Supplementary Electronic Materials for
Heritability and Heterogeneity
Peter Taylor
Biological Theory (2) 2006, 150-164.

Online Appendix 3: "The Analysis of Variance and the Analysis of Causes" Revisited

In order to amplify the themes presented in the paper, this appendix revisits some oft-cited examples from agriculture and population genetics that Lewontin (1974, 1982) uses to critique the use of heritability and AOV in human behavioral genetics. Unfortunately, because Lewontin's examples do not keep effects and measurable factors clearly distinct (*contra* theme 2), they tend to obscure the difficulties of moving from the observation of traits that differ across cultivars (genotypes) and locations (environments) to hypotheses and subsequent investigation of measurable genetic and environmental factors that correspond to those differences.

Lewontin (1982: 132-133) introduces two agricultural thought experiments to help non-specialist readers visualize why heritability within groups is not relevant to explanation of differences between groups. In one example, two set of seeds sampled from one open-pollinated cultivar are planted in two pots of washed sand. Both pots are fed with plant-growth solution, but the solution in the second pot lacks nitrogen. Lewontin observes that, because each location (pot) is uniform, variation within them will be associated with genetic differences among the sampled seeds. Lewontin calls this a heritability of 1, but the correct value of within-location heritability is 0 because a new sample of seeds grown in the same location would have no correlation with the first sample. Although all the within-location variance is associated with genetic variability, this variability is residual variance in the AOV based on model 4. (If we thought of the sample of seeds as a set of *different* cultivars and if we could clone each seed and replant it in each pot, then model 5 would be the appropriate one. In the absence of replications of these cultivars in each pot, residual variance is zero and the heritability within locations would then be 1.)

Similarly, the heritability across both locations is 0. This value seems consistent with Lewontin's observation that the difference between (the mean measurements for) locations is "totally environmental" (i.e., entirely associated with the environmental difference of nitrogen vs. no nitrogen). However, for the rerun predictability when the location (not the cultivar) is constrained to be the same in both cases, the correlation between the current data and the predicted results is less than 1. The only way to increase this correlation — that is, for the difference between the mean measurements for each location to be more strongly associated with the location-to-location (pot-to-pot) difference—is to reduce the within-location or residual variance that is associated with genetic differences among seeds in the open-pollinated cultivar. (See Table 1 for a numerical analysis of Lewontin's example.)

Table 1.
Reading from the right-hand diagram in Lewontin (1982, 132) and inventing arbitrary units for height of the plants produces Data set 3, which can be analyzed using a variant of equation 5':

$$y_{jk} = m + l_j + r_{jk} \quad (5a')$$

See text for further discussion.

			Estimates of effects		**Variance & heritability estimates**	
	location		m	3.15	σ^2_l	1.95 (75%)
cultivar	1	2	l_1	1.35	σ^2_ε	0.65 (25%)
1	5, 4, 6, 3	2, 1.8, 2, 1.4	l_2	-1.35	$h^2_{\text{within location or across both locations}}$	0
			r_{jk}	varied	rerun predictability for locations	.75

Precision about these technical issues helps to disconnect heritability, which depends on AOV of observational trials and thus difference-in-effects causes, from the idea of measurable genetic factors that differ among individuals in association with differences among individuals in the trait in question (theme 6). Moreover, when Lewontin uses his example to make conceptual points, he does not comment on the control the example presumes over which varieties to select and grow and the ability to replicate environmental factors. Such control characterizes observational and experimental crop trials, but is not available in human behavioral genetics. The example also blurs the distinction between observational and experimental crop trials as if human behavioral genetics could use experiments to generate knowledge of measurable factors.

In Lewontin's other example (1982, 132), one seed from each of two inbred cultivars is planted in a series of pots of soils taken from different locations (Figure 1). Lewontin observes that for each of the inbred cultivars there are no genetic differences across the locations (pots) and all the location-to-location (pot-to-pot) variation must be "environmental," corresponding to differences in the soils. At the same time, noting that cultivar 1 does better than (or as well as) cultivar 2 in each location, Lewontin asserts that this gap is entirely "genetic" because the cultivars experienced identical sets of locations.

The meanings of the terms "genetic" and "environmental" are ambiguous in this context. Again, it is instructive to be precise. First, notice that the advantage of cultivar 1 over cultivar 2 varies from one location to the next. In terms of AOV and difference-in-effect causes, there is a cultivar effect (i.e, a non-zero gap between the means across all locations of cultivar 1 and 2), and there are cultivar-by-location-interaction effects. In terms of measurable genetic and environmental factors, no hypotheses are obvious. What we can infer, however, is that the different within-location, between-cultivar differences are associated *either* with a different mix of genetic factors in different locations, or with the same genetic factors having a different influence. The within-cultivar differences across locations may correspond with one or many environmental factors and these factors need not be the same from one cultivar to the next. In short, the best we can say about the gap between cultivars in each location and the gap between locations for each cultivar is that they are associated with a combination of environmental and genetic factors.

