Studies in History and Philosophy of Science 91 (2022) 201–210

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Unifying heritability in evolutionary theory

Pierrick Bourrat [a,b]

[a] Macquarie University, Department of Philosophy, North Ryde, NSW 2109, Australia
[b] The University of Sydney, Department of Philosophy & Charles Perkins Centre, Camperdown, NSW 2006, Australia

ARTICLE INFO

Keywords:
Heritability
Causation
Regression
Environment
Microbiome

ABSTRACT

Despite being widely used in both biology and psychology as if it were a single notion, heritability is not a unified concept. This is also true in evolutionary theory, in which the word 'heritability' has at least two technical definitions that only partly overlap. These yield two approaches to heritability: the 'variance approach' and the 'regression approach.' In this paper, I aim to unify these two approaches. After presenting them, I argue that a general notion of heritability ought to satisfy two desiderata—'general applicability' and 'separability of the causes of resemblance.' I argue that neither the variance nor the regression approach satisfies these two desiderata concomitantly. From there, I develop a general definition of heritability that relies on the distinction between intrinsic and extrinsic properties. I show that this general definition satisfies the two desiderata. I then illustrate the potential usefulness of this general definition in the context of microbiome research.

1. Introduction

The term 'heritability' and its cognates are widely used both in everyday language and in diverse fields in the life sciences, including evolutionary biology, genetics, and psychology. When asking whether a trait is heritable in everyday language, one typically wants to know whether this trait is transmitted across generations so that offspring resemble their parents (Fox Keller, 2010, chap. 3). Following a more genetic-centred usage of the term, heritability is often associated with the process of genetic transmission. That it recurs over generations does not entail that a trait is heritable—sharing genes with one's parents must be the reason the phenotype is transmitted. (Lynch & Walsh, 1998, pp. 170–175).

Following the development of biometry and evolutionary theory (in particular of population and quantitative genetics), the technical notion of heritability acquired several formal definitions (see Jacquard, 1983, for a detailed analysis of these alternative definitions). In quantitative genetics, the heritability of a phenotype P in a population that depends on two independent variables G—the genotype—and E—an environmental deviation—so that $P = G + E$, is defined as the ratio of genotypic variance to phenotypic variance (Falconer & Mackay, 1996). From this definition, a further distinction between 'broad-sense' and 'narrow-sense' heritability is classically made. Broad-sense heritability, h^2_G, corresponds to the ratio of *total* genetic (or genotype) variance, Var(G), to phenotypic variance, Var(P). Formally, we have:

$$h^2_G = \frac{\text{Var}\ (G)}{\text{Var}\ (P)}. \tag{1}$$

Narrow-sense heritability, h^2_A, corresponds to the ratio of *additive* genetic variance, Var(A), to phenotypic variance. It follows a model in which the genotype can be separated into two factors A—additive—and I—non-additive (due to gene-gene interactions)—so that $G = A + I$. Formally, we have:

$$h^2_A = \frac{\text{Var}\ (A)}{\text{Var}\ (P)}. \tag{2}$$

Hereafter, I refer to broad-sense (Equation (1)) and narrow-sense (Equation (2)) heritability as falling under the 'variance approach' to heritability.

Both broad-sense and narrow-sense heritability can be used to assess the extent to which some parent-offspring resemblance should be attributed to genes as opposed to the environment. Broad-sense heritability is relevant when reproduction is asexual (Lynch & Walsh, 1998, pp. 592–595). Because asexual organisms transmit their *whole* genome across generations rather than some shuffled part of it (following meiosis in sexual organisms), gene-gene interactions are expected to be the same in parents and offspring (assuming no mutations). Thus, the effects they produce on a phenotype should be included in a measure of heritability if it is used as a measure of parent-offspring resemblance due to genes. Some have connected broad-sense heritability to genetic causation (see, for instance, Sesardic, 1993, 2005; Oftedal, 2005;

E-mail address: p.bourrat@gmail.com.

https://doi.org/10.1016/j.shpsa.2021.10.019
Received 13 July 2020; Received in revised form 25 October 2021
Available online 27 December 2021
0039-3681/

Tabery, 2014; Lynch & Bourrat, 2017; Tal, 2009; Bourrat, 2021). Others, following Lewontin (1974), have denied that such an interpretation is warranted (for analyses of the debate, see Oftedal, 2005; Tabery, 2014).

Narrow-sense heritability is relevant when genes are shuffled at every generation so that, in the long run, one can consider that a given allele at locus $L1$ has no more chance to interact with an allele at locus $L2$ than it has with any another allele at the same locus. Such a situation is a good approximation of what happens in sexual organisms (Falconer & Mackay, 1996, chap. 10), although that involves many caveats I will not discuss here.

In the biometric tradition, heritability, h_b^2, is defined as the regression slope of average offspring phenotype on average parental phenotype without reference to genes (Jacquard, 1983). Following the least-squares method (Lynch & Walsh, 1998, pp. 39–42), the regression slope of a variable X on another variable Y, β_{XY}, is equal to the covariance between the two variables, divided by the variance of Y, so that $\beta_{XY} = \frac{\text{Cov}\ (X,Y)}{\text{Var}\ (Y)}$. In the case of heritability, assuming asexual reproduction, we have:

$$h_b^2 = \frac{\text{Cov}\ (P', P)}{\text{Var}\ (P)}, \tag{3}$$

where P' is the average offspring phenotype coming from a parent. Hereafter, I refer to Equation (3) as the 'regression approach' to heritability.

Facing these different technical definitions of heritability, novices and sometimes more expert readers might feel uncertain about the differences and similarities between these definitions. Do they play the same explanatory role? Are they genuinely different? Do they rest on the same conceptual foundations? How do they relate to genetic causation and genetic transmission? In this paper, I propose a new definition of heritability that is motivated by the aim to address an unresolved tension existing between the variance and regression approaches. In some settings, the two approaches yield different answers about the heritability of a trait. If one assumes, as I do, that heritability is a single concept or ought to refer to a single value in a particular setting, this is problematic. In developing my new definition, I will provide novel answers to the aforementioned questions.

To meet my goal, I start by discussing the vexed topic of heritability and genetic causation. I argue that, following the variance approach, heritability can be connected to genetic causation following the interventionist or difference-making account of causation. I then link this conception of heritability to genetic transmission. Next, I present a setting in which the variance and the regression approach arrive at different conclusions about the heritability of a trait. I argue that having two different heritability values for the same setting undermines our capacity to assess the extent to which natural selection operates in this setting. From there, I set two desiderata that an adequate heritability concept ought to satisfy. I call these 'general applicability' and 'separability of the causes of resemblance.' I show that these desiderata are concomitantly satisfied by neither of the two approaches. This leads me to reject the two approaches as the preferred one in the context of evolutionary theory. I present an alternative definition of heritability that meets these two desiderata. This definition relies on the distinction between extrinsic and intrinsic properties. I elaborate on this definition for situations in which different causes of phenotype are not independent. Finally, I show the potential relevance of this general definition in the context of microbiome research.

2. The variance approach and its relations with genetic causation and transmission

In this section, I start by analysing in what sense broad-sense heritability is associated with genetic causation and extend my reasoning to narrow-sense heritability. I then show how genetic causation and transmission are connected. Before delving into the analysis, it should be noted that, when referring to the variance approach of heritability, one can focus exclusively on the notion of genetic causation rather than transmission. This is particularly the case with the use of broad-sense heritability in behavioral genetics (Bourrat, 2021; Downes & Matthews, 2020; Griffiths & Stotz, 2013; Lewontin, 1974; Lynch & Bourrat, 2017; Oftedal, 2005; Sesardic, 1993, 2005; Tabery, 2014). In this discipline, heritability is used to estimate the extent to which phenotypic variation, at the population level, is statistically associated with genetic variation when all gene-gene interactions are considered. Whether such a statistical association should be interpreted causally has been a topic of controversy.

