenting their results, Bouchard and McGue (1981) pointed out that the correlations observed in the familial IQ studies they surveyed were reasonably close to theoretically expected values, on the basis of the assumption of strictly additive gene action ($D = .5$ in the terminology of this article). These theoretically expected values are .50, .50, and .707 for the parent–child, sibling, and midparent–child correlations, respectively; one can see that the correlations predicted by the simulations gradually approach these values as D approaches .5. Recall that the observed values, the weighted averages of many studies, were .42, .47, and .50, respectively. From the graphs, one can see that although positing nonadditive gene action ($D > .5$) does tend to reduce the expected family correlations, it does not reduce them so much as to be completely out of line with observed findings. For example, from the graphs one can see that a model that posits a D value of .8 and an R value of .5 predicts values of .42, .46, and .60 for the parent–child, sibling, and midparent–child correlations, respectively. This fits the observed findings about as well as models that posit only additive gene action ($D = .5$).

It should be emphasized that attempts to very precisely fit models like those presented here are probably unwarranted, given that a number of potentially relevant factors, such as measurement error, assortative mating, and regional genetic variation, have not yet been incorporated. Assortative mating, in particular, would tend to increase both the parent–child and sibling correlations but would have little effect on the midparent–child correlation. Nevertheless, the analysis does show that the results of familial IQ studies are not necessarily at odds with models of the genetic

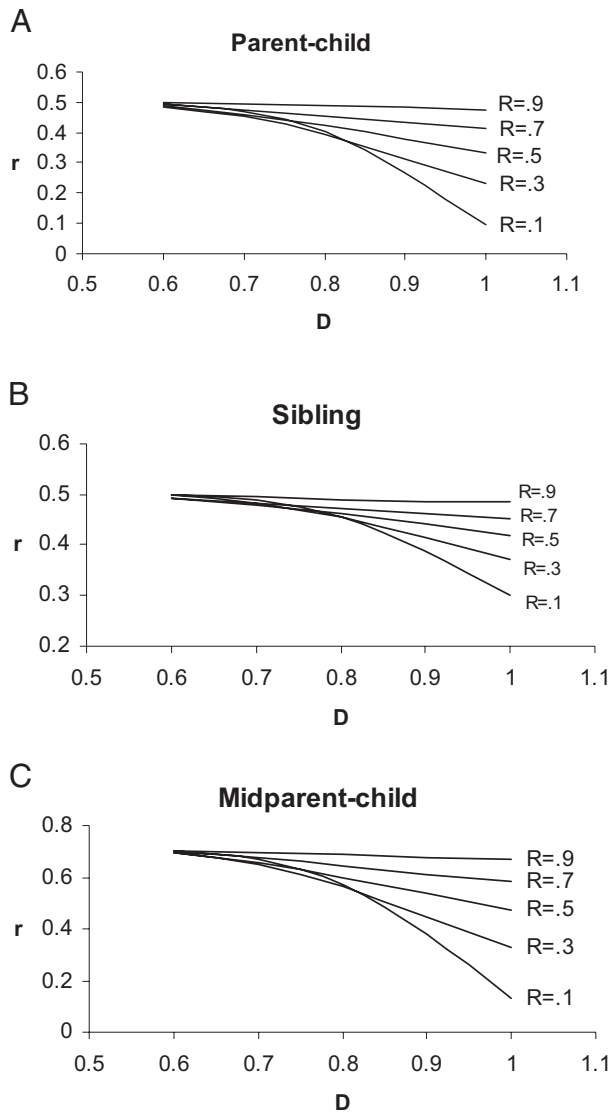

Figure 2. Expected parent–child correlations (A), sibling correlations (B), and midparent–child correlations (C). r = expected correlation; D = dominance level; R = recessive allele frequency.

structure of IQ that posit significant nonadditive gene action. Moreover, such nonadditive gene action is essential to explain inbreeding depression in IQ.

It is possible to take findings from the simulation of families and apply them to the hypothetical exercise carried out earlier using data from the Parma Valley. That analysis yielded five different models, all of which comported reasonably well with inbreeding study results, and also posited genetic differentiation within the initial generation that matched the level of genetic differentiation observed in actual genes. Table 6 presents the results of the simulation of families for the same five models presented in Tables 4 and 5. Although none of the models provides a perfect fit to the observed values from familial IQ studies, some are reasonably good. For example, the model with 70 loci, a D value of .8, and an R value of .5 predicts values of .42, .46, and .60 for the parent–child, sibling, and midparent–child correlations, which comport reasonably well with the observed findings of .42, .47, and .50.

