242-74

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

PROFESSOR JENSEN, MEET MISS BURKS

Arthur S. Goldberger

UNIVERSITY OF WISCONSIN-MADISON

PROFESSOR JENSEN, MEET MISS BURKS

Arthur S. Goldberger

December 1974

The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity pursuant to the Economic Opportunity Act of 1964, and by Grant GS-39995 of the National Science Foundation. I am deeply indebted to Glen Cain and Leon Kamin for many instructive comments. I am also grateful to Dudley Duncan, David Layzer, Paul Taubman, and Sewall Wright for helpful responses to an earlier draft. The opinions expressed in this paper are mine and should not be attributed to the institutions and individuals named above.

ABSTRACT

We critically examine the portions of Arthur Jensen's books, Genetics and Education and Educability and Group Differences, that concern Barbara Burks's 1928 study of adoptive families. Jensen cites the low correlations of children's IQs with measures of home environment as evidence that environment plays only a minor role in the determination of intelligence. We find that Burks's sample was highly selective, that her environmental measures were limited, and that Jensen has thoroughly misrepresented the content and implications of the Burks study.

PROFESSOR JENSEN, MEET MISS BURKS

Arthur S. Goldberger

Burking = murdering...stifling or quietly suppressing ...

Oxford English Dictionary

1. INTRODUCTION

In his two recent books, Arthur R. Jensen (1972a, 1973a) draws on a classic study by Barbara S. Burks (1928) to support his contention that heredity, rather than environment, plays the predominant role in the determination of intelligence.

Jensen's presentation of the Burks study is incredible, in several senses. To determine this, we need only read Jensen and then read Burks.

2. JENSEN'S REPORT

Reproduced below are the passages in Jensen's books that deal with Burks's study. For ease of reference, I have italicized and numbered selected items.

Jensen (1972a, pp. 128-130):

Direct Measurement of the Environment. Another method for getting at the relative contribution of environmental factors to IQ variance is simply by correlating children's IQs with ratings of their environment. This can be legitimately done only in the case of adopted children and where there is evidence that selective placement by the adoption agencies is negligible. Without these conditions, of course, some of the correlation between the children and their environ-
(1) mental ratings will be due to genetic factors. *There are two large-scale studies in the literature which meet these criteria. Also, both studies involved adopting parents who were representative of a broad cross-section of the U.S. Caucasian population with respect to education, occupation, and socio-economic level. It is probably safe to say that not more than 5 percent of the U.S. Caucasian population falls outside the range of environmental variation represented in the samples in these two studies.* The study by Leahy (1935) found an average correlation of 0.20 between the IQs of adopted children and a number of indices of the 'goodness' of their environment, including the IQs and education of both adopting parents, their socioeconomic status, and the cultural amenities in the home. Leahy concluded from this that the environmental ratings accounted for 4 percent (i.e., the square of r = 0.20)

of the variance in the adopted children's Stanford-Binet IQs, and that 96 percent of the variance remained to be accounted for by other factors. The main criticisms we can make of this study are, first, that the environmental indices were not sufficiently 'fine-grained' to register the subtleties of environmental variation and of the qualities of parent-child relationship that influence intellectual development, and second, that the study did not make use of the technique of multiple correlation, which would show the total contribution to the variance of all the separate environmental indices simultaneously. A multiple correlation is usually considerably greater than merely the average of all the correlations for the single variables.

A study by Burks (1928) meets both these objections.

(2) *To the best of my knowledge no study before or since has rated environments in any more detailed and fine-grained manner than did Burks'. Each adoptive home was given 4 to 8 hours of individual investigation. As in Leahy's study, Burks included intelligence measures on the adopting parents as part of the children's environments, an environment which also included such factors as the amount of time the parents spent helping the children with their school work, the amount of time spent reading to the children, and so on. The multiple correlation (corrected for unreliability)*

between Burks' various environmental ratings and the adopted children's Stanford-Binet IQs was 0.42. The square of this correlation is 0.18, which represents the proportion of IQ variance accounted for by Burks' environmental measurements. This value comes very close to the environmental variance estimated in direct heritability analyses based on kinship correlations.

(3) *Burks translated her findings into the conclusion that the total effect of environmental factors one standard deviation up or down the environmental scale is only about 6 IQ points...*

(4) *Another part of Burks' study consisted of a perfectly matched control group of parents raising their own children, for whom parent-child correlations were obtained. Sewall Wright (1931) performed a heritability analysis on these parent-child and IQ-environment correlations and obtained a heritability coefficient of 0.81.*

Jensen (1972a, pp. 173-174):

(5) ...studies of foster children which show that *the single most important factor in the child's* environment *with respect to his intellectual development is his foster mother's IQ. This variable has been shown to make the largest* independent *contribution to variance in children's IQs of any environmental factor (Burks, 1928).*

Jensen (1973a, pp. 196-197):

(6) *In a classic study, Burks (1928) estimated the effects of environment on IQ from an analysis of correlations between detailed ratings of the home environment and the IQs of adopted children. A multiple correlation (corrected for attenuation) between the actual environmental ratings and IQ was 0.42.* (The correlation between IQ and the theoretical environmental scale derived in our own twin

(7) study is 0.32). *Burks concluded from her analyses of the IQs and environments of adopted children that*

> *1. The total effect of environmental factors one standard deviation up or down the scale is only about 6 points, or, allowing for a maximal oscillation of the corrected multiple correlation (0.42) of as much as 0.20, the maximal effect almost certainly lies between 3 and 9 points.*
>
> *2. Assuming the best possible environment to be three standard deviations above the mean of the population (which, if 'environments' are distributed approximately according to the normal law, would only occur about once in a thousand cases), the excess in such a situation of a child's IQ over his inherited level would lie between 9 and 27 points -- or less if the relation of culture to IQ is curvilinear on the upper levels, as it well may be. (Burks, 1928, p. 307).*