The reasoning behind this use of heritability is the following. At birth, individuals are endowed with genes that come from their parents, and they develop in a particular environment. Within a causal interpretation of the concept, a heritability measure is supposed to tell us how much a phenotype causally depends on the genotype rather than the environment in which the individual develops. In a population, the more deviations from the mean population phenotype are due to genotypic deviations, the higher those phenotypic deviations are 'caused' by genotypic deviations. This translates into a high broad-sense heritability value. In contrast, when phenotypic deviations are small in comparison to phenotypic deviations due to the environment or there are none, broad-sense heritability is low or nil.[1]

One potential objection to the definition of 'genetic causation' as 'deviation from the mean phenotype' is that it does not correspond to the typical notion of genetic causation. When asking whether genes or the environment cause a phenotype, one might intuitively think that it means a phenotype could develop in the absence of one or the other. It should be clear, however, that it cannot be so when referring to genetic causation in the context of heritability. It does not make sense to ask what a phenotype with *no* genotype or developing in *no* environment would be. An organism, by definition, is constituted by a genotype *and* it develops in an environment. Thus, both genes and the environment are contributors to an organism's phenotype. This point has been widely acknowledged in the literature. It is referred to as the 'interactionist consensus' by Sterelny and Griffiths (1999, pp. 97–100).[2]

What do 'genetic' and 'environmental' causes of phenotype mean in the context of heritability, if not the presence or absence of the environment/genes for an individual? One possible answer is that a genetic cause can be characterised as *changes* in the values of the genotype producing *changes*[3] in the value of the phenotype in a population. *Mutatis mutandis*, an environmental cause can be characterised as a change in the value of the environment. A deviation from the mean is one way to operationalise those changes in value. Elsewhere, I provide a systematic defense of this interpretation (see Bourrat, 2021).

Such a conceptualisation of genetic causation can be framed within the difference-making account of causation (Woodward, 2003, 2010; Bourrat, 2020, 2021; Pearl, 2000; Lynch & Bourrat, 2017; Lynch, 2021).[4] Following this account, causation in a minimal sense between two variables is established when the change in the value of one variable (say X)

[1] Note that environmental deviations are measured as phenotypic deviations that remain once deviations that can be explained by individuals sharing a portion of their genotype with their kin have been taken into consideration. An immediate implication is that if parents not only pass on their genetic material (DNA) to their offspring but also other factors, these factors might be counted as 'genetic' (for developments of this idea see Bourrat & Lu, 2017; Lu & Bourrat, 2018).

[2] Fisher was very well aware of this when writing his seminal paper on the analysis of variance: 'loose phrases about the "percentage of causation" which obscure the essential distinction between the individual and population should be avoided' (Fisher, 1918, pp. 399–400).

[3] Assuming a range of possible changes (those observed in real populations).

[4] Note in passing that deviations from the mean as a way to approach causation is not at all restricted to genetic causation since it is at the basis of ANOVA and regressions that are widely used in many sciences.

produces a change in the value of another at a later time (say *Y*), provided the change in *X*—known as an 'ideal intervention' in this literature—did not lead to any other change at the time it was made. Sometimes the change is from 'presence' to 'absence,' but this need not be. It can be between different values of *X*. This account of causation has received both verbal (e.g., Waters, 2007; Woodward, 2003, 2010) and more formal treatments (e.g., Bourrat, 2019b; Griffiths et al., 2015; Pearl, 2009; Pocheville et al., 2017; Spirtes et al., 2000).

When one applies the interventionist account to the genotype-phenotype relationship in a population, an ideal intervention on genotype leading, on average, to a given change in phenotype, establishes the extent to which a variation in genes cause a variation in a phenotype in this population. The ratio of the two changes can then be used as a measure of the relative influence of genes on the phenotypic differences as opposed to environmental ones. While the connections between heritability and the interventionist account of causation are more subtle, this is, in essence, is what broad-sense heritability captures. In Bourrat (2021), I provide a detailed analysis that establishes the links between the interventionist account and heritability formally. Due to lack of space, I will not reiterate it here. However, I will respond to some potential objections to this way of conceptualising heritability.

First, one might think that there is an important impediment to applying the interventionist reasoning to any real situation. They might reason that performing an ideal intervention on the genotype (or any other variable for that matter) of an individual would require the ability to *physically* alter its genotype, observe the change in phenotype, and compare it to the phenotype, had the genotype of this individual not been changed. However, since this procedure is physically impossible, does it not then prevent one from establishing causation, not only *genetic* causation but causation *tout court*, using the interventionist account?

To respond to this potential objection, let me first note that there are at least two projects one might pursue with respect to genetic causation and heritability. One of these projects is purely conceptual. It aims to answer the following question: 'can broad-sense heritability be connected to causation (especially the interventionist account), and if yes, under which conditions?' This is the project I am concerned with here.

The second project is more epistemic and aims to answer the following question: 'assuming a phenotype under genetic influence, does an estimate of broad-sense heritability obtained from a set of data characterise this causal relationship adequately?' The two projects are connected because whether an estimate is judged adequate will depend on whether it satisfies the conditions for heritability to be linked to causation. However, one might be purely interested in the first question without needing to refer to any data or epistemic considerations.

Coming back to the concern that an ideal intervention is not physically possible, one can respond as follows: while it is true that ideal interventions are physically impossible, one can nevertheless *conceptualise* that they are possible, and that enough to have a causal explanation. Thus, one can ask what would be the effect of altering the genotype of an individual without having to worry about whether this can be done in the physical world. This fits squarely with the first of the two projects. That it is physically impossible to alter a genotype and yet yield causal explanation should not be viewed as surprising. The interventionist account is, indeed, part of a broader family of accounts of causation, namely, counterfactual accounts. The term 'counterfactual' literally means against the fact (i.e., what is observed). To estimate the extent to which genes are a cause of phenotype, one can develop methods to *emulate* an ideal intervention in the physical world, either by designing experiments or from observed data. Using these methods, one can know with some degree of confidence what the outcome of such an intervention would be without it taking place. With both experiments and observational studies, a number of methodological precautions must be taken to be able to infer causation from data. When the data is purely observational, as is typically the case with human traits, this inference is especially perilous. This point applies for *any* observational data, not solely with respect to heritability. I will not be concerned here with those precautions because estimating heritability belongs to the second of the two projects mentioned above.

A second type of objection within the project of linking heritability to causality conceptually concerns the existence of such a link in conditions of gene-environment interaction. It has been noted numerous times that heritability can be connected to genetic causation *only* when there is no gene-environment statistical interaction (i.e., gene and environment are additive causes of phenotype) (Griffiths & Stotz, 2013; Lewontin, 1974; Sesardic, 2005; Tal, 2012). One consequence of the existence of statistical gene-environment interactions (as well as different genotypes and environments in different populations) is that the causal interpretation of heritability can only be a local one—that is, one that does not extend to populations with different genotypes and environmental conditions. This means that interpreting heritability causally in situations of substantial gene-environment interaction is not warranted. This point is independent of any epistemic considerations (see Taylor, 2006, 2010, for some related problems).

There are two things to note in regard to this objection. First, this objection is not specific to heritability. The same difficulties of characterising a causal relationship with a single number will be encountered in any situation of interacting causes (Bourrat, 2021; Sesardic, 2005; Tal, 2009).