Before proceeding, it is worth discussing a frequent source of confusion related to the potential effects of heterosis. This is the practice, common in quantitative genetic analyses, of decomposing the genetic variance into additive genetic and nonadditive genetic components (e.g., Falconer & Mackay, 1996, pp. 125–126). Quite understandably, it is thought that when such analysis reveals that the additive genetic component is larger than the nonadditive component, the genes must be interacting in a mostly additive manner (D near .5 in the present framework). However, the link between the two components of genetic variance and the way the genes are interacting, the gene action, is not necessarily strong. Consider the following quote from Falconer and Mackay (1996):

> A possible misunderstanding about the concept of additive genetic variance, to which the terminology may give rise, should be mentioned here. The concept of additive genetic variance does not carry with it the assumption of additive gene action and the existence of additive variance is not an indication that any of the genes act additively (i.e., show neither dominance nor epistasis). No assumption is made about the mode of action of the genes concerned. Additive variance can arise from genes with any degree of dominance or epistasis, and only if we find that all the genotypic variance is additive can we conclude that the genes show neither dominance nor epistasis. (p. 128)

The above statement is also borne out by the results of the simulation of families shown in Figure 2. Notice that in some cases, as when the value of R is high, the expected correlations for a model with strictly additive gene action (D = .5) are nearly the same as for a model that posits genes interacting with complete dominance (D = 1.0). Because the additive and nonadditive components of genetic variance are calculated on the basis of family correlations, one would arrive at the same estimates for these components in all models, even though the gene action posited is very different. Thus, although partitioning the genetic variance into additive and nonadditive components may be helpful for other purposes in quantitative genetics, it is not particularly useful when one is trying to understand the degree of dominance with which the genes are interacting. Of note, if the model simulations predict family correlations that are in line with observed values, the model will inevitably be in line with estimates of additive versus nonadditive genetic variance, because these estimates are wholly derived from the family correlations.

Table 6
Expected Kinship Correlations for Solutions to Inbreeding Study Results

L	D	R	$r_{p\text{-}c}$	r_{sib}	$r_{m\text{-}c}$
520	.6	.5	.49	.47	.69
140	.7	.5	.46	.48	.66
68	.8	.5	.42	.46	.60
42	.9	.5	.38	.44	.54
31	1.0	.5	.33	.42	.47

Note. L = number of loci; D = dominance level; R = recessive allele frequency; $r_{p\text{-}c}$ = parent–child correlation; r_{sib} = sibling correlation; $r_{m\text{-}c}$ = midparent–child correlation.

Applicability to Other Trends

Heterosis increases the ratio of heterozygotes to homozygotes at every polymorphic locus. As such, it is capable of causing simultaneous trends in multiple polygenic traits. As mentioned earlier, many other heritable traits and conditions have undergone large changes over time in parallel with the IQ trend. It is possible to adapt the stochastic modeling used above for IQ to try to account for these other trends. In the case of some trends, the adaptation is very simple; in other cases, it is slightly more complicated.

In the case of some traits, such as height, adapting the above analysis merely involves changing the dependent variable in Equation 2 from intelligence to the new trait. Also, because traits like height are measured on an absolute scale, the presumed effects on the trait in Table 1 would not be dimensionless, as they must be in the case of intelligence, but could be measured in actual units of length like millimeters. Use of an absolute scale would also place additional constraints on the universe of plausible models, because the ratio σ/μ would have to comport with empirically observed values. There is no way to measure this ratio for intelligence.

Some of the trends do not involve continuously varying traits, in which one can observe changes in the entire distribution. Instead, they involve heritable conditions in which the population is divided into discrete groups, such as affected versus unaffected. The trends in these conditions are measured by changes in the incidence of individuals categorized as affected, according to some diagnostic criteria. Conditions like myopia, asthma, and autism belong to this group. Though still possible, modeling the trends in conditions like these requires making some additional assumptions about their underlying genetic structure, as well as some additional calculations.

The first step in adapting the above analysis to discontinuous traits is that one must change the dependent variable in Equation 2 from intelligence to something like "liability to myopia" or "genetic predisposition to autism." In this case, one would be assuming that all individuals have some genetic liability to the condition and that this liability is distributed normally in the population. Those affected would represent the upper tail of this presumed normal liability distribution, above some threshold value. Such a model of the underlying liability to a condition, referred to as a *polygenic threshold model*, is commonly used in epidemiological studies (Falconer, 1965).