(8) *The geneticist Sewall Wright (1931) later performed a*
genetical analysis, using his method of 'path coefficients,'
on Burks' data. He showed that Burks' correlation between
environment and adopted child's IQ could be broken down into
two components: the direct *effect of home environment on*
IQ and the indirect *effects of the foster parents' IQ on the*
child's environment. The direct *correlation of home environ-*
ment and child's IQ was 0.29; that is, about 9 percent of the
IQ variance was attributable to variance in home environments,
(9) independently of the intelligence of the foster parents. *The*
SD of these environmental effects thus would be equivalent
to 4.39 IQ points and the total reaction range of home environ-
ments on IQ would be approximately this value multiplied by
the number of SDs in a normal distribution, or 4.39 × 6 = 26.34
IQ points. (If the indirect effects of foster parents' IQ is
included with the direct effects of home environment, the total
(10) *reaction range is 36 IQ points). The occupational status of*
the foster parents in Burks' study spanned a wide range,
from professional to unskilled labor although a majority
were in occupations that would be classified as middle- and
upper-middle SES. The reaction range of 26 means, in effect,
(11) that *improvement of a child's home environment (without*
changing his parents' IQs) would raise the IQ 26 points for
those children who shortly after birth are moved from the
most unfavorable environment in a thousand to the most

favorable environment in a thousand. A gain of 36 points would occur if, in addition, *the child exchanged the 'worst' parents in a thousand for the 'best' parents in a thousand.*

Jensen (1973a, pp. 202-204)

Because of the lack of independence among environmental variables, we need more studies of the multiple correlation (R) between environment and IQ. Environmental measures such as family income, father's occupation, or some composite index of SES are commonly regarded as excessively 'crude' measures of the environment, with the implication that these measures fail to include important influences on IQ caused by more subtle and refined environmental variables. The important question, however, is how much more of the IQ variance is accounted for by the subtle environmental factors over and above the IQ variance already accounted for by a 'crude' environmental index, such as SES? Could one find more than five or six environmental measures which independently add significantly increments to the multiple correlation with IQ? In a study of the correlation between adopted children's
(12) IQs and environmental factors, *Burks (1928) found a correlation of 0.33 between the children's IQs and their family's income. When two quite elaborate and detailed ratings of the home environment (Whittier Home Index and Culture Index) were included, along with family income, in a multiple correlation, the resultant R was just 0.34, just 0.01 greater than for*

income alone. Similarly, mothers' vocabulary correlated with the adopted children's IQs 0.249; the multiple R between mother's vocabulary + mother's mental age + mother's education and children's IQs was 0.254. The multiple R between children's IQs and a number of environmental factors, which taken singly had correlations with children's IQs between 0.15 and 0.30, was only 0.35 (0.42 corrected for
(13) *attenuation). Significantly higher correlations between environment and the parents' own children are obtained, because parental intelligence is correlated with the environment and the children. The multiple R between the several environmental variables and children's IQs was 0.61. But since the correlation between mid-parent intelligence and child's IQ is 0.60 and between parental intelligence and environmental rating is 0.77, most of the correlation between child's IQ and environment is attributable to the parents' intelligence and the genetic correlation between*
(14) *parents and children. The multiple correlation of the environmental indices with children's IQs when the parental*
(15) *contribution is removed is only 0.183.* Even *in the case of the adopted children, the single most important environmental factor contributing to variance in children's IQs was the*
(16) *foster mother's intelligence. The single best index of the quality of the environment is probably mid-parent intelligence, since in Burks' study it correlates 0.77 with a very elaborate composite index of the quality of home environment.*

Jensen (1973a, p. 240):

The environmental contribution of parental IQ can best be assessed by means of adopted or foster children, since there is little or no genetic correlation between foster
(17) children and their foster parents. *In a study of this kind by Burks (1928), it was found that the* total *environmental contribution to the IQs of the foster children was only 17 percent* (which is close to $1 - h^2$ when h^2 is based on twin
(18) studies). *The independent* environmental *contribution of parents' intelligence (mother and father combined) was about 3 percent. Burks (1928, p. 301) states: 'We should not expect this* environmental *contribution of parental intelligence to be over four or five percent, however, because the correlations (even when corrected for attenuation) between child's IQ and foster parents' M.A. (mental age) are so very low.'* The correlation was 0.09 for foster father and 0.23 for foster mother.

3. REPRESENTATIVENESS

We begin with Jensen's items (1) and (10) which suggest that Burks's families were representative of the United States white population. Consulting Burks (1928), we find that her adoptive and control samples were confined to English-speaking couples residing in the San Francisco, Los Angeles, and San Diego areas, who were American-, British-, or north-European-born, and who were neither black nor Jewish (p. 230). Eighty-three percent of the adoptive families owned their own home (p. 268). On the 25-point "Whittier Index" of home quality, the adoptive families' average score was 23.3 (p. 269); more than one-third of the adoptive children had private tutoring in "music, dancing, drawing, etc." (p. 270). In intelligence, the adoptive parents averaged one standard deviation above the population mean (p. 305). As for "the total complex of environment," Burks's own conservative estimate was that the foster homes averaged between one-half and one standard deviation higher than the general population (p. 306).

To supplement these remarks, I have constructed Table 1, which provides a rough comparison of the occupational distribution in Burks's samples with that in the general population. Note that over half of the adoptive fathers were professionals, business owners, or managers.

And yet Jensen would have us believe that these families formed a broad cross section of American whites.[1]

Table 1

OCCUPATIONAL DISTRIBUTIONS IN BURKS (1928) AND IN THE U.S. (1930)

	(1)	(2)	
Occupation	United States	Burks Foster	Burks Control
A. Professional	7%	17%	20%
B. Managers and proprietors	7	39	32
C. Clerical	9	10	14
D. Skilled labor	13	15	11
E. Salesmen	6	8	11
F. Farmers	12	6	5
G. Semi-skilled labor	16	1	5
H. Laborers and service	30	3	3
	100%	99%	101%

Sources:

(1) U.S. Department of Commerce, Bureau of the Census, Historical Statistics of the United States, Colonial Times to 1957, Washington: Government Printing Office, 1960, pp. 75-78. Occupation of economically active population. A = professional, technical, and kindred workers; B = managers, officials and proprietors (ex. farm); C = clerical and kindred workers; D = craftsmen, foremen, and kindred workers; E = sales workers; F = farmers and farm managers; G = operatives and kindred workers; H = private household workers + service workers (ex. private household) + farm laborers and foremen + laborers (ex. farm and mine).

(2) Burks (1928, p. 267), Occupational classification of fathers. A = professional (ex. teachers) + teaching; B = business owners and managers; C = commercial employees; D = skilled labor; E = salesmen; F = ranchers + retired; G = semi-skilled labor; H = unskilled labor.

