



The mean matters: going beyond repeatability to interpret behavioural variation

Ned Dochtermann, Raphaël Royauté

► To cite this version:

Ned Dochtermann, Raphaël Royauté. The mean matters: going beyond repeatability to interpret behavioural variation. *Animal Behaviour*, 2019, 153, pp.147-150. 10.1016/j.anbehav.2019.05.012 . hal-03955462

HAL Id: hal-03955462

<https://hal.inrae.fr/hal-03955462>

Submitted on 22 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Commentary

The mean matters: going beyond repeatability to interpret behavioral variation

Research regarding among-individual variation, typically under the label of “animal personality”, often makes use of the estimation of repeatability for inference. This usage makes sense: variance standardized ratios like repeatability allow for straightforward interpretations. For example, repeatabilities greater than 0.5 demonstrate that most of the observed variation in behavior is due to differences among individuals. However, behavioral researchers are interested in a variety of questions vis-à-vis “personality” and, for some of these questions, the inferences that can be drawn from repeatability will be limited (Wilson, 2018). Here we discuss these limitations and additional metrics that may be useful to researchers.

Repeatability, defined as the among-individual variance (V_I) divided by the total variance (V_P ; (Boake, 1989; Hayes & Jenkins, 1997; Nakagawa & Schielzeth, 2010; Dingemanse & Dochtermann, 2013), is what is known as a variance standardized measure (Houle, 1992). Variance standardized measures have limitations of which potential users should be aware. First, comparing these measures between sexes, among populations, taxa, trait types etc. is problematic. This is because these ratios can differ between groups because of differences in either the numerator or denominator (or both; Fig 1 & 2). Alternatively, this ratio can be the same between groups despite differences in both the numerator and denominator. Whether repeatability is found to be similar or different

among groups is therefore ultimately uninformative without considering the specific changes in variance components.

Second, repeatability is often discussed as a metric of “individual consistency” (de Villemereuil, Morrissey et al., 2017). Unfortunately consistency on its own is difficult to interpret and it is instead consistency relative to the mean that will often be more biologically important. Put another way, how variable individuals are only makes sense on the scale of the behavior of interest. In this regard, the inferential limitations for the use of repeatability are similar to those associated with heritability. For example, as detailed by Houle (1992) and Hansen, Pélabon et al. (2011), standardizing additive genetic variances (V_A) by total phenotypic variances (V_P)—that is, calculating narrow-sense heritability—is not appropriate for most of the questions researchers ask. This assertion is based on the observation that most of the variation expressed in a trait can be genetically based (i.e. high heritability) even when there is little total variation in that trait. In such a case the observed genetic differences result in only small differences in trait values. Similarly, for repeatability, individuals can exhibit high “consistency”, and thus high repeatability, but actually vary little on the scale of the traits of interest. Therefore, repeatability says nothing about whether individuals exhibit large differences in phenotype: individuals can exhibit substantial repeatability but ultimately differ little from one another (e.g. Figure 1C versus 1D). Instead, whether repeatability—which combines additive genetic variation, non-additive genetic variation, and permanent environmental effects (Boake, 1989; Dochtermann, Schwab et al., 2015; Niemelä & Dingemanse, 2017)—represents biologically relevant variation depends on the mean of a trait and how individuals vary around that mean.

One alternative is the use of mean standardization rather than variance standardization. Mean standardization places the observed magnitude of variation directly in the context of the mean. One mean-standardization approach is to divide the among-individual variation (V_I) in a trait by the square of its mean (for applications of this approach to understanding heritability see Houle, 1992; Hansen, Pélabon et al., 2011). For V_I (Fig. 2C):

$$I_I = 100 \times \frac{V_I}{\bar{x}^2}$$

where \bar{x} is the mean of the population.

