

Revisiting scientific and social debates about heritability in light of the under-recognized implications of heterogeneity

PETER TAYLOR

Programs in Science, Technology & Values and Critical & Creative Thinking
University of Massachusetts, Boston, MA 02125, USA peter.taylor@umb.edu

Pre-circulated paper for session on “Heterogeneity and heritability: Responses from sociology, philosophy, and history of science”
at the meetings of the Society for Social Studies of Science (with participants from the History of Science Society and Philosophy of Science Association), Vancouver, BC, 2-4 November 2006.

Version 8/26/06. Comments welcome.

1. Introduction

Claims that some human trait, say, IQ test score at age 18, show high heritability derive from analysis of data from relatives. For example, the similarity of pairs of monozygotic twins (which share all their genes) can be compared with the similarity of pairs of dizygotic twins (which do not share all their genes). The more that the former quantity exceeds the latter, the higher the trait’s “heritability.” Researchers and commentators often describe such calculations as showing how much a trait is “heritable” or “genetic.” However, no genes or measurable genetic factors (such as, alleles, tandem repeats, chromosomal inversions, etc.) are examined in deriving heritability estimates, nor does the method of analysis suggest where to look for them. Indeed, even if the similarity among a set of close relatives is associated with similarity of yet-to-be identified genetic factors, the factors may not be the same from one set of relatives to the next, or from one environment to the next. In other words, the factors may be “heterogeneous.” It could be that alleles, say, AbcDe, subject to a sequence of environmental factors, say, FghiJ, are associated, all other things being equal, with the same outcomes as alleles abCDE subject to a sequence of environmental factors FgHiJ.

Some prominent geneticists have noted that heritability estimates are not helpful in identifying specific genetic factors (e.g., Rutter 2002, 4), but the possible heterogeneity of factors has

not been recognized as a significant issue (Taylor 2006a,b,c). I believe that quantitative analysis in the study of heredity (and in social sciences more generally) needs to pay more attention to the possible heterogeneity of factors that underlie patterns in data on observed traits. Of course, as any student of science in its social context knows, ideas do not realize their transformative potential simply by being correct; changes in the social structure of the field are needed if the inevitable resistance from the mainstream is to be overcome. Moreover, alternative research programs have to be opened up before many researchers begin to shift—critique is not sufficient for a dominant paradigm to be abandoned. What role, then, can historical, sociological, and philosophical studies of science, technology and society (STS) have in such change? In this paper I identify a number of angles through which I and other STS scholars might bring attention in research and policy to the implications of heterogeneity (sect. 3). One of these angles leads me to highlight some scientific fields that open up room to pay attention to heterogeneity (sect. 4). I begin by providing an overview of the conceptual critique in which heterogeneity plays a central role.

2. Overview of a Conceptual Critique

There is long and politically charged history of scientific and policy debates about the heritability of IQ test scores and genetic explanations of the differences between the mean scores for racial groups.¹ Despite the attention given to these debates by researchers and other critical commentators, including philosophers of science, the implications of certain conceptual and methodological issues in quantitative and behavioral genetics—heterogeneity is one of these—have not been well appreciated. Before sketching these issues and their implications, let me remark on the expository challenge here: In this arena colloquial notions (“surely it makes sense that some traits are more influenced by genes than environment”) intersect with technical discussion of what data analyses would be needed to assess various claims. In my critique I want to examine issues of data analysis so as to identify where problems arise in making science out of colloquial notions. Yet I also

¹ For key points in the debate, see the 1969 Harvard Educational Review article by psychometrician Arthur Jensen, which elicited a critical response from, among others, the population geneticist, Richard Lewontin (1970a,b; 1974; Jensen 1970). Jencks and Phillips (1998) reviews recent research on the black-white test score gap and Parens (2004) provides an even-handed overview of past and potential contributions of human behavioral genetics to discussions of social importance well beyond IQ tests.

want the discussion to be accessible to non-specialists because under-appreciated technical limitations point to conceptual issues worthy of consideration by all of us, especially, how we think about responsibility and causation when we talk of individuals in terms of their group membership. I am not claiming that the expository challenge is met by the overview to follow—nor by the brief illustrated talk or the more detailed expositions in Taylor (2006a,b,c,d). To make progress towards that end, I seek more give and take with various audiences.

Core Issues

I identify five issues with individual or joint implications that are under-appreciated:

1. effects distinct from measurable factors: Because the statistical Analysis of Variance (AOV) of observed traits involves no reference to measurable genetic or environmental factors involved in the development of those traits, the quantities (“genetic” and “environmental” “effects”²) estimated by an AOV cannot be equated with such factors;
2. causation related to intervention: If researchers want quantitative associations among variables to illuminate causes (in some sense of the term), these associations must be construed in relation to some class of changes or interventions that could, in principle, be made (Pearl 2000; Woodward 2003; Taylor 2005, 241ff, 2006a);
3. possible heterogeneity of factors: There is room to question the methodological assumption that, when similar responses of different individual types are observed, similar conjunctions of genetic and environmental factors (or, in epidemiology, risk and protective factors) have been involved in producing those responses;³

² A genetic or “variety” effect in an AOV can be thought of as a single value of the trait for each variety that best conveys its difference from other varieties. (Similarly for location effects.) The variety effect can be calculated as the average value of the trait for the variety over the range of locations in which varieties are observed minus the average of all varieties over all locations.

