Biological Sciences; Genetics

A note on Zuk, Hechter, Sunyaev, and Lander (2012)

[a] Department of Psychology
Simon Fraser University, 8888 University Drive, Burnaby, B.C. V5A 1S6, Canada. Email: michael_maraun@sfu.ca.
[b] Department of Psychology, Ludwig Maximilian University, Leopoldstr. 13
D-80802 Munich, Germany
Email: heene@psy.lmu.de

Michael Maraun[a]

*Moritz Heene[b]

*Corresponding author: Prof. Dr. Moritz Heene
Ludwig Maximilian University, Munich, Germany
Department of Psychology
Leopoldstr. 13
D-80802 Munich, Germany
heene@psy.lmu.de

Number of text pages: 3; number of figures: 0; number of tables: 0

Total number of characters: 5.515

Keywords: heritability; quantitative genetics; twin studies.

In our opinion, the recent paper by Zuk, Hechter, Sunyaev, and Lander (1), *The mystery of missing heritability: Genetic interactions create phantom heritability,* is not merely a fascinating analysis of the issue of phantom heritability, but, also, a paradigm example of how methods heralded as yielding estimates of heritability, both narrow and broad, should be investigated. Essentially, these authors break the insidious hold upon thinking of the standard biometric model by carefully disambiguating three fundamental issues: 1) the specification of architecture (both phenotypic and genetic); 2) the definition of heritability (which rests on the specification of variance components, these components defined with respect to the Fisherian decomposition of the genotypic value function; and 3) the mode by which estimates are produced (these classifiable into classical, coefficient based approaches and the more recent structural equation modeling approaches).

These issues have, historically, been conflated under the heading of "heritability" (h^2; (i.e., the proportion of the total phenotypic variance linearly related to gene content); the first and third, additionally, under the facile invocation of the "assumption" of the standard biometric model. In fact, the default invocation of the biometric model reduces the important issue of heritability estimation to a moot consideration of the estimation of parameters misdefined under an irrelevant architecture; thereby precludes the analysis of heritability parameter recovery under possible, realistic, architectures.

Comment [MH1]: I think we should give this definition because many readers and authors conflate heritability with heredity

Our one quibble with the paper is the suggestion it carries that what the reader is seeing in the simulations presented is phantom heritability that has arisen from epistatic interactions. In fact, the *limiting pathway* architecture featured in the paper: (a) involves direct additive impacts of environment on (locus-specific) gene content, thereby undermining the very meanings of each of *genotypic value* and *epistatic interaction*; (b) induces a variety of non-epistatic interactions, notably, (locus-specific) gene content-environment and genotype-environment interactions;.

For an n-loci architecture, the Fisherian decomposition of the genotypic value function, $\Psi'(\mathbf{g})$, which takes as its argument the genotype $\mathbf{g}$, of which, in the biallelic case, there are 3^n particular values, is (see, e.g., 2)

$$\Psi'(\mathbf{g}) = \mu_{\Psi'(\mathbf{g})} + \sum_{j=1}^{n} B(g_j) + \sum_{j=1}^{n} \delta(g_j) + \sum_{l=1}^{m} I_{(\mathbf{g})l} \,. \qquad [1]$$

In expression 1: the variable g_j, j=1...n, represents the *gene content* at locus j, or, in other words, the number of copies of the second allele (e.g., A_2, B_2, C_2, ...) present in a particular locus j-specific genotype; the $B(g_j)$ are locus-specific *breeding values*, and are equal to $\Psi'(\mathbf{g})_{\text{lin}j} - \mu_{\Psi'(\mathbf{g})}$, in which $\Psi'(\mathbf{g})_{\text{lin}j} = \alpha_j + \beta_j g_j$, is the linear predictor of $\Psi'(\mathbf{g})$ on the basis of g_j, and $\mu_{\Psi'(\mathbf{g})}$, the expectation of $\Psi'(\mathbf{g})$; the $\delta(g_j)$ are locus-specific *dominance values*, and are equal to $\Psi'(\mathbf{g}) - \Psi'(\mathbf{g})_{\text{lin}j}$; and the $I_{(\mathbf{g})l}$ are

$$m = \sum_{r=2}^{n} \binom{n}{r} 2^r = \left(3^n - 2n - 1\right) \textit{epistatic interaction terms,} \; \binom{n}{r} 2^r \text{ terms of order } r,\ r = 2..n.$$

The $I_{(g)l}$ quantify the impacts on $\Psi'(g)$ of interactions among the n (locus-specific) gene contents, g_j, $j = 1...n$.