Second, the interventionist account has some resources permitting us to partly address this problem. It allows characterising and comparing causal relationships over different 'dimensions,' such as 'invariance,' 'stability,' and 'specificity' (see Woodward, 2010, 2003; Pocheville et al., 2017, for verbal and formal attempts). To give an example, a causal relationship might be very unstable. That is, it might not hold in most backgrounds. However, following Woodward's analysis, this lack of stability does not imply the relationship is noncausal. It just scores lower on that dimension than a more stable one. The same applies to other dimensions. In my recent analysis, I show that statistical interaction and locality can be viewed from the perspective of Woodward's notions of stability and invariance (see Bourrat, 2021). Thus, despite the difficulties of characterising relationships in which there is substantial gene-environment interaction, this is possible provided that more information about the relationship is given. In other words, there is scope to develop more complex metrics based on heritability for situations of gene-environment interactions. Doing so, however, would go beyond the scope of this paper. For simplicity, my analysis will assume no gene-environment statistical interaction. This is a classical assumption made the basic 'ACE' model of behavioral genetics (see Knopik et al., 2016, appendix). In Section 5, I will take up the problem of gene-environment covariance, which is another way in which genes and environments might not be 'independent.'

Admittedly, the point that local relationships are nonetheless causal following a nuanced view of causation means that extrapolating the heritability measure from one population to another is generally unwarranted. A further response to this objection might be that it is unclear that within-population or local measures, in and of themselves, have no value, as, for instance, the successful use of heritability estimates in breeding programs attests (see Lynch & Walsh, 1998, p. 5, for a brief overview).

Having drawn a link between *broad-sense* heritability and genetic causation, can this be extended to narrow-sense heritability? The only difference between narrow- and broad-sense heritability is that the former does not account for the influence of gene-gene interactions (i.e., dominance and statistical epistasis) on phenotype. This means that narrow-sense heritability can be interpreted causally using the same reasoning as for broad-sense heritability but with a slightly different meaning of 'genetic causation.' In the context of narrow-sense heritability, 'genetic causation' refers to 'genetic causation once gene-gene interactions have been discounted.' It should be noted that there are specific epistemic problems with the attempts to eliminate all the effects of gene-gene interactions that causally influence a phenotype. For a discussion in the context of the missing heritability problem, see Zuk

et al. (2012). As with broad-sense heritability, I will not tackle these issues here.

Thus far, I have argued that the variance approach to heritability can be connected to genetic causation when measured over the lifetime of an organism—that is, how genotypes that organisms are endowed with at conception can causally influence their phenotypes. I now ask how this is related to transmission and parent-offspring similarity. Following the variance approach (whether broad-sense or narrow-sense heritability), genes happen to be transmitted from parent to offspring.[5] Assuming that, everything else being equal, the same genotype or allele has a similar effect on a phenotype in both parents and offspring (or later generations), one can deduce the extent to which a parent-offspring resemblance (assuming they live on average in the same environmental conditions) is genetic, provided we know the portion of the genotype transmitted from parents to offspring. When the resemblance is high, assuming parents transmit their whole genotype to offspring, we can conclude that this resemblance is almost entirely due to genes. This translates into a high heritability. In contrast, when the same offspring do not resemble their parent, we can deduce that genes play no role in the determination of the phenotype. In such cases, heritability is low or nil. Thus, heritability can be used both as a measure of genetic causation and also phenotypic resemblance due to genetic transmission, where the latter is inferred by the reoccurrence of genotypic effects on phenotypes at each generation.

To sum up, heritability, following the variance approach, can be regarded as an adequate measure, in some conditions, of genotypic causation *qua* context-dependent (i.e., local) difference-making causation. Further, genetic causation can serve as the basis of measuring the extent to which the similarity between parents and offspring is genetic. However, as we will soon appreciate, this conception of heritability is in tension with the regression approach. To calculate the heritability of a trait using the regression approach, one does not need to assume that any gene is causally related to a phenotype or transmitted from parent to offspring. To unveil this tension and what it implies, I next present a setting in which, depending on whether one uses the regression or the variance approach to heritability, a different value for heritability is obtained. In Section 4, I alleviate the tension between these two conceptions by proposing a more general definition that satisfies two desiderata one can reasonably demand for a sound concept of heritability.

3. Conflicting heritabilities

3.1. A problem

To show what separates the variance from the regression approach, let us examine the following biological setting.

Let us suppose a species of plant with individuals living in two possible environments: 'rich' and 'poor.' Assume that individuals in the rich environment are *ceteris paribus* taller than those living in the poor environment. Assume now a population of genetically identical individuals of these plants with offspring plants living in the same environment as their parents. Finally, assume that reproduction is asexual and that parents transmit their genetic material (DNA sequences) faithfully to their offspring.

What is the heritability of height in this population following the regression approach? Recall that it is a measure of similarity between parents and offspring in the following sense. A trait is heritable if parents which deviate by x units from the mean value of the trait in the population are more likely than other individuals with a different phenotype to produce offspring which deviate by x units from the mean value of the trait in the offspring population.

This response is in sharp contrast with the answer one would obtain using the variance approach. Whether using the broad-sense or narrow-sense definition of heritability, this answer would be that heritability is nil in the above example. Why? Simply because we assumed that the population is composed of genetic clones. Thus, the genotypic variance and consequently the additive genotypic variance in this population are both zero. It follows that the ratios of these variances to phenotypic variance, which constitute the definitions of heritability (broad-sense and narrow-sense, respectively) following the variance approach, are nil.

Thus, we have two answers for the heritability value of the same trait in the same population. These conflicting answers can have significant implications beyond the concept of heritability. To see that, assume now that tall individuals produce more offspring than short individuals. The two approaches would lead to different answers about whether height evolves by natural selection in the population since heritability is a necessary ingredient for evolution by natural selection (Bourrat, 2014, 2015b; Brandon, 1990; Godfrey-Smith, 2007; Lewontin, 1970; Mameli, 2004; Okasha, 2006). This is an unsatisfactory situation since whether evolution by natural selection occurs is a matter of fact rather than definition. If this reasoning is correct, there are two possible conclusions: either one of the two definitions should be favoured over the other, or neither of the two definitions is adequate, at least in the context of evolutionary theory. As we shall see, many authors have argued for one of the two definitions being the correct one. I will argue that this is not so, and a third definition of heritability is preferable.

3.2. Uncovering the tension

One way to uncover the origin of the tension between the regression and variance approaches to heritability is to notice that they are formally equivalent when some assumptions (in particular about the environment) are made (Okasha, 2010).

Recall the definition of heritability following the regression approach:

$$h_b^2 = \frac{\mathrm{Cov}\,(P', P)}{\mathrm{Var}\,(P)}. \tag{3}$$

Following the assumptions of the height example presented above, individuals breed true. Following the model used with the variance approach, we have $P = G + E$ and $P' = G' + E'$, where G' and E' are the offspring genotype and environmental deviation, respectively, and, since individuals reproduce asexually, we have $G' = G$. Consequently, making those assumptions, we can develop Equation (3) as:

$$h_b^2 = \frac{\mathrm{Cov}\,(G + E', G + E)}{\mathrm{Var}(P)}. \tag{4}$$

Since the covariance of a variable with itself is its variance, using the distributive property of variance and covariance, Equation (4) can be rewritten as:[6]

$$h_b^2 = \frac{\mathrm{Var}\,(G) + \mathrm{Cov}\,(G, E) + \mathrm{Cov}\,(E', G) + \mathrm{Cov}\,(E', E)}{\mathrm{Var}\,(P)}. \tag{5}$$

It can be deduced from this equation that the variance heritability and the regression heritability are equal when we have $\mathrm{Cov}(G, E) = \mathrm{Cov}(E', G) = \mathrm{Cov}(E', E) = 0$. In words, the two definitions of heritability are equivalent in a genetic context when there are no gene-environment correlations (both between the genotype of the parent and its environment, and the environment of its offspring) and no correlation between the parental and offspring environments (see also Okasha, 2010, Appendix).