³ This sense of “heterogeneity” needs to be distinguished from other uses of the term in the arena of genetics (see Kaplan 2000, 18) and statistics: statistical methods often assume equality (or homogeneity) of variances from one sample to the next; mutations in a gene may be heterogeneous in the sense that they occur at a variety of points in the gene and the clinical expression of such mutations can vary significantly; and different genetic conditions may be expressed as the same clinical entity. This last “genetic” form of heterogeneity is a restricted case of the heterogeneity I am examining in that the environmental factors acting in conjunction with the genetic factors are not considered.

4. conditionality: Because the calculation of effects in an AOV depends on averages of observations for the trait over a set of genetically replicable varieties and over a set of replicable locations or environments, effects (and thus any hypotheses or causal implications drawn from them) are conditional on those particular sets. (Similarly for coefficients calculated through “path analyses” based on additive models related to those in AOV; Lynch and Walsh 1998); and
5. experience of membership in different groups analyzed as different locations: This is entailed when analyzing data that consists of replications in observational studies that are within, not across, groups (as is always the case with human racial groups); it is not a matter of assuming disjunct kinds of causes.

These core issues, individually or in combination, give rise to a series of corollaries. Before laying them out, let me remark on my use of the agricultural terms “variety” and “location.” Unlike the more common terms “genotype” and “environment,” they do not suggest what needs, in fact, to be established, namely, that the quantities estimated through an AOV have a relationship with measurable genetic and environmental factors. Similar thinking leads me to refer to “trait” not “phenotype.” Moreover, in explicating my conceptual critique, I find it helpful to consider first the analysis of agricultural crop trials in which a number of different genetically replicable plant varieties or "cultivars" are grown in multiple plots in one or more "locations." In these analyses, the best case for illuminating genetic and environmental factors can be achieved; this provides a contrast for analyses in human genetics, which falls far short of the ideal.

Corollaries

6. Effects provide limited guidance in hypothesizing about measurable factors. In practice, effects help only when cluster analysis can be used in agricultural crop trials to group varieties by similarity in responses across all locations (Byth et al. 1976). (Such clustering minimizes heterogeneity [#3] and allows researchers to hypothesize about the group averages—about what factors in the locations elicited basically the same response from varieties in a particular variety group that distinguishes them from other groups.⁴

⁴ From Taylor (2006a): “To invent an example, suppose the cultivars in a group yielded poorly in locations where rainfall occurred in concentrated periods on poorly drained soils and the cultivars originated from particular parental stock more susceptible to plant rusts (a form of parasitic fungi). The obvious location-specific hypothesis about genetic factors

Of course, knowledge from sources other than the data analysis is always needed to help researchers generate any such hypotheses about genetic and environmental factors. See Taylor 2006a for further discussion of heterogeneity, grouping, and generation of hypotheses.)

In human quantitative genetic research, however, genetic varieties can at most be replicated in two locations (i.e., identical twins separated at birth) and these locations differ from one variety (twin pair) to the next. This means that

7. grouping of varieties by similarity of responses across locations is impossible (see #6) and thus effects provide no guidance in hypothesizing about measurable factors behind observations of human traits. By extension, heritability, which can be derived from effects, also provides no guidance in hypothesizing about measurable factors behind observations of human traits. (This also applies to heritability calculated from path analyses; see Taylor 2006a, online appendix 1, part 5.)

This is a bold conclusion. If my critique is not dismissed out of hand, I can imagine it being challenged around, for example, my focus on measurable factors—“isn’t it enough to know that genetic effects outweigh environmental in an AOV (or path analysis)?”—or around my insistence on the possibility of heterogeneity when it could be more parsimonious to assume that the same genes are behind the traits that appear similar, or around how I would interpret the variation in heritability across economic classes (Turkheimer et al. 2003). Rather than anticipate and address counter-arguments in this short paper, I want to move onto other issues, but let me note here that:

8. support is lacking, even in the ideal case of crop trials, for the contention that high heritability is an indication that measurable genetic factors have more influence on variation in the trait than measurable environmental factors (Taylor 2006a, sect. 4.2). Given this, the origins and durability of the heritability concept might seem hard to understand, until we recognize that

modulated by environmental factors is that the cultivars share genes from the parental stock that are related to rust susceptibility and this susceptibility is evident in the yields when the rainfall pattern in a location enhances rusts. After additional research comparing the cultivar and parental genomes, it may be possible to identify specific sets of genes that are shared and investigate whether and how each one contributes to rust susceptibility. (See Byth et al. 1976, 224ff for actual hypotheses after analysis of the international wheat cultivar trial referred to earlier. In their analysis the locations were also grouped, using similarity in the responses elicited across the full range of cultivars.)”