This mean-standardization can also be applied to any other estimate of variance, including residual variances (e.g. I_R for V_R ; following Houle (1992) and Hansen, Pélabon et al. (2011)) and variances due to fixed effects (I_F for V_F , with V_F calculated following (Nakagawa & Schielzeth, 2013; de Villemereuil, Morrissey et al., 2017). Mean-standardization allows direct comparison of *magnitudes* of variation across traits measured on different scales between groups (Figure 2). In contrast, repeatability only indirectly provides an indication of *relative magnitude*. There are other dimensionless mean standardizations that similarly allow the direct comparison of magnitudes, including the coefficient of variation ($CV, \frac{\sqrt{V_I}}{\bar{x}}$) which is likely more familiar than I (see also Holtmann, Lagisz et al., 2017). We prefer the use of I over CV and other measures because the former is more directly linked to understanding evolutionary responses to selection (Houle, 1992; Hansen, Pélabon et al., 2011). For example, when calculated for additive genetic variance, I estimates the percent change in a trait's mean underselection (Hansen, Pélabon et al., 2011). By using I , the study of individual variation in behavior can therefore be more

directly linked to the broader evolutionary literature. Regardless of whether I or CV is preferred, we encourage authors to report all the unstandardized variance components their statistical models estimate, along with behavioral means, so as to facilitate future meta-analyses.

Mean standardization provides several benefits that complement those of repeatability. Most notably, mean standardization allows variation to be compared across scales of measurement and traits (e.g. time allocated to parental care versus propensity to engage in agonistic interactions), without the assumptions necessary for doing so with repeatability. Mean standardization allows for different biological inferences to be made than can be made based on variance standardization alone. For example, in regards again to heritability, it is a general prediction that traits that are closely linked to fitness will exhibit lower variation and this is generally borne out when examining heritabilities (Mousseau & Roff, 1987; Stirling, Reale et al., 2002). However, when this question is asked using mean standardization it becomes clear that there is considerable additive genetic variation in fitness affecting traits and that the observation of low heritabilities is due to there also being considerable environmental variation in these traits (Houle, 1992; Hansen, Pélabon et al., 2011). Similarly, behavioral traits linked to mating, habitat selection and aggression typically show high repeatability while activity, mate preference and migration have lower repeatability (Bell, Hankson et al., 2009). However, because mean standardized ratios have not typically been reported in behavioral studies, it is not possible to determine how these results translate to magnitudes of behavioral differences among individuals. Whether traits differ in the magnitude of individual differences consequently remains an important open topic for behavioral ecologist to tackle. More generally, the use of mean-

standardization means researchers can ask new and exciting questions about how different types of behaviors differ and the magnitude of effects on behavioral variation.

As an example of how mean standardization complements repeatability, we can revisit data collected (simulated) by Wilson (2008) for a variety of morphological and behavioral traits of unicorns (see also de Villemereuil, Morrissey et al., 2017). Two populations of unicorns exhibited the same average aggression and the same repeatability of aggression ($r = 0.80$, Fig 1). However, unicorns in Population 1 exhibited less variability both among individuals and across repeated measurements of the same individuals than observed for Population 2 (Fig 1 C&D). This difference in variability is apparent when directly comparing variance components (Fig 2A) but obscured under variance standardization (Fig 2B). The use of I values recaptures this difference between the two populations at both the among-individual and within-individual levels (Fig 2C). The use of I therefore highlights that while the unicorns of both populations were equally aggressive and exhibited the same repeatability, individual unicorns in Population 2 differ from each other to a greater degree than did those in Population 1 (Fig 2C). This is an inference that could not have been drawn from the comparison of repeatabilities and suggests interesting future questions regarding the contribution of genetic and environmental factors to aggression in unicorns.

There are, of course, caveats to the use of mean standardization. First, it only applies to traits on ratio or log scales (though the evolutionary implications discussed above and by Hansen, Pélabon et al. (2011) are not applicable on log scales). Ratio scale does not refer to values expressed as ratios but rather to whether how something has an objective zero

value and if the comparison of two measurements of the same kind estimates the ratio of the difference in their magnitude (Houle, 1992; Hansen, Pélabon et al., 2011). This issue of a trait having a real zero is particularly important as researchers frequently standardize measures based on their standard deviations and centered to their mean prior to analysis (e.g. Hadfield, 2010). This approach can facilitate statistical model fitting but prohibits mean standardization and thus limits the inferences that can be made. Finally, while mean standardization can be conducted on the data scale for non-normally distributed traits following de Villemereuil, Schielzeth et al. (2016), its interpretation in such cases is not clear.