Phenotype Z is a composite function $\Psi(\Psi'(g), \gamma(E))$, the arguments of which are the genotypic value function and a function $\gamma(E)$ that takes as its argument, the entire vector of environments causally relevant to Z. There is present, for Z, in a particular population P, *gene-environment interaction* (properly, *genotype*-environment interaction) if and only if $\Psi(\Psi'(g),E)$ is non-separable; else, under the condition of separability,

$$\Psi(\Psi'(g),E) = \Psi'(g) + \gamma(E), \qquad [2]$$

and genotype and environment do not interact in bringing about Z. In a *strictly additive genetic architecture*, expression 1 contains neither dominance, nor epistatic interaction, terms, the consequence being that

$$\Psi'(g) = \mu_{\Psi'(g)} + \sum_{j=1}^{n} B(g_j) + \gamma(E). \qquad [3]$$

Now, the architecture put forth by Zuk et. al (2012) under the heading of *limiting pathway architecture* (in their notation, $[LP(k,h^2_{pathway},cr)]$, see 3, p. 16ff.) is as follows:

$$Z = \Psi(\Psi'(g),E) = \max(\Psi_1, \Psi_2,..., \Psi_k), \qquad [4]$$

in which $\Psi_j = h^2_{pathway}\alpha + \sqrt{cr(1-h^2_{pathway})}E_c + \sqrt{(1-cr)(1-h^2_{pathway})}E_u$, $j= 1...k$,

$\begin{pmatrix} \alpha \\ E_c \\ E_u \end{pmatrix} \sim N_3(0,I)$, and $0 \le h^2_{pathway}, cr \le 1$.

The max function of expression 4 induces *interactions* among the Ψ_j, the Ψ_j called, in Zuk et. al. (2, p.1193), "inputs." If Ψ_j were interpretable, under $LP(k,h^2_{pathway},cr)$, as the gene content- equivalently, locus-specific genotype- at locus j, i.e., if it were the case that $\Psi_j = g_j$, then these induced interactions would be justifiably interpretable as epistatic, in nature. However, the essential point to recognize here is that, under $LP(k,h^2_{pathway},cr)$, Ψ_j is a function of not only α, which represents gene content, but, also, the variables E_c and E_u, which represent shared and unique environmental effects, respectively.

In consequence: a) the Ψ_j *cannot* be interpreted as standing for gene content (for, by definition, gene content is purely the number of the second allele present in a locus-specific genotype); b) from a), *genotype* is specified not, as it must be, in terms of (locus-specific) gene content, but, also, in terms of environment; c) though the only function under $LP(k,h^2_{pathway},cr)$ that takes as its argument, gene content, hence, that is open to playing the role of genotypic function, is $\max(\Psi_1, \Psi_2,..., \Psi_k)$, $\max(\Psi_1, \Psi_2,..., \Psi_k)$ is prohibited from playing this role (by virtue of its taking as additional arguments, E_c and

E_u); d) the *meaning* of the concept *epistatic interaction* is fundamentally compromised (for interactions among the Ψ_j are not, as they must be in order to qualify as epistatic, interactions among the n (locus-specific) gene contents, g_j, $j = 1...n$).

Furthermore, even if each Ψ_j were specified so as to be a function of only the gene content at locus j, because $LP(k,h^2_{pathway},cr)$ is a non-separable architecture, it would (and does, in fact) induce, in addition to epistatic interactions, genotype-environment interaction. The foregoing does not, of course, invalidate the demonstration of Zuk et. al. apropos the existence of phantom heritability, but does suggest that the phantom heritability yielded under $LP(k,h^2_{pathway},cr)$ arises from a complicated *mixture* of interactions, none of which can be singled out as being epistatic, in nature. An example of an architecture that is similar to [4], but for which both all component quantities are properly defined and epistatic interactions are the *only* type of interaction produced, is

$$Z = \Psi(\Psi'(\mathbf{g}),E) = \max(g_1, g_2,..., g_k) + \gamma(E). \qquad [5]$$

In this architecture, the max function *is* the genotypic function, and each of gene content, genotype, and, consequently, epistatic interaction, is defined correctly.

References

1. Zuk, O., Hechter, E., Sunyaev, S. & Lander, E. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Science of the United States of America, 109(4): 1193-1198. doi:10.1073/pnas.1119675109

2. Lynch, M., & Walsh, B. (1998). Genetics and Analysis of Quantitative Traits. Dunerland, MA: Sinauer.

3. Zuk, O., Hechter, E., Sunyaev, S. & Lander, E. (2012). Appendix of The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Science of the United States of America, 109(4): 1193-1198. Retrieved from www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119675109/-/DCSupplemental.