9. heritability can be a useful predictor of advances through selective breeding in agricultural and laboratory settings where researchers have the ability to replicate varieties and locations (give or take some variability of weather from season to season in the field) and select among varieties for the next generation on the assumption that the environmental factors will remain unchanged. Indeed, such settings correspond to a causal construal (see #2 & 4) in which
10. differences in effects can be construed as causes if the unknown dynamics remain close to the original situation (i.e., the only difference in any such “re-run” is “noise” equal to the residual or “error” effects in the AOV model).

Further steps in analyzing data

There is a long way from analysis of observed traits to interventions based on well-founded, non-conditional claims about genetic or environmental causality. Suppose that hypotheses have been derived about measurable factors (even if, as must be the case for human traits, AOV has provided no guidance [#7]). The next step for researchers would be to investigate associations with measurable factors through regression analysis and experimental trials. In both cases, conditionality (#4) applies, now extended to conditionality on the set of factors measured as well. Indeed, measurable factors that are significant in a regression analysis can be construed as causes only in re-run situations (#10). By choosing such factors to be manipulated in experimental trials, researchers are assuming that this does not modify the structure of the dynamics as a whole (something easier to achieve in crop trials than in human social relations). Insights from these studies can, in turn, contribute to research on the ways that pathways of growth and development are affected by the genetic makeup of varieties and the environmental factors in the locations—presuming such research has been taking place. This research might, in turn, provide a basis for interventions outside the typically well-controlled conditions in which research on causes in growth and development is undertaken. The sequence of steps in this paragraph is summarized in Table 1.

Table 1: Connections from one kind of data analysis to the next

	Crop trials (varieties each replicated over a number of locations)	Human studies of twins and other relatives
Observations of a trait that differs across different varieties and locations	AOV + Cluster analysis + knowledge from sources outside data ->	AOV (& path analysis) not helpful in generating hypotheses about measurable factors
l v	hypotheses about measurable factors	(hypotheses about factors drawn from other sources)
Observed associations with measurable factors	Significant factors from regression analysis ->	Significant factors from regression analysis ->
l v	factors for testing through experimental trials	Same as on the left (but more questionable if factors can be manipulated w/out modifying structure of dynamics)
Experiments that vary measurable factors	Significant factors->	(Rare)
l v	insights for investigation of dynamics of development	?
Factors over sequence of development [rarely-realized ideal]	Significant factors in development in controlled research conditions ->	?
l v	candidates for interventions in less controlled situations	?

Differences between groups

Discussions about heritability in humans get most heated around the issue of explaining differences among groups. Given that, in my account, effects from AOV and heritability estimates provide no guidance in hypothesizing about measurable factors behind observations of human traits within one group of varieties (#7), they can provide no guidance about measurable factors associated with differences between two groups. Nothing more need be said. However, if we examine further what is involved in attempting to find genetic factors that explain differences between groups, some deeper issues can be exposed.

Consider first the case of crop trials in which the observations of the trait are used to cluster varieties by similar responses across locations (#6). By minimizing the possibility of heterogeneity (#3), researchers can hypothesize about the group averages—about what factors in the locations

elicited basically the same response from varieties in a particular variety group that distinguishes them from other groups (#6). Figure 1 conveys the relationship between factors and patterns in data that underlies such hypothesizing. Notice that it involves both genetic and environmental factors and that insights about one group in one location are related through contrasts to insights about other groups in that location and about the same group in another location.