Since the two sources do not use the same occupational classification, this table is only approximate. A closer match of the categories might be made by using the detailed job titles given in Historical Statistics and the illustrative job titles given in Burks.

4. MULTIPLE CORRELATIONS

In items (2), (6), (12), (17) Jensen informs us that when Burks regressed the adopted children's IQs on a long list of environmental variables, she found a multiple R^2 of .17 or .18 (or R = .42).

No such regression was computed by Burks. Her R^2 in fact refers to the regression of child's IQ on the following four variables: father's IQ, father's vocabulary, mother's vocabulary, and income (pp. 386-387). Before arriving at this formulation she did experiment with five additional explanatory variables: mother's IQ, father's education, mother's education, Whittier index, and Culture index. The Whittier index of home quality was the sum of scores on five 5-point items: necessities, neatness, size of home, parental conditions, and parental supervision. The Culture index was also the sum of scores on five 5-point items: parents' vocabulary, parents' education, interests of parents, home library, and artistic taste.[2] Computational facilities being what they were at the time, Burks limited herself to observing that multiple Rs using several of the five additional variables along with one of the four included variables were only slightly larger than the simple r with the included variable (p. 287). Her procedure is adequately described in Jensen's item (12). On p. 287, she expressed the conviction that "The variables finally employed no doubt yield values for the multiple correlations that attain, within one or two points in the second decimal, to what the values would have been had we used all nine variables." But we cannot verify this at present because she did not provide a full set of correlations.

With respect to Jensen's item (2), we remark that Burks's interviewers did ask about "the home instruction or attention received by the child in such matters as reading or writing, story-telling to child, number work, or nature study" (p. 229); that she tabulated the means and standard deviations for the total number of hours spent in this group of activities at various age levels (p. 269); that she reported the correlation of this variable with child's IQ (p. 278); and finally that she did not use this variable in the multiple correlations, not even experimentally.

In any event, it is worth repeating that the "detailed and fine-grained" environmental measures which, according to Jensen, accounted for 17 percent of the variation in IQ scores, turn out to be: father's IQ, father's vocabulary, mother's vocabulary, and income.

5. PARENTAL INTELLIGENCE

Jensen tells us in items (5) and (15) that of all Burks's environmental variables, it was mother's IQ that had the largest correlation with adopted child's IQ.

This is simply not true. On p. 278 Burks tabulated the simple correlations of some twenty-five environmental variables with adopted child's IQ. Among the entries are: mother's vocabulary .23, Whittier index .21, Culture index .25, income .23, home-ownership .25, number of books in child's library .32. For mother's mental age (that is, IQ) the entry is .19. Again on p. 285 she tabulates the simple correlations (now corrected for attenuation) of ten environmental variables with adopted child's IQ.

Among the entries are mother's vocabulary .25, Whittier index .24, Culture index .29, income .26. For mother's mental age, the entry is .23.

Now Jensen uses the adjective "independent" in (5), which suggests that he may be referring to partial rather than simple correlations. I cannot locate such partial correlations in Burks, nor can I find anything else in Burks to support Jensen's assertion. Indeed, as Jensen himself reports in (12), she found that mother's IQ adds little once mother's vocabulary has been introduced as an explanatory variable.

We proceed to item (18) which claims that the independent environmental contribution of parental IQ to child's IQ was about 3 percent. In the context of the sentences that precede it, this item appears to tell us that when mother's and father's IQs were dropped from the list of variables explaining adopted child's IQ, the R^2 fell by .03 from .17. As we already know, mother's IQ was not included in that multiple regression; nor can I locate any other regression in Burks that produces the 3 percent figure. If we read (18) in the context of the sentence which follows it, we get the impression that Burks calculated 3% and then compared it with the 4 or 5% obtained in some other regression. Actually, the latter figure was computed as follows (pp. 301-302). For the adoptive families, the simple correlations of child's IQ with father's and mother's IQ were .09 and .23. Summing the squares of these, and making an arbitrary deduction to allow for the fact that some of this correlation is not causal but merely attributable to the correlation of parental IQ with other environmental factors, she arrives at "four or five percent". Whatever be the merits of Burks's arithmetic. I see no route by which Jensen can have arrived at three percent.

Now consider item (13), which purports to describe the results of a multiple regression for Burks's control group--which consisted of "natural" (i.e. non-adoptive) families.[3] Let C = child's IQ, P = parental IQ, and E = set of environmental variables. Jensen appears to say that with $r_{CP} = .60$ and $r_{PE} = .77$, the multiple correlation of C on P and E was $R_{C(P,E)} = .61$. Where do his figures come from?

On p. 287 Burks gives .61 as the control group multiple correlation of child's IQ on: father's IQ, mother's IQ, father's vocabulary, and the Whittier index; but the intercorrelations among the explanatory variables are not given there. We turn instead to pp. 300-301 where she reports and analyzes a control group multiple regression of child's IQ on two explanatory variables: midparent IQ and the Whittier index. From her presentation we can extract $r_{CP} = .6036$, $r_{PE} = .7653$, $r_{CE} = .4771$, and thus $R_{C(P,E)} = .6041$. Since the first two correlations round off to .60 and .77, and the multiple correlation rounds off to almost .61, we may have located Jensen's source.

But note that E now contains only the Whittier index, a single measure of environment. This is hardly compatible with the characterizations that Jensen has scattered so liberally through the paragraph in which item (13) appears: "subtle environmental factors," "five or six environmental measures," "elaborate and detailed ratings of the home environment," "a number of environmental factors," "the several environmental variables," "the environmental indices."

Item (14) also refers to the control group regression. It seems to say that, after controlling on P, the multiple correlation of C and E, that is $R_{CE\cdot P}$, equals .183. Now, when P is removed from the regression above, only a single explanatory variable remains, namely the Whittier index E. Thus "<u>multiple</u> correlation of the environmental <u>indices</u>" is a peculiar description. Furthermore, the partial correlation of C and E after controlling on P is not .183, but rather

$$r_{CE\cdot P} = (r_{CE} - r_{CP}r_{PE})/\sqrt{(1-r_{CP}^2)(1-r_{PE}^2)} = .030.$$

Where in the world did Jensen find .183?