It can also be deduced that variance heritability is a *component* of regression heritability in biological systems, since we have:

[5] Depending on whether organisms are asexual or sexual, not all genes are transmitted from parent to offspring.

[6] (Okasha, 2010, Appendix, arrives at the same form, except he uses the *narrow*-sense rather broad-sense heritability).

$$h_b^2 = h_G^2 + \frac{\mathrm{Cov}(G,E) + \mathrm{Cov}(E',G) + \mathrm{Cov}(E',E)}{\mathrm{Var}(P)}. \quad (6)$$

In the height example above (Fig. 1), G and G' are constant because individuals are clones. This means that Var(G), Cov(G, E), and Cov(E', G) are all nil. However, since the environment of the parents is the same as that of their offspring, we have Cov(E', E) > 0. Thus, heritability of height, when measured with the regression approach, is positive in our example solely due to the parent-offspring environmental covariance being positive.

We now have a clear picture of the relationship between the two approaches to heritability. However, this does not tell us whether one ought to be chosen over the other. In the next section, I argue that neither of the two approaches is satisfactory in general. They both fall short of one of two desiderata one can reasonably demand for a general concept of heritability tailored to evolutionary theory. From there, I propose a general account of heritability that satisfies the two desiderata.

4. The best of both worlds?

4.1. Neither the variance nor the regression definition

My analysis thus far has revealed that a single trait can have a different heritability depending on which approach is used. Facing this tension, one might be tempted to argue that one is better than the other. Some authors in the heritability literature have indeed argued that one approach is more fundamental than the other. For instance, some quantitative geneticists, such as Lynch and Walsh (1998, pp. 170–171), consider that regression heritability is only an *estimate* of variance heritability.[7] Their rationale is that if heritability is a measure of *genetic* parent-offspring resemblance and the regression definition can include other terms, it cannot be regarded as defining the genetic component of the resemblance. From a purely genetic perspective, this sounds right.

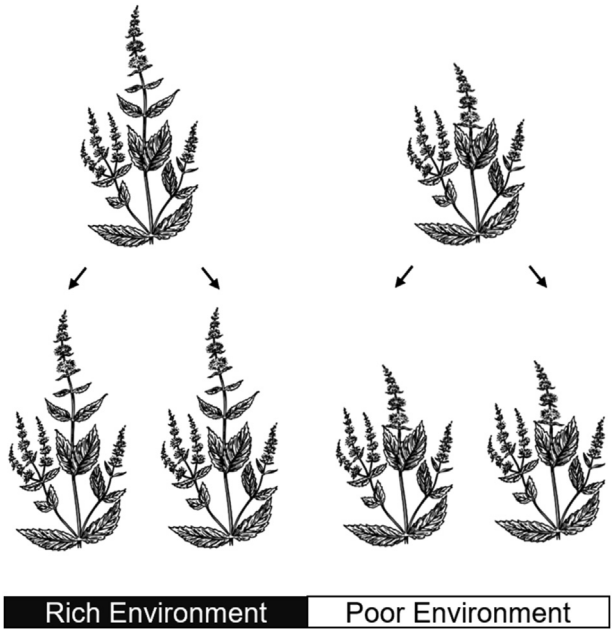

Fig. 1. Setting leading to conflicting heritability measures. Individual plants are genetic clones, but they live in different environmental patches ('rich' or 'poor'). If there is a correlation between the parental and offspring environment and the phenotype (here height) depends on the variation in the environment, applying the regression definition yields a positive answer while applying the variance (broad-sense) definition leads to a nil heritability.

However, the requirement that heritability refers to genetic variance prevents one from using the notion of heritability in a nongenetic context, as pointed out by Okasha (2010), Rice (2004, pp. 203–205), and Mameli (2004). For instance, using the variance approach, the heritability of a cultural trait would be nil because cultural entities are nongenetic. Yet, one might believe that cultural traits, and other epigenetic traits considered broadly, are heritable and can be selected. Restricting the definition of heritability to *genetic* variance arbitrarily limits the legitimate usage of this concept in different domains. Further, the regression definition of heritability is the one recovered from the most abstract equations of evolutionary change, such as the Price equation (Price, 1970), and thus the one truly relevant in the context of evolutionary theory. These considerations lead Rice (2004, p. 204) to consider that 'it is more appropriate to think of all measures of additive genetic variance as estimators of parent-offspring covariance.'[8]

It should also be pointed out that considering a definition of heritability in terms of genetic variance requires that everyone agree about what 'genetic' refers to when talking about genetic variance. Unfortunately, as discussed in Griffiths and Neumann-Held (1999), Griffiths and Stotz (2006, 2013), Moss (2003), and Lu and Bourrat (2018), such is not the case. One area in which not making this distinction clear has created some confusion is the literature on the missing heritability problem and heritability estimates derived from genome-wide association studies (see Bourrat & Lu, 2017; Bourrat et al., 2017; Bourrat, 2020; Turkheimer, 2011; Matthews & Turkheimer, 2021; Yang et al., 2010; Maher, 2008).

Because the variance approach only refers to genes, some have deemed the regression approach to be more general (see Godfrey-Smith, 2007; Okasha, 2010; Rice, 2004). Although these authors have a point, it is important to note that a sound definition of heritability cannot *purely* be a matter of parent-offspring similarity if heritability is used to separate environmental from nonenvironmental causes of phenotypic variation. Arguing the contrary would amount to claiming that the distinction between environmental and nonenvironmental causes of phenotypic variation is irrelevant to heritability (and consequently natural selection), which is unsound. Yet, using the regression approach to heritability does not allow for such a distinction.

With all things considered, I argue that neither the variance nor the regression approach can be regarded as the *true* definition of heritability. The regression approach is general so that it can be applied to any situation in which there are a parent and an offspring population. Still, it is partly unsatisfactory because it only relies on parent-offspring similarity without an explicit model of the causes of similarity. The variance approach relies on an explicit model, but it excludes any nongenetic factors.

Before moving further, I should note that I have assumed thus far that heritability ought to refer to a single concept or definition when the word is used within evolutionary theory. One reason for this assumption is because, as argued above, I consider that whether natural selection occurs should not depend on the definition one uses. However, some might argue that the variance and regression approach to heritability are not *definitions* of heritability but merely statistical tools that permit one to *infer* the true heritability of a trait. The idea here, roughly, is that heritability is a single concept (as I have assumed thus far), but there are different methods to infer it. This question leads us back to the distinction between epistemic and conceptual considerations made earlier. While I argue that there are currently different *concepts* of heritability, someone might disagree and claim that what I call 'concepts' with precise definitions are only *tools* that capture imperfectly, and with different

[7] They argue that it is an estimate of narrow-sense heritability because they frame their discussion considering diploid sexual organisms. In the context of my discussion, regression heritability would be an estimate of broad-sense heritability.

[8] Note here again that Rice uses additive genetic variance because he has in mind sexual organisms. This term would be replaced by *total* genetic variance in our example of asexual organisms.

imperfections, a single concept. However, I disagree with this 'pragmatic' position. As mentioned above, the regression approach does not refer to genes or any factor in particular, while the variance approach does so explicitly. Both approaches have thus different referents. I assume that a single referent corresponds to a single concept.