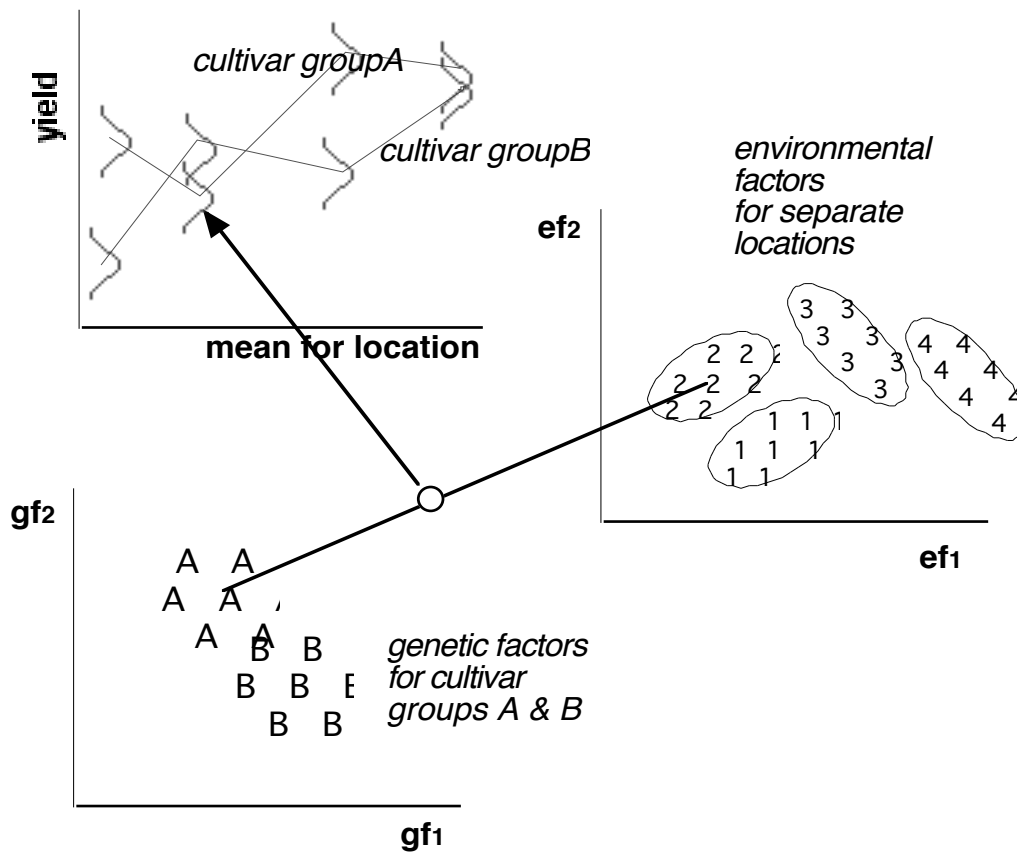


Figure 1. Patterns in data in relation to homogeneous genetic and environmental factors.

However, if varieties are not grouped by similarity of responses across locations, the possibility of heterogeneity of factors (#3) remains. The relationship between factors and patterns in data that underlies any hypothesizing in this situation is very difficult to disentangle (figure 2). To undertake such hypothesizing is akin to hypothesizing about the difference between group averages as if the spread (variance) of within-cultivar-group effects were noise (figure 3). Such a typological worldview, whether held deliberately or inadvertently, warrants interpretation (see sect. 3.1 & 3.3).

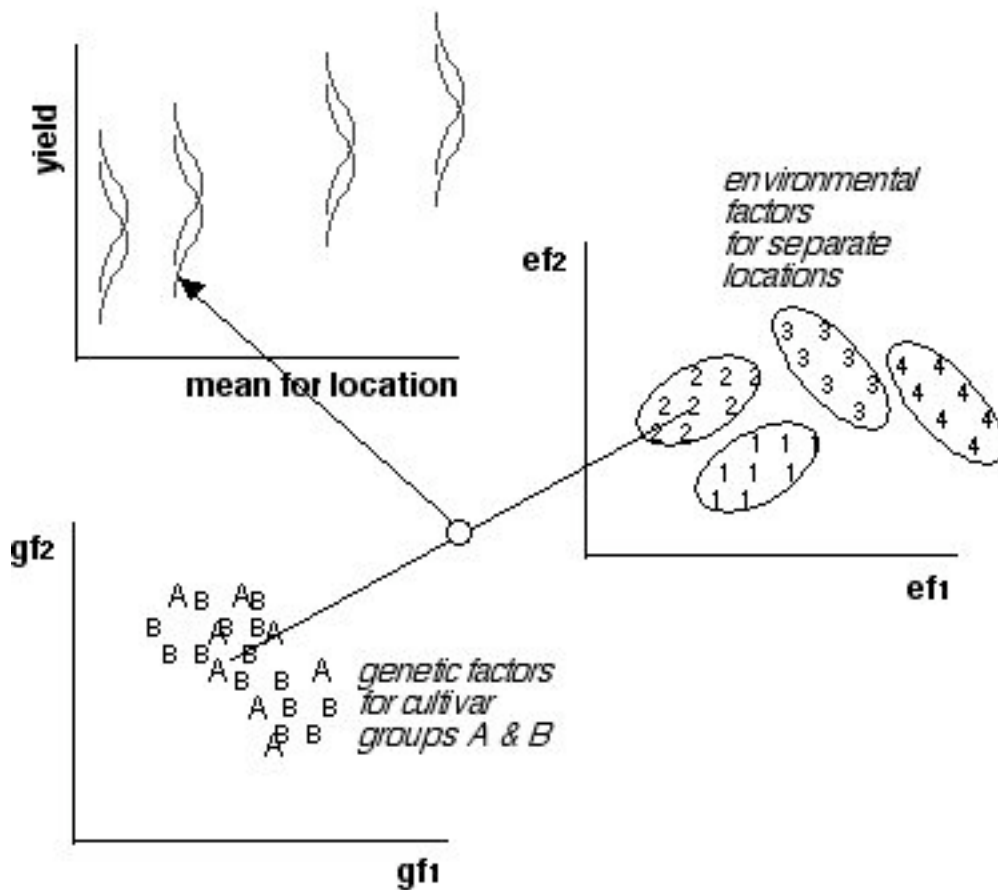


Figure 2. Patterns in data in relation to heterogeneous genetic factors. Cultivar groups A and B have not been formed by cluster analysis and are different groups from those in Figure 1.