The use of variance standardized measures like repeatabilities versus mean standardization has been discussed a great deal in the evolutionary literature, as have the points that we make above (e.g. Houle, 1992; Hansen & Houle, 2008; Hansen, Pélabon et al., 2011; Holtmann, Lagisz et al., 2017). However, given the manner in which repeatability is frequently discussed in animal behavior studies, these points require continued emphasis (see also Wilson, 2018). Importantly, we are not suggesting that behavioral ecologists stop using repeatability, but rather want to emphasize that variance ratios like repeatability *and* mean standardized estimates allow for different inferences. The study of “animal personality” and behavioral variation continues to ask increasingly sophisticated questions regarding when and why behavioral variation is manifested. This expanding range of questions necessitates an expanded quantitative toolbox as not all questions will be properly addressed with repeatability. Mean standardization is one such tool. Which tool should be used fundamentally comes down to the questions being asked.

Thus, given their differences, what questions are appropriate for available approaches? Variance ratios like repeatability allow inferences regarding sources of variation while mean standardized estimates allow inferences regarding the magnitude of variation (Table 1; see also Nakagawa & Schielzeth, 2010; de Villemereuil, Morrissey et al., 2017; Wilson, 2018). For example, if a researcher is interested in whether most variation among individuals is attributable to developmental differences, comparing variance ratios and direct comparison of variances is most appropriate (e.g. Royauté & Dochtermann, 2017; Rudin, Simmons et al., 2018; Royauté, Garrison et al., 2019). If, instead, a researcher is interested in whether individuals exhibit large differences in their behaviors, mean standardization will be most appropriate. Researchers should carefully consider their questions when drawing inferences from these estimates. We encourage researchers to present the most complete information possible. This complete information includes not only derived values like repeatability and mean standardized estimates but their constituent components as well.

References Cited

- Bell, A.M., Hankson, S.J. & Laskowski, K.L. 2009 The repeatability of behaviour: a meta-analysis. *Animal Behaviour*. **77**: 771-783.
- Boake, C.R.B. 1989 Repeatability - Its Role in Evolutionary Studies of Mating-Behavior. *Evol Ecol*. **3**: 173-182.
- de Villemereuil, P., Morrissey, M.B., Nakagawa, S. & Schielzeth, H. 2017 Fixed effect variance and the estimation of repeatabilities and heritabilities: Issues and solutions. *J Evol Biology*.
- de Villemereuil, P., Schielzeth, H., Nakagawa, S. & Morrissey, M. 2016 General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics*. **204**: 1281-1294.

- 157 Dingemanse, N.J. & Dochtermann, N.A. 2013 Quantifying individual variation in behaviour:
158 mixed-effect modelling approaches. *J Anim Ecol.* **82**: 39-54.
- 159 Dochtermann, N.A., Schwab, T. & Sih, A. 2015 The contribution of additive genetic variation
160 to personality variation: heritability of personality. *Proc R Soc B.* **282**: 20142201.
- 161 Hadfield, J.D. 2010 MCMC Methods for Multi-Response Generalized Linear Mixed Models:
162 The MCMCglmm R Package. *Journal of Statistical Software.* **33**: 1-22.
- 163 Hansen, T.F. & Houle, D. 2008 Measuring and comparing evolvability and constraint in
164 multivariate characters. *J Evol Biol.* **21**: 1201-1219.
- 165 Hansen, T.F., Pélabon, C. & Houle, D. 2011 Heritability is not evolvability. *Evolutionary
166 Biology.* **38**: 258.
- 167 Hayes, J.P. & Jenkins, S.H. 1997 Individual variation in mammals. **78**: 274-293.
- 168 Holtmann, B., Lagisz, M. & Nakagawa, S. 2017 Metabolic rates, and not hormone levels, are a
169 likely mediator of between-individual differences in behaviour: a meta-analysis. **31**:
170 685-696.
- 171 Houle, D. 1992 Comparing evolvability and variability of quantitative traits. *Genetics.* **130**:
172 195-204.
- 173 Mousseau, T.A. & Roff, D.A. 1987 Natural selection and the heritability of fitness
174 components. *Heredity.* **59**: 181-197.
- 175 Nakagawa, S. & Schielzeth, H. 2010 Repeatability for Gaussian and non-Gaussian data: a
176 practical guide for biologists. *Biol Rev.* **85**: 935-956.
- 177 Nakagawa, S. & Schielzeth, H. 2013 A general and simple method for obtaining R² from
178 generalized linear mixed-effects models. **4**: 133-142.
- 179 Niemelä, P.T. & Dingemanse, N.J. 2017 Individual versus pseudo-repeatability in behaviour:
180 lessons from translocation experiments in a wild insect. **86**: 1033-1043.
- 181 Royauté, R. & Dochtermann, N.A. 2017 When the mean no longer matters: developmental
182 diet affects behavioral variation but not population averages in the house cricket
183 (*Acheta domesticus*). *Behav. Ecol.* **28**: 337-345.
- 184 Royauté, R., Garrison, C., Dalos, J., Berdal, M.A. & Dochtermann, N.A. 2019 Current energy
185 state interacts with the developmental environment to influence behavioural
186 plasticity. *Anim. Behav.* **148**: 39-51.
- 187 Rudin, F.S., Simmons, L.W. & Tomkins, J.L. 2018 Social cues affect quantitative genetic
188 variation and covariation in animal personality traits.