Another position, seemingly compatible with what I have argued so far, might be to claim that, because the two definitions of 'heritability' have different referents, there cannot be a unified concept of heritability. In principle, I have no dispute with this pluralist approach. One might want to use the word 'heritability' in different contexts for different purposes. However, I consider that within a single field or area of research, and particularly in evolutionary theory, the word 'heritability' ought to correspond to a single concept. As I illustrated above with the plant example and quotes from prominent evolutionary theorists, this is not so, and this lack of unity can be the source of tension about whether natural selection operates in a population. Some might be comfortable with this idea. I am not, and as I show below, a general definition of heritability that yields a unique answer as to whether natural selection operates in a population exists.

4.2. Two desiderata for heritability

Faced with the inadequacies of both the variance and the regression approach to capture a general definition of heritability within evolutionary theory, we can delineate two desiderata for such a definition:

1. **General applicability**—that is, applicability beyond genetics (satisfied by the regression but not the variance approach).
2. **Separability of the causes of resemblance**—that is, an explicit model for separating environmental from other causes of phenotype: namely, causes that can be attributed to the entities studied rather than their environment (satisfied by the variance but the not the regression approach).

In this section, I develop a framework that allows for a definition satisfying both desiderata. To do so, let us start by defining the phenotype, P, of an entity, as the outcome of two additive and independent types of factor, namely intrinsic factors, I, and extrinsic factors, $\mathcal{E}$. These two assumptions guarantee that, at the population, there is no statistical interaction between I and $\mathcal{E}$ and the two variables are not correlated. Formally, we have:

$$P = I + \mathcal{E}. \tag{7}$$

Mutatis mutandis, we define the average offspring phenotype, P', as:

$$P' = I' + \mathcal{E}'. \tag{8}$$

What do I mean by 'intrinsic' and 'extrinsic' factors? The distinction between intrinsic and extrinsic properties is seemingly intuitive, yet difficult to define precisely in metaphysics (Marshall, 2016; Weatherson & Marshall, 2017). Following and updating the work of Godfrey-Smith (2009) and Bourrat (2015a, 2017, 2019a) on a different but related topic, I will use the distinction pragmatically. This should not be taken, however, as meaning that this pragmatic version of the distinction is easy to pin down. We will see in the next section that some difficulties surrounding a generalised and causal concept of heritability are due precisely to the difficulties with capturing this distinction.[9]

By 'intrinsic factor,' I mean any property of an entity that does not depend on the existence and arrangement of other things (including other entities). Thus, an extrinsic factor, by deduction, is a property that depends on the existence or arrangement of other things. By 'dependence,' I mean here both supervenience and causal dependence.[10] Intrinsic and extrinsic factors permit us to distinguish properties that should be attributed to the entities studied from those that should be regarded as part of their environment.

That said, I use the distinction in a relative sense. This is so because, in a strict or absolute sense, there are no intrinsic properties of entities that are part of evolving populations, as I defined them. Why? Notice that for any member of a lineage to have *any* phenotype, they *depend* on the existence of at least one parent. Further, at any point in time, any biological individual requires some resources to survive and exhibit a phenotype. Thus, *any* biological phenotype *depends* on the existence of some resources. Therefore, an absolute notion of intrinsically is a nonstarter: any property of a biological entity is an extrinsic property. This last remark is related to the point made in Section 2 about the interactionist consensus—namely, that when a trait is said to be entirely 'genetic' or 'environmental,' it does not mean that an individual bearing this trait could have no genes or live in no environment.

Although we cannot define an intrinsic property in an absolute sense, we can define it in a relative sense as follows. An intrinsic property is one that does not depend on the arrangement and existence of other things *differently* than it does for any other entity of the population considered, which is equivalent to a *ceteris paribus* clause. This notion is relative because the same property might be considered as intrinsic or extrinsic depending on the parameters of a population. *Mutatis mutandis*, a relative extrinsic property is an entity's property that depends on the arrangement and existence of other things *differently* than it does for any other entity of the population considered.[11]

To illustrate the difference between intrinsic and extrinsic relative properties, let us assume a population in which every individual is composed of a certain percentage of water. Although the availability of water might vary in different environments, suppose that the value observed in all members of this population in normal conditions does not vary within this range of conditions. In this population, the property 'water composition' is a (relative) intrinsic property following my definition. Although each individual depends on the presence of water in the environment to survive, the water composition of an individual is not sensitive to the amount of water in a normal environment. It is not an absolute intrinsic property because, in conditions in which there is little water or no water, a plant's water composition would be impacted by the lack of water in the environment. In a population in which there are dramatic variations in water availability, water composition would be considered as an extrinsic property.

The distinction between (relative) intrinsic and extrinsic properties permits us to define in general terms (i.e., not gene-centred) a model in which intrinsic properties influencing a phenotype are those properties that should be associated with heritability. In contrast, extrinsic properties should be excluded from the definition of heritability. Of course, the genotype of an individual is one of its intrinsic properties.

From the explicit model presented in Equations (7) and (8), we can input these expressions in the regression definition of heritability (Equation (3)). Assuming I and $\mathcal{E}$ are independent, we have:

$$h_b^2 = \frac{\mathrm{Cov}\,(I' + \mathcal{E}', I + \mathcal{E})}{\mathrm{Var}\,(P)}. \tag{9}$$

[9] Note that Taylor (2012) provides a 'gene-free' definition of heritability using the classical tools of quantitative genetics. In doing so, Taylor demonstrates some abuse of language when the terms 'genetic' and 'environmental' variance are used in these disciplines, while no genes need to be postulated to derive the equations.

[10] In that sense, I depart from some accounts of intrinsic properties in metaphysics, in which intrinsicality is independent of supervenience and causal dependence (see Marshall, 2016).

[11] Northcott and Piccinini (2018), in the context of the debate over acquired versus innate characteristics, use the notion of *intrinsic factor at the time of origin* to characterize an innate factor. By 'intrinsic', they mean '*inside* the organism.' Although the framework they derive from it partially overlaps with the notion of relative intrinsic property proposed here, it bears some differences.

Once developed, we have:

$$h_b^2 = \frac{\mathrm{Cov}\,(I', I) + \mathrm{Cov}\,(I', \mathcal{E}) + \mathrm{Cov}\,(\mathcal{E}', I) + \mathrm{Cov}\,(\mathcal{E}', \mathcal{E})}{\mathrm{Var}\,(P)}. \quad (10)$$

Following the reasoning above on the intrinsic/extrinsic distinction, only intrinsic properties can be associated with heritability so that we can safely eliminate the fourth term of the numerator since it concerns only extrinsic properties. Further, since intrinsic and extrinsic properties are assumed to be independent, this means that the two covariance terms $\mathrm{Cov}\,(I', \mathcal{E})$, and $\mathrm{Cov}\,(\mathcal{E}', I)$ are nil. Finally, the covariance between the average intrinsic character of the offspring and the intrinsic character of the parent can be rewritten, following the least square theory (see Lynch & Walsh, 1998), as the regression coefficient of average offspring intrinsic character on parental intrinsic character times the variance in parental intrinsic character, so that $\mathrm{Cov}\,(I', I) = \beta_{I'I}\mathrm{Var}\,(I)$.

From there, we can define the heritability of a trait as:

$$h^2 = \beta_{I'I}\frac{\mathrm{Var}\,(I)}{\mathrm{Var}\,(P)}. \quad (11)$$

The term $\beta_{I'I}$ measures the extent to which the intrinsic properties of the entities of a population are passed on to the next generation. Note that this term can be further decomposed in a number of factors such as DNA, epigenetic marks, and so forth. If all factors are passed on perfectly, as is the case (or nearly so) with genetic material in the case of asexual organisms, then $\beta_{I'I} = 1$. What is true of genes in modern organisms might, however, not have been true in the deep past or for other inheritable intrinsic factors, such as epigenetic factors. Thus, one can expect that, generally, $\beta_{I'I} \neq 1$

Equation (11) satisfies the two desiderata of general applicability and separability of the causes of resemblance outlined at the beginning of this section. It is not overly restrictive like the classical variance approach that precludes cultural traits and traits that depend on nongenetic intrinsic factors from having a non-nil heritability. Yet, it relies on an explicit model for separating environmental from nonenvironmental causes of phenotype based on the intrinsic/extrinsic property distinction. Recall that the lack of an explicit model is the main lacuna of the regression approach.