A shift toward more random mating would be capable of causing upward shifts in the presumed liability distributions of discontinuous traits in the same way that it is capable of causing shifts in the distribution of continuously varying traits. As the underlying liability distribution shifts upward, a larger portion of the population will fall above the threshold to be considered affected, thereby causing the incidence to rise. For a given change in the incidence of a condition, it is a simple matter to estimate the magnitude of the shift that would have to take place in its underlying liability in order to cause the observed rise in incidence. One need only make reference to a table of areas under the normal curve with the earlier and later incidence levels.

As an example, a series of studies in Sweden in the 1980s reported an apparent rise in the incidence of autism, from 4.0/10,000 in 1980 to 11.6/10,000 in 1988 (Gillberg, Steffenburg, & Schaumann, 1991). Such an increase in incidence could be explained by an upward shift in an underlying normal liability to autism of 0.32 standard deviations in the 8 years (z threshold $+3.35$ to z threshold $+3.03$), or about 0.4 standard deviations per decade. This estimate assumes that there were no changes in factors like diagnostic criteria or ascertainment methods and also assumes no change in the variance of the underlying distribution. It should be noted that in the specific case of autism there is still considerable debate among experts as to whether the apparent increases that have appeared in a number of studies are real, due to artifacts, or both (see, e.g., Gernsbacher, Dawson, & Goldsmith, 2005). Obviously, the above analysis would be applicable only if the trend were eventually found to be real.

Similarly, an observed increase in asthma from 3.0% to 8.2% in 22 years (Upton et al., 2000) could be accounted for by an upward shift of 0.22 standard deviations per decade. An observed increase in myopia, from 26.3% to 43.3% in approximately 10 years (Storfer, 1999), could be accounted for by a shift over time in the underlying liability to myopia of 0.46 standard deviations per decade.

It is interesting to note that the rates at which the liability distributions would appear to be shifting are similar to the rates at which the distributions of continuous traits like height and IQ have shifted; these have generally ranged from 0.2 to 0.5 standard deviations per decade. In the context of a heterosis hypothesis, this would suggest that the discontinuous traits have genetic characteristics similar to the continuous traits with respect to things like the number of loci involved, the dominance level, and the ratio of dominant to recessive genes in the gene pool.

It should be noted that a major problem for the analysis of discontinuous traits is that their incidence will be affected by changes in both the mean and the variance of the liability distributions over time. However, because the trends are measured by changes in a single variable, the incidence, there is no way to disentangle effects brought about by changes in the mean of the distribution from effects caused by changes in the variance. The examples cited above simply assumed no change in variance; however, future, more detailed modeling efforts would have to take into account the very real possibility that the variance might be changing. This will likely mean that the universe of plausible genetic structures for the conditions in question may be less constrained and therefore less specific.

Growth Saltations

In the case of at least one of the polygenic traits that has undergone change, height, it may be possible to observe the effects of individual genes on the phenotype, which opens up a number of opportunities for testing a heterosis hypothesis. Recent investigations have found that when height is measured frequently, such as on a daily basis, it is possible to discern very fine grained growth spurts, or saltations (Hermanussen et al., 1998; Lampl, 2002). For example, a child will sometimes go for a week or two with no discernible growth, followed by one or two days in which as much as a centimeter of growth occurs, followed by another extended period of stasis. Of interest, Lampl (2002, p. 267) cited unpublished data suggesting that monozygotic twins resemble one another in the pattern of these growth spurts. The similarity of monozygotic twins in the timing of saltations raises the possibility that each spurt might represent the action of a single gene, or at most a small number of genes. Such a one-gene/one-saltation

hypothesis generates several testable predictions in the context of the present effort to model the genes that influence height.

Within the context of the simple framework outlined in Table 1, a heterosis hypothesis would predict that the trend in height should manifest itself primarily as an increase in the frequency of saltations, rather than an increase in the average amplitude of saltations. In fact, unless the genes that influence height interact with complete dominance ($D = 1.0$), the trend in height should actually be accompanied by a decline over time in the average amplitude of saltations, reflecting the increased number of heterozygous loci and the decrease in double dominants. Unfortunately, testing such a prediction would be difficult, requiring frequent testing of a large number of individuals over a long period of time.

A more feasible test might be to compare children in tall populations that have already experienced many decades of height gains with children in shorter, developing countries, whose current environment is comparable to past environments in the now developed populations. In general, a heterosis hypothesis would predict that the difference should be observed mainly as a difference in the frequency, rather than the amplitude, of saltations. Also, it would be interesting to compare inbred children with noninbred children. Again, a genetic framework like that posited here would predict that the height difference between the inbred and noninbred groups should appear mainly as a difference in the frequency of saltations.