189 Stirling, D.G., Reale, D. & Roff, D.A. 2002 Selection, structure and the heritability of
190 behaviour. *J. Evol. Biol.* **15**: 277-289.

191 Wilson, A. 2008 Why h^2 does not always equal VA/VP ? **21**: 647-650.

192 Wilson, A. 2018 How should we interpret estimates of individual repeatability?

193

194 Tables and Figures

195

196 Table 1. Variance and mean standardized measurements of variation and example interpretations for which each might be

197 appropriate.

Metric		Interpretation	Reference
Variance Standardized Measures			
Adjusted Repeatability	$\frac{V_I}{V_I + V_R}$	Proportion of variation due to intrinsic differences among individuals (i.e. due to genetic and permanent environmental effects)	(Nakagawa & Schielzeth, 2013; de Villemereuil, Morrissey et al., 2017)
Unadjusted Repeatability	$\frac{V_I}{V_I + V_F + V_R}$	Proportion of total variation that is attributable to differences among individuals conditioned on measurable, fixed effects (VF). Fixed effects might include either procedural/experimental controls or biological	(Nakagawa & Schielzeth, 2013; de Villemereuil, Morrissey et al., 2017)
Mean Standardized Measures			
Mean-Scaled Individual Variation (I_I)	$100 \times \frac{V_I}{\bar{x}^2}$	Percentage of variation relative to the mean explained by differences in trait expression among individuals	this manuscript
Mean-Scaled Fixed Effect Variation (I_F)	$100 \times \frac{V_F}{\bar{x}^2}$	Percentage of variation relative to the mean explained by differences in trait expression due to fixed effects	this manuscript
Mean-Scaled Residual Variation (I_R)	$100 \times \frac{V_R}{\bar{x}^2}$	Percentage of variation relative to the mean explained by differences in trait expression due to unmeasured sources of variation	(Houle, 1992; Hansen, Pélabon et al., 2011)

198

Figure 1. Populations can have the same repeatability yet have drastically different magnitude of among-individual variance relative to their means. For example, two populations of unicorns studied by Wilson (2008) had the same average aggression and the same repeatability despite different amounts of variability (A versus B). This can be more clearly observed (C versus D) if one hundred randomly selected individuals from population 1 have their aggression measured three times (lines in C) and one hundred randomly selected individuals from population 2 are likewise measured three times (lines in D). The comparison of individuals from population 1 (C) and from population 2 (D) demonstrates that, despite the equal repeatability, individuals from population 2 exhibit much larger differences from each other than do unicorns in the first population.