Returning to the example presented in Fig. 1, what is the heritability of height in this example following the definition of heritability presented in Equation (11)? Recall that, following the variance approach, heritability is nil in this example because there is no genetic variation, but that it is positive following the regression approach because a correlation between parental and offspring environmental richness exists. This covariance (because it refers to extrinsic properties of individuals at both generations) is not part of the definition in Equation (11); thus, heritability should be considered nil in this example.

Despite Equation (11) representing a general way to approach the concept of heritability, it relies on a strong assumption—namely, that intrinsic and extrinsic factors are independent. In other words, it relies on the assumption that there is no causal influence between intrinsic and extrinsic properties of the objects studied. What if this assumption is relaxed? In the next section, I show that this creates some difficulties for computing heritability and propose a solution to it.

5. Dependence between intrinsic and extrinsic properties

The definition of heritability proposed in Equation (11) assumes a model in which I and $\mathcal{E}$ are independent. When we relax this assumption (i.e., when we allow for them to covary), a potential problem emerges for this equation to represent an adequate definition of heritability. The problem is that we cannot assume anymore that the terms $\mathrm{Cov}\,(I, \mathcal{E}')$ and $\mathrm{Cov}\,(\mathcal{E}, \mathcal{E}')$ are nil in Equation (11). Should we include or exclude these covariance terms in our general definition of heritability?

Before answering this question, let us first talk about the term $\mathrm{Cov}\,(\mathcal{E}, I')$, which we will consider nil even when I and $\mathcal{E}$ are not independent. Recall that, following the intrinsic/extrinsic distinction proposed in the previous section, properties that vary as the result of environmental change are never considered intrinsic. They are, by definition, extrinsic. Thus, any causal influence from $\mathcal{E}$ to I' is, by definition, nil,[12] assuming that intrinsic properties are defined such that they do not vary under changes in the extrinsic properties of the parental entities. This means that any non-nil $\mathrm{Cov}\,(\mathcal{E}, I')$ term would have to be spurious (i.e., purely coincidental or due to other factors in the background). However, the intrinsic properties of the parent can have a causal influence on extrinsic properties of their offspring. This influence could manifest in both $\mathrm{Cov}\,(I, \mathcal{E}')$ and $\mathrm{Cov}\,(\mathcal{E}, \mathcal{E}')$. To assess whether these terms should be included in the heritability definition, we should know the causal origin of the covariance. Following our distinction between intrinsic and extrinsic properties, any covariance between variables at the two generations involving an extrinsic property with a causal origin in I should be included in the heritability definition. In contrast, any covariance with a causal origin in $\mathcal{E}$ should be excluded. There are at least three ways by which intrinsic properties can lead to a non-nil inter-generational covariance involving extrinsic properties (see the causal graph in Fig. 2 inspired by path analysis[13] first proposed by Wright, 1921): 1) some parental intrinsic properties influence some offspring extrinsic properties that, in turn, influence the offspring phenotype (red pathway in Fig. 2); 2) some parental extrinsic properties influenced by some parental intrinsic properties influence some offspring extrinsic properties, which, in turn, influence the offspring phenotype (blue pathway in Fig. 2); and 3) some offspring extrinsic properties influenced by offspring intrinsic properties influence the offspring phenotype in the same way as the parental phenotype is influenced by extrinsic properties with a parental intrinsic influence (green dashed double-headed arrow in Fig. 2).[14]

Based on these considerations, we can now define a parental phenotype P as the outcome of three causes:

$$P = I + \mathcal{E}_i + \mathcal{E}_e, \quad (12)$$

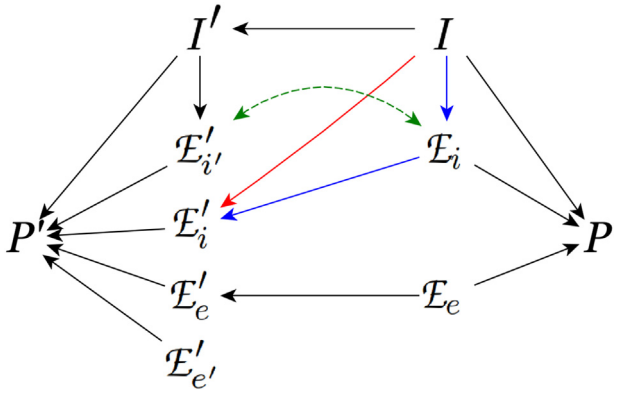

Fig. 2. Causal graph representing three pathways involving a dependence between intrinsic and extrinsic factors between two generations. Each of these dependencies potentially leads to a non-nil inter-generational parent-offspring covariance. Full arrows represent causal relationships (including transmission of factors, such as genetic transmission). The green dashed arrow represents the correlation between extrinsic factors at the parental generation influenced by intrinsic parental factors and extrinsic properties at the offspring generation influenced by intrinsic offspring factors. See main text for explanation.

[12] The causal influence from I' to $\mathcal{E}$ is also nil since causes always precede their effects.

[13] This causal graph is inspired by the causal graphs presented by Otsuka (2016), who also proposes an introduction to path analysis.

[14] Strictly speaking, this arrow would not be part of a causal graph; it is simply in the figure to illustrate the correlation.

where $\mathcal{E}_i$ and $\mathcal{E}_e$ are extrinsic properties that have an intrinsic and extrinsic causal origin, respectively.[15]

Mutatis mutandis, we define the average offspring phenotype as:

$$P' = I' + \mathcal{E}_i' + \mathcal{E}_e' + \mathcal{E}_{i'}' + \mathcal{E}_{e'}', \tag{13}$$

where $\mathcal{E}_{i'}'$ and $\mathcal{E}_{e'}'$ are the offspring extrinsic properties that have a *parental* intrinsic and extrinsic causal origin, respectively.

Implementing these terms in (10), we have:

$$h_b^2 = \frac{\mathrm{Cov}\left(I' + \mathcal{E}_i' + \mathcal{E}_e' + \mathcal{E}_{i'}' + \mathcal{E}_{e'}', I + \mathcal{E}_i + \mathcal{E}_e\right)}{\mathrm{Var}(P)}. \tag{14}$$

From there, we assume that I and its direct influence $\mathcal{E}_i$ are independent of $\mathcal{E}_e'$ and $\mathcal{E}_{e'}'$ since, by definition, the latter terms are influenced by extrinsic properties. We also assume that I is independent of $\mathcal{E}_{i'}'$, since the latter is, by definition, only influenced by offspring intrinsic properties. Further, following the definition of an intrinsic property, we assume that $\mathcal{E}_i$ is independent of I' and of its direct influence $\mathcal{E}_{i'}'$ because, as noted above, the environment (extrinsic to the organism) cannot influence an intrinsic property (and, consequently, its effects). Finally, we discount any covariance term involving an extrinsic causal origin (i.e., $\mathcal{E}_e$, $\mathcal{E}_e'$, and $\mathcal{E}_{e'}'$). This is so to satisfy the desideratum of separability of causes of resemblance, which is not satisfied by the regression approach. With these assumptions, once Equation (14) is developed, we obtain a general definition of heritability as:

$$h^2 = \frac{\mathrm{Cov}(I', I) + \mathrm{Cov}\left(\mathcal{E}_i', I\right) + \mathrm{Cov}\left(\mathcal{E}_i', \mathcal{E}_i\right) + \mathrm{Cov}\left(\mathcal{E}_{i'}', \mathcal{E}_i\right)}{\mathrm{Var}(P)}. \tag{15}$$