The occurrence of identifiable growth saltations also offers the possibility of testing the plausibility of models of the genes that influence height, such as those presented earlier. According to the simple framework posited in the present article, if each saltation is caused by the genes at a single locus, then the number of saltations that an average individual experiences up until adulthood should be approximately equal to the average number of loci at which individuals inherit either one or two dominant alleles: that is, $L_{DD} + L_{DR}$. One would probably have to change the assumption that all loci have an equal effect on height; hypothesized variation in the effects of different loci would have to match observed variation in the amplitudes of growth saltations. The simple framework presented here would also predict that sex differences in height should show up as a difference in the amplitude of saltations rather than in their frequency. Because the present model assumes all loci are autosomal, boys and girls should be affected by the same number of loci and therefore experience the same overall frequency of saltations. Thus, one can see that as with other relevant empirical findings, the need to comport with growth saltation studies will likely constrain the universe of plausible genetic models.

As mentioned already, the remarkable similarity in height of monozygotic twins reared apart, particularly in the earliest studies, conducted in populations that went on to experience large subsequent increases in height, poses a serious problem for environmental hypotheses for the trend. Such results suggest that any presumed environmental deprivations must have stunted past populations with a remarkable level of uniformity and pervasiveness, hardly varying at all among households. The study of growth saltations offers the possibility of measuring this uniformity and pervasiveness in much greater detail. Specifically, if one could identify a pair of monozygotic twins currently being reared apart in a developing country, both twins could be measured daily in their respective homes. If the twins were to resemble each other in the timing and amplitude of their growth saltations, it would mean either that no stunting is occurring in the population or that both twins are somehow being stunted by the same amount every week, or even every day, despite living in different homes. Although such a test would require a unique set of circumstances, it would only entail the measurement of two children for several months. It might also be interesting to take pairs of monozygotic twins who are currently being reared together in a developing country and separate them for a short while, although ethical considerations would limit the extent to which one could vary their respective environmental conditions while measuring them.

Developmental Modeling

In all of the models presented in this article, it has been assumed that every individual's phenotype is influenced by the same number of gene loci. This assumption, however, is probably not correct in the case of children. Because different genes will likely have their effect on the phenotype at different ages, it would probably be more appropriate to assume that the phenotype in childhood is influenced by fewer genes than the adolescent or adult phenotype. It should be possible to incorporate the likely developmental nature of genetic influences into the modeling of polygenic traits.

For example, recall that in the simulation of families, alleles at each locus were randomly passed on from parents to children. It would be a simple matter to allow only the alleles at some of the loci to be passed on, so that the phenotypes of the children would then be influenced by only a fraction of the loci that influence the adult phenotype. By gradually increasing the fraction of loci passed on to the children, one could simulate developmental changes in the trait in question. One could estimate likely changes in the different family correlations as children get older. One could also calculate within-individual correlations at different ages and compare these with observed findings. In the case of traits like height that are measured on an absolute scale, the model would also need to comport with observed developmental changes in both the mean and standard deviation of the trait.

Incorporating developmental processes into the modeling process will also influence the interpretation of inbreeding study results as well as data on time trends, both of which will likely be influenced by the age of the subjects. Recall that in the models presented above, positing more loci predicts larger inbreeding depression and larger Flynn effects, all other things being equal. In general, then, one would initially expect studies involving younger children to show smaller inbreeding depression and smaller Flynn effects, due to the fact that the phenotype in childhood is likely influenced by fewer loci than the adult phenotype. This may provide a partial explanation for the apparent finding of larger Flynn effects in adult samples as compared to children (Flynn, 1998, p. 27).

Heterosis Hypothesis Predictions

This last section provides a summary of the various testable predictions of the heterosis hypothesis that have been mentioned throughout the article. The first four predictions apply to an observed trend in any trait or condition, be it a gross physical trait like height, a behavioral trait like IQ, a threshold trait like asthma or myopia, or a biochemical trait like the level of a particular hormone. The last prediction deals with traits like height, or

possibly protein levels, in which it may be possible to discern the effects of individual genes, as manifested by things such as growth spurts or spikes in hormone levels. It must be noted that all of the predictions below are premised on a strong version of the hypothesis in which heterosis is the primary cause of the trend in question and differences in the genes are the primary cause of variance in the trait. Obviously, to the extent that other factors may be responsible for the trends or contribute substantially to variance, the following expectations may not be fully met.