After diligent search, I have arrived at the following conjecture. With all variables standardized, Burks (p. 301) obtains the partial regression coefficients ("beta-weights") $b_{CP\cdot E} = .5757$ and $b_{CE\cdot P} = .0367$. She then decomposes the multiple R^2 into

$$R^2_{C(P,E)} = b^2_{CP\cdot E} + b^2_{CE\cdot P} + 2\, b_{CP\cdot E}\, b_{CE\cdot P}\, r_{PE}$$

$$(.6041)^2 = (.5757)^2 + (.0367)^2 + 2\,(.5757)(.0367)(.7653)$$

$$.3649 = .3314 + .0013 + .0322.$$

She labels the three terms on the right as: "parental contribution," "contribution of environment <u>other</u> than parental intelligence," and "joint parental and environmental contribution over and above separate contribution of each." If we sum the last two terms -- or equivalently subtract the first term from the left-hand side -- we get .0013 + .0322 = .3649 - .3314 = .0335, which is precisely the square of .183. I have no idea why Jensen believes that this measures the correlation of C and E when P is removed.

As far as I can see, $\sqrt{R^2_{C(P,E)} - b^2_{CP\cdot E}}$ does not measure any correlation whatsoever.

With item (16) we reach the close of Jensen's remarkable paragraph, which, incidentally, appears in a chapter he entitles "Multiple and partial correlation methods." Here we meet $r_{PE} = .77$ once again. This time its magnitude is offered as evidence that midparental intelligence is the "single best index of the quality of the home environment." But surely E is even better than P as an index of E?[4]

6. HERITABILITY ANALYSIS

Items (4) and (8) refer to the analyses of Burks's data that were undertaken by the distinguished geneticist Sewall Wright.[5]

In (8), Jensen would have us believe that Wright decomposed the adoptive group r_{CE} into a direct effect of E on C, and an indirect effect of P on E. Wright did no such thing, nor would it make any sense to do so. What Wright (1931, p. 160) did was construct a simple model in which child's IQ, C, is directly determined by environment, E, and heredity, H (an unobserved variable). For the adoptive children, he assumes that H is uncorrelated with E and with parental IQ,P, while E and P are correlated with each other. In this model, we can express r_{CP} as the product of r_{PE} and the "path coefficient" (i.e., beta-weight) running from E to C. But no decomposition of r_{CE} is possible, and indeed the .29 which Jensen reports so emphatically as the <u>direct</u> correlation between E and C is simply r_{CE} itself.

In (4), Jensen tells us that Wright produced .81 as the estimate of heritability (= proportion of variance in IQ accounted for by variation in heredity) from Burks's data. What Wright actually did can be sketched as follows. For the control children, child's IQ is again directly determined by E and H, but now H, E, P are all intercorrelated. Taking the adoptive-group and control-group equations along with five observed correlations and several plausible assumptions, Wright obtains .90 as the estimate of the path coefficient running from H to C. And the square of this, namely .81, estimates the proportion of the variation in IQ that is attributable to variation in heredity. So far, so good.

However, as Wright observes, this model attributes to heredity, H, which is not measured, all effects that cannot be attributed to measured environment. If so, the heritability estimate may be sensitive to the choice of a measure for E. Indeed, a simple manipulation of Wright's (1931, p. 160) formulas will show that his estimate of p, the path coefficient running from H to C, is calculated as

$$p = \sqrt{1-q^2}\,(-q\,r + \sqrt{q^2r^2 + 1-2q^2})/(1-2q^2),$$

where q and r are, respectively, the adoptive-group and control-group correlations of child's IQ with environment. Thus in his model, the estimate of p is completely determined by the two r_{CE}'s. Now, the environmental measure that Wright used was the Culture index, a single variable reflecting certain aspects of the parents' vocabulary, education, interests, home library, and artistic taste. With that measure for E, he has q = .29 and r = .49, and the formula above gives p = .90. But there is nothing sacred about the Culture index as a measure of environmental influences on intelligence, so there is nothing sacred about .29 and .49 as values for q and r.[6] For example, we have already seen that Burks found an adoptive-group multiple correlation of .42 between C and a set of four environmental variables, and that she also found a control-group multiple correlation of .61 between C and a slightly different set of four environmental measures. For illustrative purposes, we can take q = .42 and r = .61 as values for the correlations of child's IQ with environment. When these new values are inserted in the formula above, we find p = .82; that is, we get $p^2 = .68$ rather than $p^2 = .81$ as our estimate of heritability. It is not surprising to find that a more refined measure of environment leads to a lower estimate of heritability, in a model that attributes to heredity all effects that are not attributable to measured environment.

Moreover, in the same nine-page article, Wright (1931, pp. 161-163) provides a lower estimate of heritability from Burks's data. The lower estimate comes from a second model in which environment is still measured by the Culture index alone, but the effects not attributable to measured environment are allocated between G (additive genotype) and M (a residual that includes non-additive genotype and genetic-environment interactions along with unmeasured environment). The path coefficient running from G to C is estimated as .71; squaring this yields .49 as the second estimate of heritability. To some extent, the reduced value arises because of the switch from broad to narrow heritability. But Wright does not rationalize it in that manner. Rather (p. 162) he clearly states that the first estimate is intended as an upper bound, the second as a lower bound. On two subsequent occasions, in reviewing his analysis of Burks's data, he emphasized this point: Thus, Wright (1934, pp. 185-188) wrote:

> [The first model is] doubtless too simple since heredity is represented as the only factor apart from the measured environment. Any estimates of the importance of hereditary variation will thus be maximum...., [In the second model, we] attempt at obtaining a minimum estimate of heredity. ... The path coefficient for influence of hereditary variation lies between the limits + .71 (if dominance and espistatis are lacking) and + .90.

And Wright (1954, p. 23) wrote

> The results are reasonable [for the first model] except that H undoubtedly includes more than heredity ....

It is strange that Jensen was able to collapse Wright's elaborate analyses into an unqualified conclusion that the heritability coefficient was 0.81.

7. ENVIRONMENTAL EFFECTS

The remaining items directly concern the implications of Burks's study for social policy.