Transforming covariances into regressions, we get:

$$h^2 = \frac{\beta_{I'I}\mathrm{Var}(I) + \beta_{\mathcal{E}_i'I}\mathrm{Var}(I) + \beta_{\mathcal{E}_i'\mathcal{E}}\mathrm{Var}(\mathcal{E}_i) + \beta_{\mathcal{E}_{i'}'\mathcal{E}_i}\mathrm{Var}(\mathcal{E}_i)}{\mathrm{Var}(P)}. \tag{16}$$

The first term on the right-hand side of this equation represents the part of the parent-offspring regression for the phenotype due to the parental intrinsic properties being transmitted to offspring such as genes. It captures the black arrow from I to I' in Fig. 2. The second term on the right-hand side represents the part of the parent-offspring regression due to the direct influence of parental intrinsic properties on offspring extrinsic properties. It captures the red path in Fig. 2. The third term on the right-hand side represents the part of the parent-offspring regression due to the indirect effect of parental extrinsic properties under the influence of intrinsic properties on offspring extrinsic properties. It captures the blue path in Fig. 2. Finally, the fourth term on the right-hand side represents the part of the parent-offspring regression due to the extrinsic properties under intrinsic influence reoccurring between the two generations due to the transmission of intrinsic properties. It captures the green double-headed dashed arrow in Fig. 2.

To illustrate a causal influence $I \rightarrow \mathcal{E}_i'$ (red pathway in Fig. 2) that would manifest as $\mathrm{Cov}\left(I, \mathcal{E}'\right) \neq 0$, let us take the example presented in Fig. 1 and modify it slightly. We could imagine now that, when a plant develops, it changes the soil composition, for instance, because its leaves fall and are then degraded. In turn, this changes the richness of the environment, but only after one generation since it takes time to degrade the leaves.[16] We could further imagine that there is variation between different plants in the way they modify the offspring environment due to some intrinsic factors (including genetic ones). For instance, some leaves could be richer in the amount of an essential nutrient for plant growth, or contain a toxin that eliminates competition. In this case, a parent-offspring resemblance due to this difference should be considered as heritable. Why? Because the causal origin of the covariance is intrinsic.[17]

To illustrate the covariance due to the causal influence of intrinsic properties on the environment at both the parental and the offspring generation, $I \rightarrow \mathcal{E}$ and $I' \rightarrow \mathcal{E}'$, respectively (green dashed double-headed arrows representing the covariance in Fig. 2), that would manifest as $\mathrm{Cov}(\mathcal{E}, \mathcal{E}') \neq 0$, we can take again the example presented in Fig. 1. However, now imagine that the leaves degrade at a much faster rate so that a plant can benefit from some of the nutrients before reproducing and, thus, have a different phenotype from the one it would have had, had the nutrients coming from leaf degradation not been taken into consideration. Assuming that the offspring of the plant are endowed with the same intrinsic material, they would also produce the same change in soil composition and phenotypic effects. Had only the differences in phenotype made by the intrinsic properties directly on phenotype been taken into consideration at both the parental and offspring generation, the heritability value obtained would have been lower.

Finally, to illustrate a causal influence $I \rightarrow \mathcal{E}_i \rightarrow \mathcal{E}_i'$ (blue pathway in Fig. 2) that would also manifest as $\mathrm{Cov}(\mathcal{E}, \mathcal{E}') \neq 0$, we could imagine (as previously) that the leaves degrade at a much faster rate and lead to a change in the plant's height. However, another effect would also be that the leaves attract earthworms, which further change the properties of the soil only after one generation, leading to a positive parent-offspring environment covariance.

These thought experiments are consistent with recent findings. Some plant ecologists have indeed found that the litterfall of some species of trees can modify the soil composition in a way that leads to favourable conditions for the trees of that species (Bigelow & Canham, 2015; Olson, 2019). It has also been argued that this type of interaction could be under selection pressure (Schweitzer et al., 2018).

These examples connect with the literature on niche construction (Odling-Smee et al., 2003), developmental systems theory (Oyama et al., 2003), and the extended evolutionary synthesis (Laland et al., 2014, see also; Mameli, 2004). Two aims of these approaches are i) to move away from a purely gene-centred approach to evolutionary theory, and ii) to take into account the multigenerational interactions between an organism and its environment. My definition of heritability in terms of intrinsic and extrinsic properties is consistent with i. The multigenerational covariance effects on phenotype of intrinsic properties by environmental mediation is consistent with ii.

I should also mention that my definition is consistent with the framework on indirect genetic effects (IGEs) in quantitative genetics. This framework permits us to account for heritability components by considering that some of the causes leading to phenotypic variation can be indirect: that is, due to the interactions of individuals with their environment (potentially across multiple generations).[18] These components of heritability are classically not accounted for when traditional estimates of heritability are used (See Walsh and Lynch 2018, chap. 22. for a review of the literature on IGEs, see also Bijma 2011) Using the intrinsic/extrinsic distinction, some intrinsic effects are indirect. In situations of covariance between intrinsic and extrinsic factors, the indirect effects of intrinsic factors should be accounted for as, I have argued.

In the next section, I show that, surprisingly, my approach can also be relevant to another area: namely, microbiome research.

[15] Note that this decomposition neglects the possible extrinsic influence originating from intrinsic factors of generations anterior to the parental generation. I will consider that they are negligible here.

[16] Of course, the delay could be longer or shorter, as we shall see.

[17] Similar considerations are developed in Lynch and Bourrat (2017) in the context of the variance approach to heritability and gene-environment correlations.

[18] That the individuals of a population interact does not imply that this would translate as a statistical interaction. For more details on the distinction between these two notions of interaction, see Tabery (2014).

6. Heritability & the microbiome

In recent years, whether the gut microbiome (including of humans) should be regarded as part of the environment of a multicellular organism or as constituting, together with a multicellular organism, an individual that can be regarded as a unit of selection (a holobiont) has been debated in both the biological and philosophical literature (see, for instance Skillings, 2016; Bordenstein & Theis, 2015; Rosenberg & Zilber-Rosenberg, 2016; Moran & Sloan, 2015; Douglas & Werren, 2016; Gilbert et al., 2012; Roughgarden et al., 2017; Lloyd & Wade, 2019). I will not participate here in the debate, although my views on this matter can be found in Bourrat and Griffiths (2018). Rather, what interests me here is whether the potential effects of the microbiome on complex traits of the host should be regarded as heritable (for a review on the potential influence of microbiome composition on numerous human-health traits, see Gilbert et al., 2016).

Before proceeding, it should be noted that microbiome research is a young discipline. This means that many of its findings will likely be challenged one way or another in the coming years as new and more robust data is collected.

The question of the heritability of microbial composition effects and related questions have been asked by several authors. Some consider that the effects of the microbiome on the host's traits should be included in heritability measures (e.g., Sandoval-Motta et al., 2017; van Opstal & Bordenstein, 2015). Others disagree (e.g., Douglas et al., 2020), arguing that a substantial part of microbiome composition is driven by environmental factors (for a recent study, see Rothschild et al., 2018). Yet, others consider that microbiome composition is itself a complex human trait (e.g., Goodrich et al., 2017) partly under the influence of the host genotype. Faced with these conflicting views about the role of microbiome composition on host traits, how should we conceive of microbiome composition associated with phenotypic variation in light of the foregoing analysis?