1. In populations in which the trend in question is occurring, one should observe an increase over time in the frequency of heterozygotes. More specifically, there should be more families in which the parents are more homozygous than their children, on average, at a randomly selected set of gene loci. The reason to look intergenerationally within families rather than across the entire population is that some segments of the population may be relatively more or less homozygous, on average, owing to things like one segment being descended from a very small founding population. Differential fertility rates in the different segments could then cause a rise or fall in overall heterozygosity rates that would obscure what is going on within families.

2. The trait in question should display directional dominance in the same direction as the trend. Although studies involving indirect measures of likely genetic heterozygosity like parental relatedness or parental birthplace may be informative, studies should ideally examine the genes of the subjects directly. Specifically, families in which there are greater intergenerational increases in heterozygosity should account for a greater share of the intergenerational changes in the trait. Here again, it is preferable to look within families because different segments of the population might differ in both their average heterozygosity rate and their mean values of the trait, creating a spurious correlation between genetic heterozygosity and the trait unrelated to the existence of directional dominance.

3. It should be possible to develop a model of the genetic structure of the trait whereby observed genetic changes can account for both the observed phenotypic changes over time and the magnitude of directional dominance observed in the trait. It is important to note that the proposed genetic structure of the trait should not predict outcomes that conflict with other relevant empirical findings, such as the observed correlations among family members or features of the development of the trait throughout childhood.

4. The trait in question should not be observed within families as one goes from earlier born to later born siblings.

5. In traits like height for which it may be possible to discern the effects of individual genes, the number of growth saltations should be approximately equal to the number of loci at which the genes have a positive effect on height in the proposed model of the genetic structure of the trait. Differences in height between tall and short populations, as well as differences between inbred and outbred groups, should be caused primarily by differences in the frequency, rather than the amplitude, of growth saltations. By contrast, sex differences should be the result of differences in the amplitude of saltations, assuming that most of the genes that influence height are autosomal. Also, monozygotic twins should resemble one another in both the timing and the magnitude of the saltations, regardless of whether they are living in an affluent or poor population and whether they are being reared together or apart.

Conclusion

For more than two decades now researchers have searched in vain for the environmental cause of the Flynn effect, and it seems unlikely that such a cause will be identified soon. The reason for such pessimism is that the IQ paradox renders all environmental hypotheses implausible, untestable, or both. For example, one can look to identifiable factors like nutrition and education, but given the high heritability of IQ, the associations between cause and effect will not persist beyond the ecological level. Any postnatal factor, even one with the characteristics of a factor X, is problematic because it must have either a shared-family component, a nonshared component, or both. However, MZA twin and adoption studies suggest the cause does not have a large shared component, and birth order studies suggest it does not have a large nonshared component. The only environmental factors with much hope of resolving the paradox are shared prenatal factors, or environmental factors that are so highly correlated with the genes as to be indistinguishable from them. Even these face the problem of the stability of IQ heritability over time. Moreover, even if factors of this type are responsible for the Flynn effect, their very nature makes it difficult to disentangle their effects from those of genetic factors, making them inherently difficult to identify or measure.

If, however, the Flynn effect is genetic in origin, the IQ paradox is not only resolved, it is eliminated; there is nothing paradoxical to explain. And the only plausible genetic cause yet proposed is heterosis. Perhaps most important, and in contrast to most environmental hypotheses, a heterosis hypothesis generates numerous predictions that can be tested empirically. In a world of limited resources, scientists rarely have the option of pursuing in great depth every hypothesis put forward to explain a particular phenomenon. Instead, it is usually necessary to engage in a kind of cost–benefit analysis, focusing mainly on those lines of inquiry that seem most likely to advance our understanding of the phenomena in question. In light of its plausibility, testability, and potential to explain multiple phenomena, a heterosis hypothesis for the Flynn effect represents a working hypothesis superior to any yet proposed. It is important to point out that even if the hypothesis is eventually disproved, the process of testing it is likely to provide clues to the eventual cause, whatever that may be, as well as clues to the genetic structure of the traits and conditions in question. Considering that there is so much to be gained and so little to lose, the testing of a heterosis hypothesis should be the primary focus of future efforts to better understand the cause of the worldwide rise in IQ test scores.

References

Agrawal, N., Sinha, S. N., & Jensen, A. R. (1984). Effects of inbreeding on Raven matrices. *Behavior Genetics, 14,* 579–585.

Alroy, J. (1998, May 1). Cope's rule and the dynamics of body mass evolution in North American fossil mammals. *Science, 280,* 731–734.

Anderson, A. M. (1982, May 20). The great Japanese IQ increase. *Nature, 297,* 181–182.