Figure 2. Total (A) and proportional (B) variation explained by among- and within-individual sources of variation (C) mean standardized estimates (error bars are the middle 95% of estimates based on resampling from Figure 1 A and B) in each population of unicorns from Figure 1. Despite equal repeatability (B), population 2 has a greater magnitude of among-individual variation relative to its mean (C). Thus, variation in this trait for this population is expected to be more strongly influenced by genetic and permanent environment sources of variation.

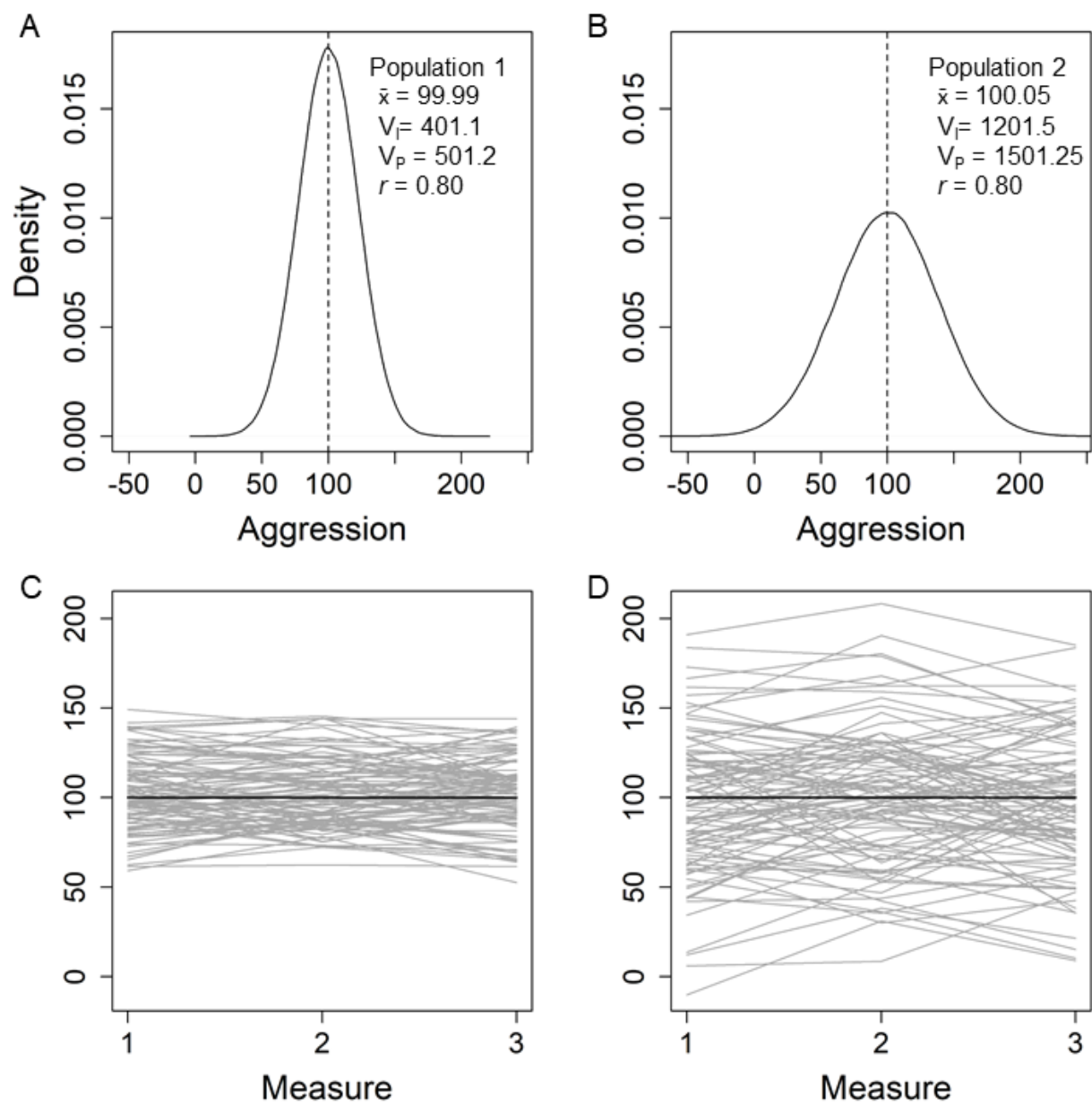
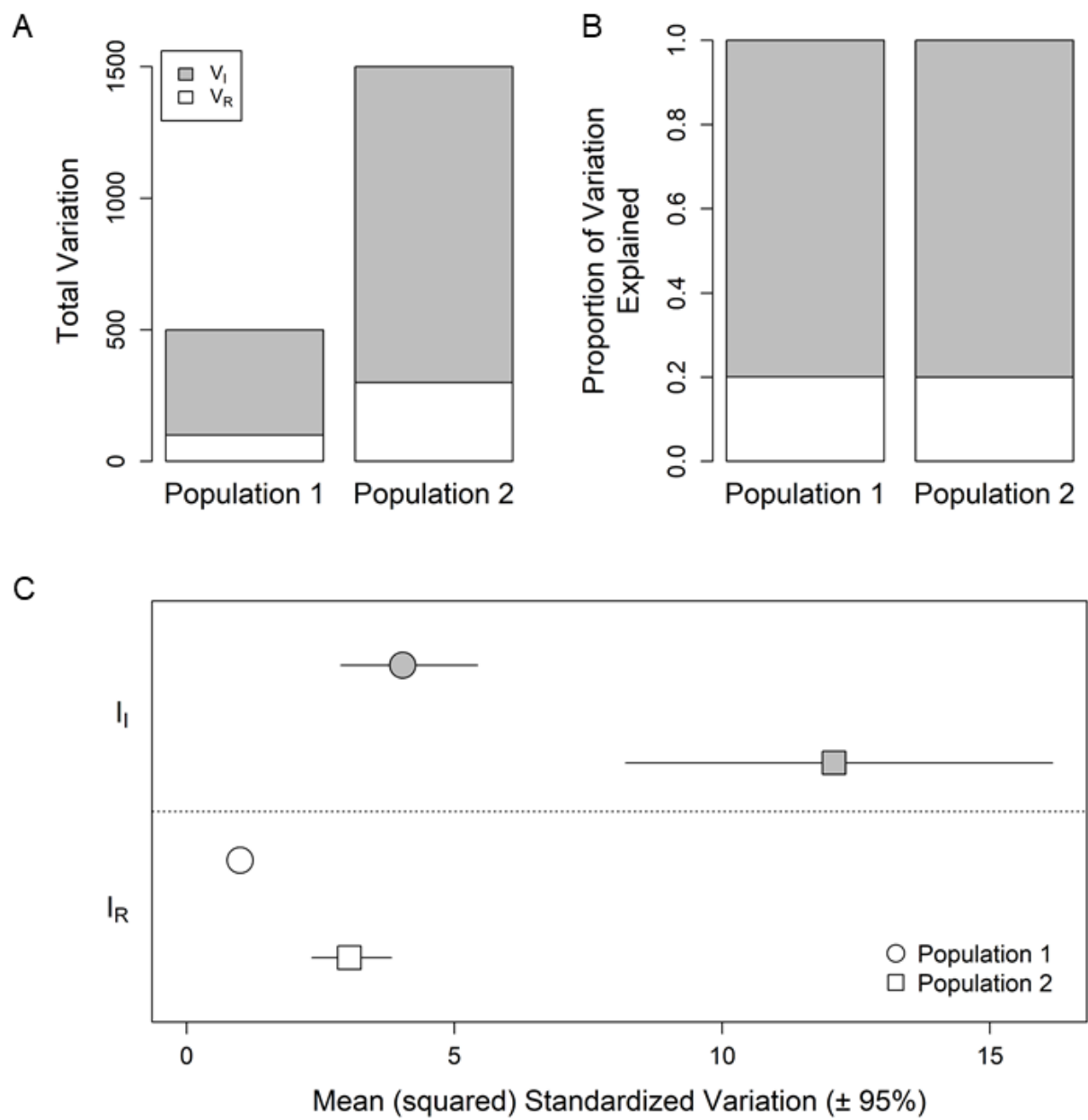


Figure 1.



221 Figure 2.
222