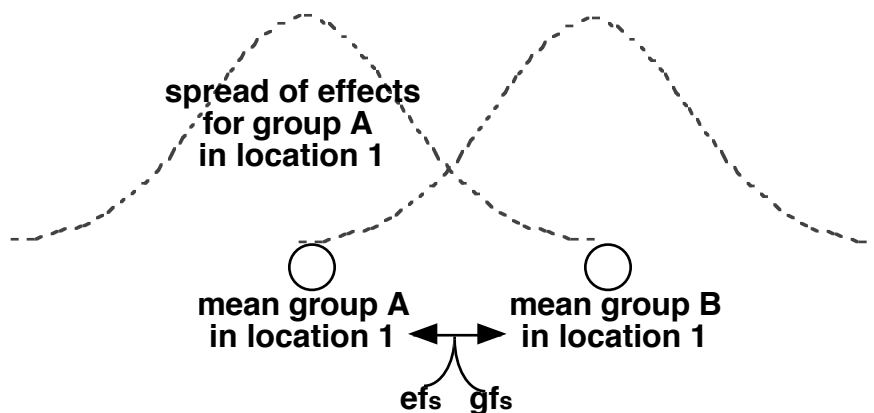


Figure 3. One approach to generating hypotheses about measurable factors. gfs and efs refer to measurable genetic and environmental factors associated with the observed difference between means (averages). The spread of effects is not taken into account.

If the possibility of heterogeneity has not been minimized, then, by extension, it must also be difficult to gain insights from one group that shed light on factors in other groups or on factors for the first group in other locations. The prospects become even worse when replications are limited to within one location only, as is the case for human racial groups (see #5). In that case two bell curves from two different pairs in figure 2 would have to serve as the basis for any hypothesizing about the genetic and environmental factors. Neither the contrasts within a location nor the contrasts for a single group across locations would be available to guide (or constrain) the researchers.

Lindman's (1992) textbook illustrates a cautionary note about "nested" AOV (i.e., when each variety is replicated in one location only) with an example of high school students' test scores in algebra viewed in relation to their teacher and school. The students within a school were randomly assigned to a teacher in their usual school. Lindman notes that a significant location (school) effect "is likely to be interpreted as due to differences in physical facilities, administration, and other factors that are independent of the teaching abilities of the teachers themselves... [However, d]ifferences between teachers in different schools are part of the [location or school] effect, and the observed differences between schools could be due entirely to the fact that some schools have better teachers [or] some schools have smarter children attending them" (Lindman 1992, 194).

Lindman could have added that the observed differences between schools could be due entirely to combinations of factors, such as students responding worse to teachers whose attention is distracted because their school's administrators insist more on detailed documentation of student performance and so on. The point is that

11. nested analysis cannot help researchers hypothesize about the difference between the average scores of varieties replicated only within locations (i.e., teachers [as replicated in student test scores] in the two schools). Researchers can do no better than to conduct a separate AOV for each combination of a set of varieties and location—or, in the context of racial differences, for each combination of group of individuals and experience of membership in different racial groups.

Although Lindman's note and the preceding discussion and diagrams center on AOV, possible heterogeneity of factors (#3), the limitations of nested analysis (#11), and membership in different groups analyzed as different locations (#5), it might apply equally well to drawing hypotheses and insights from regression analysis and experimental trials (see Table 1). If so, the idea

that genetic factors might be able to explain differences between groups lacks a method of data analysis through which it could be assessed well. Accounting for the persistent appeal of non-assessable formulations in science and popular discussion would then warrant STS and other social scientific interpretation. These and other questions motivate the section to follow.

3. Some Possible Angles for STS Investigations

In this section I identify a number of angles through which STS scholarship might bring attention in research and policy to the implications of heterogeneity.

a. Conceptual reconstruction and extensions

The possible heterogeneity of factors is not mentioned as an issue in the extensive entry on heredity and heritability in the Stanford Online Encyclopedia of Philosophy (Downes 2004) or the key sources cited therein (e.g., Sarkar 1998; Kaplan 2000). Clearly, there is room to draw philosophers into debate about the conceptual oversights or missteps I have identified in sect. 2 (as well as in Taylor 20006a, b, c, d) and to refine, rethink, or extend my arguments and their conceptual basis. I have already encountered skepticism about the relevance of agricultural methods to the analysis of human variation. My response has been to note that human heritability estimation is based on data that are less ideal than agricultural crop trials so it cannot somehow, miraculously, allow researchers to support claims about more general notions of genetic or environmental causality.