In items (3) and (7) Jensen reports Burks's own conclusions about the potential effects of environmental change upon intelligence. Her basic estimate, namely that a standard-unit change in environment would produce a 6-point change in IQ, was obtained as follows (pp. 306-308). An IQ-environment correlation for adopted children, namely the now-familiar multiple R of .42, was interpreted as a standardized regression coefficient: changing environment by one standard unit will change IQ by .42 standard units. Then multiplying .42 by the standard deviation of IQ scores, namely 15 points, gave 6 points. Her alternative estimates, namely 3 and 9 points, were calculated in the same manner, except that .22 and .62 were used, arbitrarily, instead of .42. Finally, she multiplied 3 and 9 by three to depict the effects of a three-standard-unit change, arriving at 9 and 27 points respectively.

In item (9), Jensen has refined Burks's arithmetic. He is using .29 (the simple correlation of adopted child's IQ with the Culture index) in place of .42, and 15.1 in place of 15 as the standard deviation of IQ scores, and thus gets 4.39 in place of Burks's 6 points. He then multiplies 4.39 by six to depict the effect of a six-standard-unit change, which brings him to the marvelously precise figure of 26.34 points. The basis for his alternative figure of 36 points escapes me.

Finally, we have item (11), which is Jensen's vivid portrayal of a six-standard-unit change in environment, since "one in a thousand" is the probability that a normal variable lies more than three standard deviations above (or below) its mean.

It is hard to take this arithmetic seriously. "The environment" is being measured by income and three test scores (Burks) or by a single crude index (Jensen). Putting that aside, the inferences are being made from a nonrepresentative sample. In constructing their estimates, Burks and Jensen implicitly take the sample standard deviation as the unit of measurement for environment, yet their conclusions purport to tell us about the population. If environmental variation was substantially less in Burks's samples than in the population at large, the Burks-Jensen arithmetic will inevitably lead to substantial understatements of the potency of environmental change.

As we have seen, Burks's samples were not at all representative of the population, having been selected from the upper ranges of the environmental distribution. Variation within those upper brackets is presumably less than it is across the full distribution. To suggest orders of magnitude, let us use Burks's own guess that in her samples the total complex of environment averaged between one-half and one standard deviation above the population mean. In a normal distribution with mean μ and standard deviation σ , we get a group in which the mean is $\mu + (1/2)\ \sigma$ by selecting the top 69% of the distribution; the standard deviation within that group is $.7\sigma$; see Kelley (1947, pp. 295-298) for the relevant formulas. Thus a conservative guess might be that the standard deviation of environment in Burks's samples was .7 as large as it was in the general population. If so, a population standard unit was 1.4 times as large as a sample standard unit, and we need not hesitate to raise the Burks-Jensen estimates of environmental effects by, say, 50%, on this ground alone. (Or, for that matter, if we take the mean in Burks's samples to be $\mu + \sigma$ rather than $\mu + (1/2)\sigma$, the same argument would lead us to double the Burks-Jensen estimates.) An environment that was the "most unfavorable... in a thousand" in Burks's samples may not have been all that extreme in the population.

To replace our conjectural arithmetic, it would be nice to have direct information on the truncation of environmental variation in Burks's data. But such information is rather difficult to come by. She presents sample standard deviations for many of her variables, but the corresponding population values are not readily available. There are a few isolated exceptions. The Barr occupational scale "comprises the combined judgment of thirty raters upon the grade of intelligence which each of 100 representative occupation demands on the average"; its standard deviation for Burks's adoptive families was about 75% as large as it was in the California communities from which her families were drawn (pp. 249, 255, 274). For the Whittier index, I have been unable to locate population figures. But for each of its five component 5-point items, the sample means are so high and the sample standard deviations so low as to indicate that virtually all the families scored at the 4- or 5-point level (p. 269).

With respect to income variation, the evidence that I have displayed in Table 2 appears to point in a contrary direction. The high means and medians confirm that Burks's families came from the upper socioeconomic brackets, but the high standard deviations seem to say that environmental variation was amplified rather than truncated. To resolve this point, we should recognize that the income variation in Burks's samples occurred at high income levels. There is no reason to presume that a change from say $10,000 to $15,000 income is as stimulating to children's IQ as a change from $1,000 to $6,000. In economic jargon, it is plausible that there are "diminishing returns" to increases in income, so that the responsiveness of IQ to income changes is less at high income levels than it is at low- and middle- income levels. If so, the large variation of income when measured in dollars is quite consistent with a small variation of income when measured in IQ-relevant units.[7]

Of Burks's adoptive families, about 63% had one child, 24% two children, and 13% three children (pp. 270, 276). Thus, the number of siblings, which is presumably a relevant environmental variable, seems to have been less variable in Burks's sample than in the general population.[8] All of Burks's families were intact, that is both parents were alive and living together; this aspect of the environment, which is conceivably relevant to children's achievement, must have shown <u>some</u> variation in the population at large. Another factor that we may presume the adoptive families had in common is one that not all families share: the desire for a child.

Table 2

INCOME STATISTICS IN BURKS (1928) AND IN THE U.S. (1929)
(Income measured in thousand dollars)

	(1)	(2)	
	United States	Burks Foster	Burks Control
Median	1.7	3.6	3.0
Mean	2.3	6.2	4.1
Standard Deviation	2.3	7.4	3.1

Sources:

(1) My calculation from tables in Historical Statistics of the United States, pp. 165-166, using interpolation and price level adjustment.

(2) Burks (1928, p. 268).

Reasonable men may differ in the weights they attach to these various bits of evidence concerning environmental variation in Burks's samples. However, there is no doubt that the environments provided by her families failed to represent those provided across the population at large. The burden of proof rests on Jensen who wishes to persuade us that the responsiveness of IQ to environment in a nonrepresentative sample is indicative of its responsiveness in the population.

8. IQ DISTRIBUTIONS

Burks herself called attention to the implications of selectivity on p. 222, saying that

> It should be emphasized at this point that whatever tendencies and conclusions can be found in this study are valid only for populations as homogeneous in racial extraction, social standards, and educational opportunities as that from which are subjects are drawn. The distribution of homes of the children studied in this investigation was probably nearly as variable in essential features* as homes of the general American white population (though somewhat skewed toward a superior level). It was not as variable, however, as if the homes of southern negroes, poor mountain whites, or Philippine Negritoes had been included; and consequently, home environment cannot be expected to have as large a proportional effect upon the

> mental differences of the children we studied as though they were being reared in families unselected as to race or geographical location throughout the world.