Barber, N. (2005). Educational and ecological correlates of IQ: A cross-national investigation. *Intelligence, 33,* 273–284.

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

Billy, G. (1980). Microevolution and exogamy. In I. Schwidetzky, B. Chiarelli, & O. Necrasov (Eds.), *Physical anthropology of European populations* (pp. 59–68). The Hague, the Netherlands: Mouton.

Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990, October 12). Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science, 250,* 223–228.

Bouchard, T. J., Jr., & McGue, M. (1981, May 29). Familial studies of intelligence: A review. *Science, 212,* 1055–1059.

Brand, C. (1987, August 27). Keeping up with the times. *Nature, 328,* 761.

Brinkman, H. J., Drukker, J. W., & Slot, B. (1988). Height and income: A new method for the estimation of historical national income series. *Explorations in Economic History, 25,* 227–264.

Capron, C., & Duyme, M. (1996). Effect of socioeconomic status of biological and adoptive parents on WISC–R subtest scores of their French adopted children. *Intelligence, 22,* 259–275.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cavalli-Sforza, L. L., Moroni, A., & Zei, G. (2004). *Consanguinity, inbreeding, and genetic drift in Italy.* Princeton, NJ: Princeton University Press.

Colom, R., Lluis-Font, J. M., & Andres-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence, 33,* 83–91.

Comings, D. E. (1996). *The gene bomb.* Duarte, CA: Hope Press.

Cook, E. H., Jr., & Leventhal, B. L. (1996). The serotonin system in autism. *Current Opinion in Pediatrics, 8,* 348–354.

Dahlberg, G. (1942). *Race, reason and rubbish.* London: Allen & Unwin.

Damon, A. (1965). Stature increase among Italian-Americans: Environmental, genetic, or both? *American Journal of Physical Anthropology, 23,* 401–408.

Deary, I. J. (2001). *Intelligence: A very short introduction.* Oxford, England: Oxford University Press.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108,* 346–369.

Dickens, W. T., & Flynn, J. R. (2002). The IQ paradox is still resolved: Reply to Loehlin (2002) and Rowe and Rodgers (2002). *Psychological Review, 109,* 764–771.

Drukker, J., & Van Meerten, M. (1995). Beyond Villerme and Quetelet: The quantitative relation between sex- and age-specific height and real per capita income. In J. Komlos (Ed.), *The biological standard of living on three continents: Further explorations in anthropometric history* (pp. 25–58). Boulder, CO: Westview Press.

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

Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics, 29,* 51–76.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Essex, England: Longman.

Ferak, V., Lichardnova, Z., & Borjinova, V. (1968). Endogamy, exogamy and stature. *Eugenics Quarterly, 15,* 273–276.

Fisher, R. A. (1928). The possible modification of the wild type to recurrent mutations. *American Naturalist, 62,* 115–126.

Fisher, R. A. (1931). The evolution of dominance. *Biological Reviews, 6,* 345–368.

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

Flynn, J. R. (1994). IQ gains over time. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 617–623). New York: Macmillan.

Flynn, J. R. (1998). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve*: *Long-term gains in IQ and related measures* (pp. 25–66). Washington, DC: American Psychological Association.

Gernsbacher, M. A., Dawson, M., & Goldsmith, H. H. (2005). Three reasons not to believe in the autism epidemic. *Current Directions in Psychological Science, 14,* 55–58.

Gillberg, C., Steffenburg, S., & Schaumann, H. (1991). Is autism more common now than ten years ago? *British Journal of Psychiatry, 158,* 403–409.

Gottfredson, L. S. (1997). Why g matters. *Intelligence, 24,* 13–23.

Gould, S. J., & Eldredge, N. (1977). Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology, 3,* 115–151.

Hanson, R., Smith, J. A., & Hume, W. (1985). Achievements of infants on items of the Griffiths scales: 1980 compared to 1950. *Child Care, Health, and Development, 11,* 91–104.

Helgason, A., Yngvadottir, B., Hrafnkelsson, B., Gulcher, J., & Stefansson, K. (2005). An Icelandic example of the impact of population structure on association studies. *Nature Genetics, 37,* 90–95.

Hermanussen, M., Thiel, C., von Buren, E., De Los Angeles, M., De Lama, R., Perez Romero, A., et al. (1998). Micro and macro perspectives in auxology: Findings and considerations upon the variability of short term and individual growth and the stability of population derived parameters. *Annals of Human Biology, 25,* 359–385.