The agricultural case, moreover, seems to be well suited for clarifying discussion of the kinds of realizable intervention built into inferences about causality. Given the emphasis in recent philosophy and in social science to the causation-intervention relationship (Pearl 2000, Woodward 2003), there should be interest in, if not immediate acceptance of, my argument that heritability estimation and the AOV on which it is based presuppose a circumscribed sense of causality in which everything is kept close to the original situation. Indeed, we could investigate whether the limitations I see in causal inference from quantitative data in heritability studies have wider relevance in social science and epidemiology. We could consider, in particular, whether the “close to the original situation” condition applies to any attempt to move from patterns based on observations through inferences about causes to policy interventions. How can that condition be reconciled with the

likelihood that many policy interventions, if implemented, alter the structure of the relations that produced the observations, patterns, and, thus, the causal inferences?

Another conceptual issue is to explore the relationship between inattention to the possibility of heterogeneity and a typological worldview. Notice that, even as Lindman performs the valuable role of cautioning readers about nested analyses (end of sect. 2 above), he perpetuates the typological worldview in referring to “the observed differences between schools” when he means the observed differences between averages for schools. It is commonplace to use expressions of the kind “men are taller than women,” “men tend to be taller than women,” or “men are, on average, taller than women.” Is more than linguistic convenience involved in people not using the more precise statement: “the average of men’s heights is greater than the average of women’s heights”? Indeed, we could describe the pattern in the data even more precisely as “the variation among men’s heights centers at a point greater than the variation among women’s heights.” Would this, in turn, help us keep in mind that, in principle, the factors producing that pattern could vary among men and women and need not include factors solely possessed by one sex or the other? In other words, that there need not be something essentially of each group that leads to differences in their averages (see also c below).

Finally, we could also investigate a more modest question, one of the sociology of knowledge, concerning the ways that discussion among philosophers of science might have obscured the relevance of heterogeneity, say, through visual and verbal conventions that emphasize types over variation, or by over-reliance on scientists to set the terms of issues on which philosophers focus their efforts in conceptual reconstruction.

b. History of translation from agriculture and laboratory breeding to human genetic analysis.

Heritability estimation was first used in selective breeding in agricultural and laboratory settings where researchers have the ability to replicate varieties and locations. Indeed, when agricultural researchers compare varieties and make recommendations to farmers and when they select among varieties for the next round of crop trials, they do so on the assumption that the environmental factors will remain more or less unchanged. For observation of human traits, however, such replicability of varieties and environmental factors is not possible. This observation opens up the historical question of how such restrictive conditions were discounted or forgotten in the translation of heritability estimation from selective breeding to human genetics. For example, when

Wright (1920) presented his original formulation of heritability estimation he used the notation E to refer to “environmental factors that are common to litter mates” of guinea pigs that he bred. To translate heritability estimates into predictions of future changes under selective breeding, these “factors”⁵ had to remain constant from one generation to the next. “E” now, however, is used to denote environmental factors without reference to Wright’s restricted conditions. One part of an historical investigation would be to trace Wright’s notation from its origin through its adoption in human genetics, where it has become commonplace in discussions of the relative influence of G (genes) and E (environment) in accounting for the variation among individuals and groups.

c. Racialized imaginaries in the analysis of differences among groups.

Another historical question concerns the persistent interest in explaining differences among the averages for groups defined on racial grounds. Because the ranges within those groups are large and overlapping, any finding about the differences between averages for groups is hard to do anything with unless individuals are treated on the basis of their group membership. What else can people do with the patterns researchers find in observations of human relatives assigned to groups? (Recall that the possible heterogeneity of factors makes heritability estimates within groups irrelevant for developing or supporting hypotheses about differences between averages for groups; sect. 2; see also Taylor 2006b.) When researchers do not address heterogeneity, are they making typological or essentialist assumptions? Does a racially essentialist imagination facilitate the transfer of conventional statistical tools from agricultural to human research? Might it be possible to pinpoint paths not taken or objections not taken up in scientific and public debates about group average differences, then interpret these blind spots in terms of the persistence of racial types (as against overlapping variation) as an organizing category in American social and scientific thought? Similarly, might the transfer of tools from selective breeding to analysis of human variation (see area b above) speak to persistence of eugenic hopes and fears? Just who is empowered to do something as a result of analysis of group differences—and who is given license not to have to do anything? (See area a on causation as intervention.) Such questions are more speculative and invite a more interpretive, cultural approach than the concept-centered philosophy and history of areas a & b above.

⁵ Strictly, the effects from the AOV or path coefficients—no measurable factors were involved in the analysis.

d. Engagement of STS scholars with scientists.

It is routine in STS to portray scientists shaping society as they establish knowledge or make technologies work, but STS scholars are often not reflexive about what we aim to do with our own knowledge claims, i.e., our interpretations of science (Taylor 2005). This issue becomes especially apt when our interpretations point to shortcomings in the science. Are we envisaging that our critical social/historical interpretations will influence working scientists or that it is more effective to insert STS perspectives into the education of future scientists? When should we “go native” among the scientists in their “labs” or become active citizens or consultants in policy debates? The answers must depend on the particular situation of each STS scholar and the colleagues they influence.