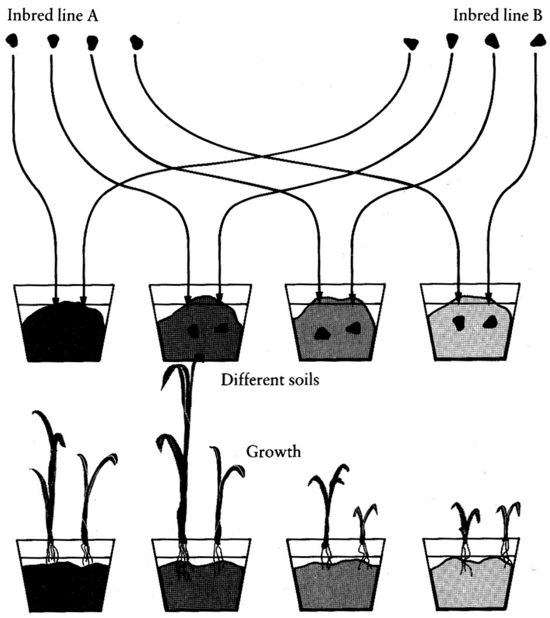

Figure 1.
Variation in two inbred cultivars planted in a series of pots of soils taken from different locations (from Lewontin 1982, 132).

In terms of rerun predictability, the absence of replication means that heritability for each cultivar is not a very interesting quantity in this case. Within one location, this heritability must be 1; across locations it must be 0 — even if the cultivars are not inbred. On the other hand, taking both cultivars into account, the heritability across locations is greater than zero and the rerun predictability when the location is constrained to be the same in both cases is less than 1 — not, respectively, 0 and 1 as might be naively expected if these heritability and rerun predictability measures corresponded to the terms "genetic" and "environmental" as used by Lewontin and others following him. In summary, this example, like the preceding one, blurs distinctions in ways that lend plausibility to lines of thinking in which measurable genetic and environmental factors are separable and in which insight about those factors could follow from learning that heritability estimates (and difference-in-effect causes) are high or low. (See Table 2 for a numerical analysis of this second example.)

In an earlier, much cited essay. "The analysis of variance and the analysis of causes," Lewontin (1974) argues, in effect, that, because any AOV is conditional (theme 1), it cannot shed light on causes beyond the local combination of genetic types and locations observed. He supports this argument with diagrams of "norms of reaction" that summarize the response of a cultivar or genetic type when some environmental factor is varied. Norms of reaction for different cultivars that vary in slope and position can confound any attempt to extrapolate the relative ranking of genetic types (or cultivars) observed over part of the range of the environmental factor to the full range. Most of Lewontin's diagrams are schematic, with the measured trait plotted against an unspecific environmental factor E, but one diagram shows real data on viability of strains of fruit flies plotted against temperature. In either cultivar having a single continuous environmental factor as the horizontal axis in both the schematic and real cases reinforces the idea that location effects in an AOV can be readily translated into environmental factors with continuous gradients (*contra* theme 2). Like agricultural breeders, but even more so, population geneticists who study fruit flies have control over genetic types and environmental conditions and can readily envisage generating such plots. However, using diagrams of norms of reaction to make conceptual points about the AOV in relation to quantitative or behavioral genetics steers us away from visualizing the difficulties in using AOV to expose equivalent environmental factors in humans.

Table 2.

Reading from figure 1 (i.e, the left-hand diagram in Lewontin 1982, 132) and inventing arbitrary units for the height of the plants produces Data set 4, which can be analyzed using a different variant of equation 5':

$$y_{ij} = m + c_i + l_j + cl_{ij} \qquad (5b')$$

See text for further discussion.

					Estimates of effects		**Variance & heritability estimates**	
	location				m	2.75	σ^2_l	2.56 (69%)
cultivar	1	2	3	4	c_1	.75	σ^2_c	0.56 (15%)
1	4	7	2	1	c_2	-.75	σ^2_{cl}	0.56 (15%)
2	3	3	1	1	l_j	.75, 2.25, -1.25, -1.75		
					cl_{1j}	-.25, 1.25, -.25, -.75		
					cl_{2j}	$-cl_{1j}$	$h^2_{\text{across locations}}$	.15

References

Lewontin RC (1974) The analysis of variance and the analysis of causes. American Journal of Human Genetics 26: 400—411.

Lewontin RC (1982) Human Diversity. New York: Freeman Press.