Her contention that environment was fully variable in her samples runs counter to the many indications of superiority previously noted. The only evidence she offers is in the footnote to which the asterisk above leads:

> *This seems probable because the variability in intelligence of both the control and foster children coming from these homes is as large as that of unselected children.

Her reasoning, presumably, is that if environmental variation had been limited in her sample, and if environment is an important determinant of IQ, then the variation of her children's IQ test scores would have been limited as well.

The IQ test that Burks used was the 1916 Stanford-Binet. For this test, the only "population" data that I have located are those in Terman *et al*. (1917). They refer to the original sample on which the test was standardized -- 905 school children aged 5-14 years. This spans the same age range as Burks, and we may take Terman's IQ distribution as the population against which Burks's is to be assessed.

Table 3 sets out the data. We note that mean IQ was somewhat higher in Burks's samples than in the "population", while (as Burks had remarked) the standard deviation was about the same.[9] In view of the many indications

Table 3

IQ DISTRIBUTIONS IN TERMAN (1917) AND IN BURKS (1928)

(1) Terman		(2) Burks		
IQ Bracket	Percent	IQ Bracket	Foster Percent	Control Percent
		35-44	1%	
		45-54	1	
56-65	*	55-64	0	
66-75	2%	65-74	1	
76-85	9	75-84	2	2%
86-95	20	85-94	11	5
96-105	34	95-104	27	17
106-115	23	105-114	28	22
116-125	9	115-124	19	29
126-135	2	125-134	7	12
136-145	1	135-144	1	9
		145-154	1	3
		155-164	1	1
	100%		100%	100%
Mean	101		107	115
Standard Dev.	15		15	15
N	905		214	105

* = less than one-half

Sources:

(1) Terman (1917, pp. 40, 42): Distribution of intelligence quotients of 905 unselected children, ages 5-14 years. Mean and standard deviation calculated by me from Terman's frequency distribution.

(2) Burks (1928, p. 264): Intelligence distribution of children, in I.Q. Mean and standard deviation reported by Burks.

Burks's table is in terms of five-point intervals; I have aggregated them to facilitate comparison with Terman, whose table is in terms of ten-point intervals. Note that the interval end-points are not quite the same in the two sources.

of superior environment, the high mean is not surprising. But the untruncated standard deviation is puzzling if we believe that environment is a major influence on IQ scores.[10]

A closer look at the Terman study (pp. 32 - 41) reveals that the 1916 Stanford-Binet test was not fully standardized for age, and that the age distribution in Terman's group was substantially different from that in Burks's samples. That opens up the possibility that the 15-point standard deviation in Burks was something of an artifact, being the result of a mixture of age-specific means and standard deviations. To explore this possibility I have constructed Table 4, which gives the means and standard deviations of IQ by age in Terman along with the age distributions in Terman's group and in Burks's samples.[11] The mean IQ has a downward trend, and the standard deviations fluctuate. We can generate a hypothetical population by using Burks's age distribution in conjunction with Terman's age-specific means and standard deviations.[12] If this is done one finds that about 4 points in Burks's means and about 1 point in her standard deviations are attributable to the age composition, primarily to the over-representation of 5-year olds. That is to say, if Terman's children had had the age composition of Burks's samples, their IQ mean would have been 105 (rather than 101) and their IQ standard deviation would have been 16 rather than 15).

After these admittedly crude calculations our puzzle remains. If environment is a major influence on IQ scores and if the environment in Burks's samples was as selective as we have argued, why didn't her children's IQs average still higher and vary still less than they did, as compared with an unselected group?

Table 4

IQ AND AGE IN TERMAN (1917) AND BURKS (1928)

	(1)		(2)	(3)	
	Terman		Terman	Burks Foster	Burks Control
Age	IQ Mean	IQ St. Deviation	Age Distributions		
5	111	14	6%	30%	28%
6	104	13	13	12	14
7	104	12	10	9	10
8	102	12	11	14	13
9	100	12	12	11	7
10	104	12	10	8	8
11	102	15	9	5	7
12	100	16	9	5	7
13	97	14	11	4	5
14	98	11	9	2	1
			100%	100%	100%

Sources:

(1), (2) Terman (1917, pp. 33-37). My calculations from Terman's histograms.

(3) Burks (1928, pp. 263).

A partial answer may be provided if we take a closer look at Terman's sampling design. Consulting Terman (pp. 10-11, 28-30), we find the following. Terman's children were all in school, residing in the San Francisco Bay, Los Angeles, Santa Barbara, and Reno areas. All were within two months of a birthday. The schools were in communities of "average social status" and were "middle-class". Furthermore:

> few children attending them were either from very wealthy or very poor homes. The only exception to this rule was in the case of Reno.... The large majority [even there]... were from homes of average wealth and culture...
>
> ...None of the children was foreign-born and only a few were of other than Western European descent...
> Spanish, Italian and Portuguese children were eliminated from our study of distribution, for the reason that in western cities children of these nationalities are likely to belong to unfavorably selected classes. We are justified in believing, therefore, that the distribution of intelligence among our subjects is less influenced by extraneous factors than has been the case in other studies of this kind.

It seems fair to conclude that Terman's "unselected" group was itself drawn from homes with environments that were better and less variable than those in the general American population. If so, the fact that the

IQ distribution in Burks's samples was not much different from that in Terman is consistent with the position that environment is a major influence on IQ scores that did not receive its due in Burks's samples.[13]

9. ANOTHER STUDY

As we have seen, Jensen has made repeated use of Burks's study to support his position that environment plays only a minor role in the determination of intelligence. In the same context he has used two other studies of adopted children's intelligence, Leahy (1935) and Skodak and Skeels (1949); see Jensen (1972a, pp. 15-17, 129, 154, 213-214; 1973a, p. 241; 1973b). But one such study is missing from his reports, namely a 115-page article by Freeman, Holzinger, and Mitchell (1928).[14]

Is it possible that the Freeman article did not meet the stiff criteria that Jensen set out in his first paragraph? In the Freeman study of adoptive families in Illinois, the sample size was similar to those in Burks and Leahy, considerable detail on home environment was obtained, and the occupational distribution was no less representative than those of Burks and Leahy. Freeman _et al_. consider selective placement (pp. 179-185); their evidence against its having occurred is rather similar to that in Burks (pp. 248-254). The Freeman study did not include a control group. Furthermore the Freeman children were placed at later ages than the Burks and Leahy children, and included black children placed in black families. Thus Jensen may have set the Freeman study aside on the grounds that selective placement was operating.