I envisage both indirect and direct influence. On the indirect side, I plan to subject the critique outlined in sect. 2 to the scrutiny of appropriate audiences and workshops and try to interest others in delving deeper into some of the areas of STS investigation outline above. This dialogue session (formally at the Society for Social Studies of Science but drawing participants from the concurrent meetings of the History of Science Society and the Philosophy of Science Association) is a start.⁶ I will consider my engagement with STS scholars to be a success if collaborations and associations arise through which we support each other in such emergent (and thus, at first, risky) research. The more direct engagement with scientists is the subject of the section to follow.

4. Scientific Fields Opening Up Room to Pay Attention to Heterogeneity

During my interviews of scientists and STS scholars who have critiqued the work of heritability researchers a theme recurred, namely, that, even when the researchers acknowledge the criticisms, their subsequent research shows no change. This is not an argument against the developing critiques and STS interpretations (such as those in sects. 2 & 3), but it does motivate me to use my interests and skills in quantitative analyses in the life sciences more directly to draw attention to—and maybe even stimulate—the development of alternative research programs. To pursue this positive side of the critique outlined in section 2, I have identified four areas of inquiry that have potential to contribute to research that does not obscure the possible heterogeneity of factors

⁶ A longer, interactive workshop is being planned for the 2007 meetings of the International Society for History, Philosophy and Social Studies of Biology, the hosts of which are the ESRC Centre for Genomics in Society, at the University of Exeter, England.

(described below). I have begun to visit or correspond with key researchers, who have shown their receptiveness to exchanges with outsiders to their disciplines. My goal is to explore and become able to articulate to others the possibilities and limits of generating empirically validated models of developmental pathways whose components are heterogeneous and differ among individuals at any one time and over generations.

a. Life-course analyses in education.

Woodhead (1988) summarizes studies explaining how the IQ test score increases produced by Head Start preschool programs tend to be transient, but in the long term, through social support systems initiated or enhanced during the Head Start years, the children end up with significantly higher high school graduation rates, employment, and many other socially valued measures. Ou (2005) has put that conclusion on a quantitative basis in finding associations among preschool participation and other measures taken through the course of schooling and development to adulthood.⁷

b. Multivariate developmental models of mental illness.

Ou (2005, 604) remarked on her model's limitations in the areas of "the correlational nature of the data, possible alternative models, and generalizability." It might also be noted that the factors in Ou's analysis would traditionally be labeled environmental. In a different context, factors that could be labeled genetic were incorporated by Kendler et al. (2002) in a comprehensive developmental model that accounted for 52% of the variance in liability to episodes of major depression. Although the models of Kendler, like those of Ou, provide a picture of development that is rich and plausible, clarification is warranted of the class of changes or interventions in which it makes sense to construe the factors in the models as causes (see sect. 2). Indeed, Kendler et al. (2002, 1133) show admirable reserve in concluding that their "results, while plausible, should be treated with caution because of problems with causal inference, retrospective recall bias, and the limitations of a purely additive statistical model." Interestingly, they did not remark on the absence of variables that correspond to therapeutic interventions (as if to suggest that these had no effect on the etiology of depression or its preceding risk factors) or to social changes that have led to the rising incidence of

⁷ E.g., basic skills scores at kindergarten, classroom adjustment (age 9-10), parent involvement (age 8-12), abuse/neglect reports (age 4-12), school quality (age 10-14), number of school moves (age 10-14), commitment to school (age 10 or 15), grade retention (by age 15), achievement (age 15), highest grade completed (by age 22).

depression. Such omissions would seem important to rectify in any analogous modeling of IQ test scores and other outcomes subject to educational influence.

Kendler et al. (2002) take an additional step in characterizing different paths to the outcome to be explained, namely, depression, e.g., "Paths Reflecting a Broad Adversity/Interpersonal Difficulty Pathway to Major Depression." Although the paths are identified by eye, the exercise opens up the possibilities of identifying paths that operate heterogeneously across social groups, across individuals within any social grouping, and, in relation to the Flynn effect (large increases in average IQ scores over time), heterogeneously across generations.

c. Life course analyses in epidemiology.

In a field initiated by the epidemiologist Barker at the University of Southampton, a large number of researchers are now studying associations between nutritional deficits during critical periods in utero and diseases of late life, including heart disease and diabetes. The integration of fetal origins and subsequent influences is now taking place under the label of "life course epidemiology" (Kuh and Ben-Shlomo 2004).

d. Statistical innovations in life course epidemiology and its applications.

Gilthorpe and his students and colleagues have highlighted the statistical challenges in interpreting associations between early life influences and diseases of later life (Head et al. 2005). West and Gilthorpe are developing alternative statistical analyses that enable them to characterize different pathways of growth over the lifecourse (which, in my terms, makes it easier to visualize the possible heterogeneity of factors underlying responses).