Hulse, F. (1964). Exogamy and heterosis. In G. W. Lasker (Ed.), *Yearbook of physical anthropology* (pp. 241–257). Washington DC: American Association of Physical Anthropologists.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–98.

Jensen, A. R. (1983). Effects of inbreeding on mental-ability factors. *Personality and Individual Differences, 4,* 71–87.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jensen, A. R., & Sinha, S. N. (1993). Physical correlates of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 139–242). Norwood, NJ: Ablex.

Kacser, H., & Burns, J. A. (1981). The molecular basis of dominance. *Genetics, 97,* 639–666.

Kane, H., & Oakland, T. D. (2000). Secular declines in Spearman's g: Some evidence from the United States. *Journal of Genetic Psychology, 161,* 337–345.

Kehle, T. J., Bray, M. A., Theodore, L. A., Zhou, Z., & McCoach, D. B. (2004). Emotional disturbance/social maladjustment: Why is the incidence increasing? *Psychology in the Schools, 41,* 861–865.

Kimura, M. (1983). *The neutral theory of molecular evolution.* Cambridge, England: Cambridge University Press.

Komlos, J. (Ed.). (1994). *Stature, living standards, and economic development: Essays in anthropometric history.* Chicago: University of Chicago Press.

Komlos, J. (Ed.). (1995). *The biological standard of living on three continents: Further explorations in anthropometric history.* Boulder, CO: Westview Press.

Lampl, M. (2002). Saltation and stasis. In N. Cameron (Ed.), *Human growth and development* (pp. 253–270). Amsterdam, the Netherlands: Academic Press.

Loehlin, J. C. (2002). The IQ paradox: Resolved? Still an open question. *Psychological Review, 109,* 754–758.

Loehlin, J. C., Horn, J. M., & Willerman, L. (1989). Modeling IQ change: Evidence from the Texas Adoption Project. *Child Development, 60,* 993–1004.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence, 31,* 543–566.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits.* Sunderland, MA: Sinauer.

Lynn, R. (1998a). The decline of genotypic intelligence. In U. Neisser (Ed.), *The rising curve*: *Long-term gains in IQ and related measures* (pp. 335–364). Washington, DC: American Psychological Association.

Lynn, R. (1998b). In support of the nutrition theory. In U. Neisser (Ed.), *The rising curve*: *Long-term gains in IQ and related measures* (pp. 207–218). Washington, DC: American Psychological Association.

Mingroni, M. A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence, 32,* 65–83.

Mueller, W. H. (1986). The genetics of size and shape in children and adults. In F. Falkner & J. M. Tanner (Eds.), *Human growth: A comprehensive treatise* (Vol. 3, pp. 145–168). New York: Plenum Press.

Murray, C. (1997). IQ and economic success. *Public Interest, 128,* 21–35.

Neisser, U. (1998). *The rising curve*: *Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Newman, H. H., Freeman, F. N., Holzinger, K. J. (1937). *Twins: A study of heredity and environment.* Chicago: University of Chicago Press.

Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science, 3,* 346–353.

Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture, and cognitive development from 1 to 16 years: A parent–offspring adoption study. *Psychological Science, 8,* 442–447.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 1. General Overview.* Oxford, England: Oxford Psychologists Press, and San Antonio, TX: Psychological Corporation.

Rodgers, J. L. (1999). A critique of the Flynn effect: Massive IQ gains, methodological artifacts, or both? *Intelligence, 26,* 337–356.

Rodgers, J. L., Cleveland, H. H., van den Oord, E., & Rowe, D. C. (2000). Resolving the debate over birth order, family size and intelligence. *American Psychologist, 55,* 599–612.

Rowe, D. C., Jacobson, K. C., & Van den Oord, J. C. G. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70,* 1151–1162.

Rowe, D. C., & Rodgers, J. L. (2002). Expanding variance and the case of historical changes in IQ means: A critique of Dickens and Flynn. *Psychological Review, 109,* 759–763.

Rutter, M., & Silberg, J. (2002). Gene environment interplay in relation to emotional and behavioral disturbance. *Annual Review of Psychology, 53,* 463–490.

Rutter, M., & Smith, D. J. (Eds.). (1995). *Psychosocial disorders in young people: Time trends and their causes.* Chichester, England: Wiley.

Scarr, S., & Weinberg, R. A. (1978). The influence of "family background" in intellectual attainment. *American Sociological Review, 43,* 674–692.

Schreider, E. (1969). Inbreeding, biological and mental variations in France. *American Journal of Physical Anthropology, 30,* 215–220.