Consulting the Freeman article suggests an alternative explanation of Jensen's failure to cite it: The IQ-environment correlations ran somewhat higher than in the Burks sample. Specifically, on pp. 177-179, Freeman et al. report the following simple correlations with adoptive child's IQ: Father's IQ .37, mother's IQ .28, father's occupation .37, mother's vocabulary .37, parents' education .42, and parental rating (a single scale somewhat similar to the Whittier index) .49.[15]

10. ANOTHER SCHOLAR

In the great IQ debate, Jensen's unreliable report of the Burks study has acquired a life of its own. For example, Herrnstein's (1973, pp. 182-184) treatment, which I have discussed elsewhere (Goldberger, 1974), is rather reminiscent of Jensen's.

Another scholar who has adopted Jensen's report is H. J. Eysenck. In his 1971 book, Race, Intelligence, and Education, Eysenck wrote:

> In a famous study on these lines Burks spent between four and eight hours in investigating each adoptive home, very carefully rating all environmental variables which had been suggested as possibly relevant to the determination of high IQs. He included the adopting parents' intelligence as part of the children's environment, as well as such factors as the amount of time the

> parents spent helping the children with their school work, the amount of time spent reading to them, and so on. The proportion of IQ variance accounted for all these environmental factors combined was 18%, which agrees well with the figure of 80% for the influence of heredity; the two add up to just about 100%. It should perhaps be added that the population sampled in this study was broadly representative of the American white environments, excluding only perhaps an extreme 5%; thus it cannot be said that these results are due to a lack of variability in environmental determinants. (pp. 63-64)

More recently, in his 1973 book, The Measurement of Intelligence, Eysenck wrote:

> The point of Burks' paper is a very simple one. Having located foster children assigned on what amounts to a random principle to their foster parents, she looked into the circumstances prevailing in the foster home, taking great care to include in her survey as many measurable features of the environment as possible; she then correlated these features with the IQ of the children involved, to determine the degree to which these features could be said to determine IQ. She also combined all the environmental aspects to determine the total amount which they might be

> said to contribute to IQ variance; the figure she arrived at was 17%. Thus the most thorough study of the influence of environmental variation on IQ variance gives a figure which neatly complements the 80% figure for genetic influence. (pp. 290-291)

Apart fro... remarking that by 1973 Eysenck had read Burks's article and correctly determined her sex, we forgo further comment.

11. CONCLUSION

We have dissected Jensen's treatment of Burks because it occupies a central place in his argument that environmental improvement will not succeed in raising intellectual ability. The low IQ correlations found for genetically unrelated individuals on the one hand and the high IQ correlations found for genetically identical individuals on the other hand, constitute the bulk of the evidence for his argument. It appears that Jensen's report of the Burks study is unreliable, and that the Burks study itself cannot support strong conclusions. Similar problems arise with respect to the other kinship studies, as Bronfenbrenner (1972) and Kamin (1974) have demonstrated.

Suppose that Jensen, instead of writing the long report that we reproduced in Section 2, had summarized the content and implications of the Burks study for us as follows:

> About a half-century ago, 200 white children who had been adopted by middle- and upper-class families in California were tested. Correlating the children's IQ scores with their parents' income, IQ, and vocabulary scores produced an R^2 of only .17. Taking this in conjunction with similar evidence found in similar studies, and suppressing the contrary evidence found elsewhere, we must conclude that environmental improvement cannot succeed in eliminating racial differences in IQ.

If Jensen had written that, where would the great IQ debate be today?

FOOTNOTES

[1]Does the Leahy study cited by Jensen compensate for the limitations of Burks? Leahy's observations covered about 200 foster families and a corresponding number of matched control families. All were nonfarm residents of Minnesota, c^f north-European extraction, and non-Jewish. Forty percent of the fathers were professionals or business managers, twelve percent were slightly-skilled or day laborers (p. 279). Leahy (p. 259) stated that

> In our earliest considerations of a population we conceived a research group which would sample the population of adoptive homes distributed from a socioeconomic standpoint as male occupations are distributed in the general population. Because of the limited number of children placed in homes of the laboring class this plan had to be abandoned.

We have seen that about 5% of Burks's samples, and none of Leahy's, were farm families; over 20% of the American population lived on farms during the 1920-1930 s.

[2]The detailed scales were given by Burks (pp. 231-235); some excerpts can be found in Goldberger (1974). At the risk of slight exaggeration, we may say that removing family portraits from the walls and jazz from the record collection would have raised the Culture index as much as attending college for four years.

[3]Jensen's switch from the adoptive group in (12) to the control group in (13) may have escaped the reader; "the parents' own children" reads like the natural children of the adoptive parents. There were indeed seven cases in which Burks tested a natural child along with his adoptive sibling (p. 280), but Jensen can hardly have been referring to them.

[4]It is conceivable that Jensen has here misconstrued Wright's (1931, p. 161) statement that in Burks's data, "It appears that midparental IQ is a much better index of home environment than of child's heredity."

[5]For a survey of some of Wright's work and its relevance to causal modeling in the social sciences, see Goldberger (1972).

[6]Because a full set of intercorrelations were not provided by Burks, Wright felt compelled to employ only a single environmental variable.

[7]A simple way to formulate the diminishing-returns idea is to specify that IQ varies linearly with the logarithm of income rather than with income itself. Suppose further that log-income is normally distributed in the population. Then we can use the figures in column (1) of Table 2 to estimate the parameters of the log-income distribution in the U.S. population. Doing so, we obtain (roughly) $\mu^* = .5$ and $\sigma^* = .8$ as the mean and standard deviation of the natural logarithms of income. (For the relevant formulas, and for empirical evidence on lognormality, see Aitchison and Brown (1957, pp. 7-9, 87-90, Chapter 11).) After application of the truncated-normal formulas to this log-income distribution, the figures in columns (2) of Table 2 permit the following interpretation.