In the spirit of being reflexive about what we aim to do with our own knowledge claims (sect. 3, area d), I will consider my engagement with these fields to be a success if I can: a) stimulate the scientists to be more explicit about the ways their methods and models address the possible heterogeneity of factors and to pay attention to each other's research; and b) identify the place and direction of most leverage for my personal combination of interests and skills in STS and in quantitative analyses in the life sciences. In the long term a collaborative proposal might emerge with some of these scientists and/or other STS scholars that furthers the transformative potential that attention to heterogeneity has, I believe, for quantitative analysis in the study of heredity and social sciences more generally, for policy-making based on such research, and for popular discussion that

no longer resorts to ideas that cannot be assessed concerning the relative contribution of genes and environment to human traits and differences among groups.

References

- Byth, D. E., R. L. Eisemann, I. DeLacy. (1976). "Two-way pattern analysis of a large data set to evaluate genotypic adaptation." Heredity 37(2): 215-230.
- Downes, S. (2004) Heredity and Heritability in the Stanford Encyclopedia of Philosophy (<http://plato.stanford.edu/entries/heredity/>, viewed 11 May 2006)
- Flynn, J. R. (1994). IQ gains over time. Encyclopedia of Human Intelligence. In R. J. Sternberg. New York, Macmillan: 617-623.
- Head, R., G. Ellison, M. Gilthorpe Eds. (2005). Statistical challenges facing the foetal origins of adult disease hypothesis, 3rd International Congress on the Developmental Origins of Health and Disease, Toronto.
- Jencks, C., and M. Phillips, Eds (1998) The Black-White Test Score Gap. Washington, DC: Brookings Institution Press.
- Jensen, A. (1969) How much can we boost IQ and scholastic achievement? Harvard Educational Review 39: 1-123.
- (1970) Race and the genetics of intelligence: A reply to Lewontin. Bulletin of the Atomic Scientists 26: 17-23.
- Kaplan, J. (2000). The Limits and Lies of Human Genetic Research. New York: Routledge.
- Kendler, K., C. Gardner, C. Prescott. (2002). "Towards a comprehensive developmental model for major depression in women." American Journal of Psychiatry 159: 1133-1145.
- Kuh, D. and Y. Ben-Shlomo, Eds. (2004). A Life Course Approach to Chronic Disease Epidemiology. Oxford: Oxford University Press.
- Lewontin, R. (1970a) Race and intelligence. Bulletin of the Atomic Scientists 26: 2—8.
- (1970b) Further remarks on race and the genetics of intelligence. Bulletin of the Atomic Scientists, 26: 23-25.
- (1974) The analysis of variance and the analysis of causes. American Journal of Human Genetics 26: 400-411.
- Lindman, H. R. (1992). Analysis of Variance in Experimental Design. New York, Springer-Verlag.

- Lynch, M. and B. Walsh (1998). Genetics and Analysis of Quantitative Traits. Sunderland, MA, Sinauer.
- Ou, S.-R. (2005). "Pathways of long-term effects of an early intervention program on educational attainment: Findings from the Chicago longitudinal study." Applied Developmental Psychology 26: 478-611.
- Parens, E. (2004) Genetic differences and human identities: On why talking about behavioral genetics is important and difficult. Hastings Center Report (January-February): S1-S36.
- Pearl, J. (2000) Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press.
- Rutter, M. (2002). "Nature, nurture, and development: From evangelism through science toward policy and practice." Child Development 73(1): 1-21.
- Sarkar, S. (1998). Genetics and Reductionism. Cambridge, Cambridge University Press.
- Taylor P. (2005) Unruly Complexity: Ecology, Interpretation, Engagement. Chicago: University of Chicago Press.
- (2006a) "Heritability and heterogeneity: I. The limited relevance of heritability in investigation of genetic and environmental factors," to appear in Biological Theory, 1(2): 150-164.
- (2006b) "Heritability and heterogeneity: II. The irrelevance of heritability in explaining differences between means for different human groups or generations," to appear in Biological Theory, 1(4).
- (2006c) "The analysis of variance is an analysis of causes (of a very circumscribed kind)," International Journal of Epidemiology, 35: 527-531.
- (2006d) "Exchange around the knowledge claim, 'Intelligence is 75% genetic'," <http://sicw.wikispaces.com/Intelligence75percentGenetic> (viewed 21 August '06)
- Turkheimer, E., A. Haley, M. Waldron, B. D'Onofrio, I. Gottesman (2003). "Socioeconomic Status Modifies Heritability of IQ in Young Children." Psychological Science 16(6): 623-628.
- Woodhead, M. (1988). "When psychology informs public policy." American Psychologist 43(6): 443-454.
- Woodward, J. (2003). Making Things Happen: A Theory of Causal Explanation. New York: Oxford University Press.
- Wright, S. (1920). "The relative importance of heredity and environment in determining the piebald pattern of guinea pigs." Proceedings of the National Academy of Science 6: 320-332.