Schull, W. J., & Neel, J. V. (1965). *The effects of inbreeding on Japanese children.* New York: Harper & Row.

Shapiro, H. L. (1936). *The heritage of the Bounty: The story of Pitcairn through six generations.* New York: Simon & Schuster.

Shih, Y. F., Hsiao, C. K., Chen, C. J., Chang, C. W., Hung, P. T., & Lin, L. L. K. (2001). An intervention trial on the efficacy of atropine and multi-focal glasses in controlling myopic progression. *Acta Ophthalmologica Scandinavica, 79,* 233–236.

Sibbald, B. (1997). Familial inheritance of asthma and allergy. In A. B. Kay (Ed.), *Allergy and allergic diseases* (Vol. 2, pp. 1177–1186). London: Blackwell Science.

Steckel, R., & Floud, R. (Eds.). (1997). *Health and welfare during industrialization.* Chicago: University of Chicago Press.

Stein, Z., Susser, M., Saenger, G., & Marolla, F. (1972, November 17). Nutrition and mental performance. *Science, 178,* 708–713.

Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. *Psychological Bulletin, 125,* 392–409.

Storfer, M. (1999). Myopia, intelligence, and the expanding human neocortex: Behavioral influences and evolutionary implications. *International Journal of Neuroscience, 98,* 153–276.

Sundet, J. M., Barlag, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence, 32,* 349–362.

Tanner, J. M. (1981). *A history of the study of human growth.* Cambridge, England: Cambridge University Press.

Tanner, J. M. (1990). *Foetus into man: Physical growth from conception to maturity.* Cambridge, MA: Harvard University Press.

Tanner, J. M. (1994). Introduction: Growth in height as a mirror of the standard of living. In J. Komlos (Ed.), *Stature, living standards, and economic development: Essays in anthropometric history* (pp. 1–6). Chicago: University of Chicago Press.

Tasbihsazan, R., Nettelbeck, T., & Kirby, N. (1997). Increasing mental development index in Australian children: A comparative study of two versions of the Bayley Mental Scale. *Australian Psychologist, 32,* 120–125.

Teasdale, T. W., & Owen, R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence, 28,* 115–120.

Teasdale, T. W., & Owen, R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn effect in reverse. *Personality and Individual Differences, 39,* 837–843.

Upton, M. N., McConnachie, A., McSharry, C., Hart, C. L., Smith, G. D., Gillis, C. R., & Watt, G. C. M. (2000). International 20 year trends in the prevalence of asthma and hay fever in adults: The Midspan family study surveys of parents and offspring. *British Medical Journal, 321,* 88–92.

Van Wieringen, J. C. (1986). Secular growth changes. In F. Falkner & J. M. Tanner (Eds.), *Human growth: A comprehensive treatise* (Vol. 3, pp. 307–331). New York: Plenum Press.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Psychological Corporation.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. S. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of Flynn effect. *Intelligence, 32,* 509–537.

Williams, W. (1998). Are we raising smarter children today? School- and home-related influences on IQ. In U. Neisser (Ed.), *The rising curve*: *Long-term gains in IQ and related measures* (pp. 125–154). Washington, DC: American Psychological Association.

Wilson, R. S. (1986). Growth and development of human twins. In F. Falkner & J. M. Tanner (Eds.), *Human growth: A comprehensive treatise* (Vol. 3, pp. 197–211). New York: Plenum Press.

Wolanski, N., Jarosz, E., & Pyzuk, M. (1970). Heterosis in man: Growth in offspring and distance between parents' birthplaces. *Social Biology, 17,* 1–16.

Wright, S. (1934). Molecular and evolutionary theories of dominance. *American Naturalist, 63,* 24–53.

Zajonc, R. B., & Mullally, P. R. (1997). Birth order: Reconciling conflicting effects. *American Psychologist, 52,* 685–699.

Received February 24, 2006
Revision received November 1, 2006
Accepted January 3, 2007 ■

Correction to Mingroni (2007)

In the article "Resolving the IQ Paradox: Heterosis as a Cause of the Flynn Effect and Other Trends," by Michael A. Mingroni (*Psychological Review,* 2007, Vol. 114, No. 3, pp. 806–829), an equation appearing on page 821 in the text and in Figure 1 was incorrect. The equation $(\sigma_{out} - \sigma_{in})/\sigma_{out}$ appearing on the third line from the bottom of the left column of text; as the *y*-axis labels in Figure 1, Panels A, B, and C; and in the next-to-last and last lines of the caption to Figure 1 should be $(\sigma_{in} - \sigma_{out})/\sigma_{out}$ in all these places.