Burks's control-group families were essentially randomly drawn from the top half of the income distribution; her adoptive families were still more selective but also included a few outliers. (Burks herself remarked (p. 275) that there were "a few extremely high incomes" in the adoptive group.) The standard deviation in the top half of a normal distribution is .6 of its value in the full population. Thus the large sample variation in income is quite compatible with a small sample variation in logarithmic income. If the diminishing-returns idea is correct, then it is the latter truncation rather than the former amplification that is relevant to estimating income effects from Burks's data.

The careful reader may have noted that at the end of item (7) Jensen himself called attention to the possibility of nonlinear response. It is remarkable that he would have us believe that it implies that the sample-estimated effects may be biased upwards.

[8] Curiously enough, Burks did not use family size or number of siblings, as an environmental measure.

[9] I was surprised to find that Terman does not actually give the mean and standard deviation. To calculate those statistics I used the crude procedure that treats all observations in an interval as though they were located at the midpoint of the interval. On p. 42 Terman does tabulate a fitted normal distribution along with his empirical distribution, but fails to say what the μ and σ of the fitted distribution were. His entries for the fitted distribution are more or less consistent with a μ between 100 and 101, and a σ between 14 and 15.

[10]The pair of abnormally low-scoring adopted children account for a full point of their group's standard deviation. Presumably those two children were not in school; that points out one respect in which Burks's sample was less selective than Terman's. Jensen, it must be noted, does not mention the high IQ means in Burks, although he devotes an entire article (1973b) to explaining away the high IQ means found in the Skodak-Skeels (1949) study of adoptive children.

[11]Here again Terman does not provide the means and standard deviations, but only the histograms. I followed the procedure described in n.9. My calculations are thus only rough and were inhibited by the fact that there are internal inconsistencies in Terman's charts; for example, for 12-year olds (p. 36) the percentages add up to 107. Freeman et al. (1928, pp. 190-193) call attention to the inadequate standardization of the 1916 Stanford-Binet and to the inconsistencies in Terman's charts. Their tabulation (p. 191) of the age-specific means in Terman's group differs slightly from mine.

[12]Burks does not tabulate IQ by age for her samples; on p. 247 she reports the age-IQ correlations: -.10 for the adoptive children and +.09 for the control children.

[13]A final note on the 1916 Stanford-Binet: Burks (pp. 230-231) used this test also for the parents, with some adjustment to the official scale. If my reading of Terman (pp. 8-9, 49) is correct, the sample on which the test was standardized for adults consisted of 30 business men "of moderate success and of very limited educational advantages," and 32 high school

juniors and seniors aged 16 to 20. (Also tested were 150 migrating unemployed men who were temporary residents at a hobo hotel in Palo Alto; but their scores were apparently not used for standardization).

[14]This article appears in the same volume as -- indeed is the chapter which immediately precedes -- Burks's article. Data from the Freeman study do underlie so e of the medians given in Jensen's (1972, p. 124; 1973c) tables of kinship correlations.

[15]In summarizing their analyses, Freeman et al. (pp. 209-211) emphasized the strength of environment, while Burks (pp. 308-309) emphasized the strength of heredity. The Freeman sample also covered some natural siblings of the adopted children, and some pairs of adopted children; the significance of such data has recently been noted by Kamin (1974, pp. 123 - 124).

REFERENCES

J. Aitchison and J.A.C. Brown (1957), The Lognormal Distribution, Cambridge University Press.

U. Bronfenbrenner (1972), "Is 80% of intelligence genetically determined?," pp. 118-127 in Influences on Human Development, Hinsdale: The Dryden Press.

B. S. Burks (1928), "The relative influence of nature and nurture upon mental development: a comparative study of foster parent-foster child resemblance and true parent-true child resemblance," National Society for the Study of Education: Twenty-Seventh Yearbook: Nature and Nurture, Part I, Their Influence on Intelligence, Bloomington, pp. 219-316.

H. J. Eysenck (1971), Race, Intelligence, and Education, London: Temple Smith.

H. J. Eysenck (1973), The Measurement of Intelligence, Lancaster: Medical and Technical Publishing Co.

F. N. Freeman, K. J. Holzinger and B. C. Mitchell (1928), "The influence of environment on the intelligence, school achievement, and conduct of foster children," National Society for the Study of Education: Twenty-Seventh Yearbook: Nature and Nurture, Part I, Their Influence on Intelligence, Bloomington, pp. 103-217.

A. S. Goldberger (1972), "Structural equation methods in the social sciences", Econometrica, Vol. 40, November 1972, pp. 979-1001.

A. S. Goldberger (1974), "Mysteries of the meritocracy," University of Wisconsin Institute for Research on Poverty: Discussion Paper 225-74, October 1974.

R. J. Herrnstein (1973), I.Q. in the Meritocracy, Boston: Little, Brown.

A. R. Jensen (1972a), Genetics and Education, New York: Harper & Row.

A. R. Jensen (1972b), "The IQ controversy: a reply to Layzer," Cognition, Vol. 1, 1972, pp. 427-452.

A. R. Jensen (1973a), Educability and Group Differences, New York: Harper & Row.

A. R. Jensen (1973b), "Let's understand Skodak and Skeels, finally," Educational Psychologist, Vol. 10, 1973, pp. 30-35.

L. Kamin (1974), The Science and Politics of I.Q., Potomac: Erlbaum Associates.

T. L. Kelley (1947), Fundamentals of Statistics, Cambridge: Harvard University Press.

A. M. Leahy (1935), "Nature-nurture and intelligence," Genetic Psychology Monographs, Vol. 17, pp. 241-305.

M. Skodak and H. M. Skeels (1949), "A final follow-up study of one hundred adopted children," Journal of Genetic Psychology, Vol. 75, 1949, pp. 85-125.

L. Terman et al. (1917), The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence, Educational Psychology Monographs No. 18, Baltimore: Warwick & York.

S. Wright (1931), "Statistical methods in biology," *Journal of the American Statistical Association*, Vol. 26, March 1931, Supplement, pp. 155-163.

S. Wright (1934), "The method of path coefficients," *Annals of Mathematical Statistics*, Vol. 5, September 1934, pp. 161-215.

S. Wright (1954), "The interpretation of multivariate systems," Chapter 2, pp. 11-33 in O. Kempthorne *et al*., editors, *Statistics and Mathematics in Biology*, Ames: Iowa State College Press.