First, when asking whether microbiome composition is heritable, one needs to know whether the microbiome is considered as intrinsic or extrinsic to the host. To this question, there is, in general, no right or wrong answer for, recall, this distinction is ultimately a *relative* one within the context of biology. This choice must be made by the researchers involved in this field on an individual basis grounded in some independent considerations. For some species, we have good reasons to lump together a multicellular organism and its microbiome into a single entity or individual—namely, when the two entities represent a functional unit and the association is obligatory and specific. With other species, it is much less clear what the benefits would be in doing so (see Bourrat & Griffiths, 2018).

In any case, a lack of agreement between researchers on whether the microbiome should be considered as an intrinsic part of the entities forming a population in a particular case will render *impossible* a comparison of the heritability of different traits obtained from these researchers. This is so because the measure obtained would not abide by the same conventions for what is considered as intrinsic and extrinsic. This point, I believe, has previously been underappreciated. Equation (15) and the relevance of the distinction between intrinsic and extrinsic factors in a general definition of heritability permits us to acknowledge this point better.

If the microbiome is considered intrinsic (and is passed on across the generations of the multicellular organism), host phenotypic differences made by the microbiome ought to be included in the heritability value. If, on the contrary, the microbiome is considered extrinsic to the host, there are two sets of possible situations to consider. In the first set, microbiome composition is correlated with and *caused* by some intrinsic properties of the host. In such cases, even though the microbiome is extrinsic to the individuals of the population, the difference it makes to the phenotype should be included in the heritability value via the terms $\beta_{\mathcal{E}'_i\mathcal{E}_i}\mathrm{Var}(\mathcal{E}_i)$ and $\beta_{\mathcal{E}'_i\mathcal{E}_i}\mathrm{Var}(\mathcal{E}_i)$ of Equation (16). In the second type of situation, the covariance between extrinsic properties across generations does not have an intrinsic origin. In such situations, the differences made should not be considered as heritable.

The analysis provided here highlights the point that, without a clear causal model, a heritability value for a trait in which microbiome composition is involved will generally be meaningless. The model based on the extrinsic/intrinsic distinction proposed here could represent a general framework that permits different researchers to use a common set of conventions and clearly assess where any of their disagreement(s) lies.

Further, having a clear causal model could be beneficial for interventions, including in humans. If microbiome composition is mostly driven by environmental factors (e.g., diet) that are independent of any intrinsic factors, it will likely be more effective to intervene on these factors in the long term than if microbial composition depends on the host's intrinsic properties (e.g., its genome).

7. Conclusion

The main goal of this paper was to propose a definition of heritability that satisfies two desiderata for it to be consistent within the field of evolutionary theory. The first desideratum is that the notion of heritability should be applicable beyond classical biological situations. The second is that it should rely on a distinction between environmental causes and causes that are attributed to the entities bearer of the phenotype. After having presented the two classical approaches (variance and regression) to heritability used in evolutionary theory, I showed that each can only meet one of the two desiderata. I then proposed a general model based on a pragmatic distinction between intrinsic and extrinsic properties that satisfies both desiderata. From there, I discussed the difficulties associated with cases involving a correlation between intrinsic and extrinsic properties and proposed a way to deal with these cases. Finally, I showed the relevance of my analysis and definition in a concrete biological context, namely, the heritability of complex traits influenced by the microbiome composition.

Acknowledgments

I thank two anonymous reviewers and the editors of this special issue —in particular, Davide Serpico and Kate Lynch—for their constructive feedback. Finally, I thank Marianne McAllister who proofread the manuscript.

References

Bigelow, S. W., & Canham, C. D. (2015). Litterfall as a niche construction process in a northern hardwood forest. *Ecosphere, 6*(7), art117. https://doi.org/10.1890/ES14-00442.1

Bijma, P. (2011). A general definition of the heritable variation that determines the potential of a population to respond to selection. *Genetics, 189*(4), 1347–1359. https://doi.org/10.1534/genetics.111.130617

Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biology, 13*(8), Article e1002226. https://doi.org/10.1371/journal.pbio.1002226

Bourrat, P. (2014). From survivors to replicators: Evolution by natural selection revisited. *Biology and Philosophy, 29*(4), 517–538.

Bourrat, P. (2015a). Distinguishing natural selection from other evolutionary processes in the evolution of altruism. *Biological Theory, 10*(4), 311–321.

Bourrat, P. (2015b). How to read 'heritability' in the recipe approach to natural selection. *The British Journal for the Philosophy of Science, 66*(4), 883–903.

Bourrat, P. (2017). Explaining drift from a deterministic setting. *Biological Theory, 12*(1), 27–38.

Bourrat, P. (2019a). In what sense can there be evolution by natural selection without perfect inheritance? *International Studies in the Philosophy of Science, 32*(1), 39–77.

Bourrat, P. (2019b). Variation of information as a measure of one-to-one causal specificity. *European Journal for Philosophy of Science, 9*(1), 11. https://doi.org/10.1007/s13194-018-0224-6

Bourrat, P. (2020). Causation and single nucleotide polymorphism heritability. *Philosophy of Science, 87*(5), 1073–1083. https://doi.org/10.1086/710517

Bourrat, P. (2021). Heritability, causal influence and locality. *Synthese, 198*(7), 6689–6715. https://doi.org/10.1007/s11229-019-02484-3

Bourrat, P., & Griffiths, P. E. (2018). Multispecies individuals. *History & Philosophy of the Life Sciences, 40*(2). https://doi.org/10.1007/s40656-018-0194-1

Bourrat, P., & Lu, Q. (2017). Dissolving the missing heritability problem. *Philosophy of Science, 84*(5), 1055–1067. https://doi.org/10.1086/694007

Bourrat, P., Lu, Q., & Jablonka, E. (2017). Why the missing heritability might not be in the DNA. *BioEssays, 39*(7).

Brandon, R. N. (1990). *Adaptation and environment*. Princeton, NJ: Princeton University Press.

Douglas, G. M., Bielawski, J. P., & Langille, M. G. I. (2020). Re-evaluating the relationship between missing heritability and the microbiome. *Microbiome, 8*(1), 87. https://doi.org/10.1186/s40168-020-00839-4

Douglas, A. E., & Werren, J. H. (2016). Holes in the hologenome: Why host-microbe symbioses are not holobionts. *mBio, 7*(2), e02.99–115. https://doi.org/10.1128/mBio.02099-15

Downes, S. M., & Matthews, L. (2020). Heritability. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy, spring 2020 edn*. Metaphysics Research Lab, Stanford University.

Falconer, D. S., & Mackay, T. F. (1996). *Introduction to quantitative genetics* (4th ed.). Essex: Longman.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*(2), 399–433.

Fox Keller, E. (2010). *The mirage of a space between nature and nurture*. Durham, NC: Duke University Press.

Gilbert, J. A., Quinn, R. A., Debelius, J., Xu, Z. Z., Morton, J., Garg, N., Jansson, J. K., Dorrestein, P. C., & Knight, R. (2016). Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature, 535*(7610), 94–103. https://doi.org/10.1038/nature18850

Gilbert, S. F., Sapp, J., & Tauber, A. I. (2012). A symbiotic view of life: We have never been individuals. *The Quarterly Review of Biology, 87*(4), 325–341. https://doi.org/10.1086/668166

Godfrey-Smith, P. (2007). Conditions for evolution by natural selection. *Journal of Philosophy, 104*, 489.

Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford ; New York: Oxford University Press.

Goodrich, J. K., Davenport, E. R., Clark, A. G., & Ley, R. E. (2017). The relationship between the human genome and microbiome comes into view. *Annual Review of Genetics, 51*(1), 413–433. https://doi.org/10.1146/annurev-genet-110711-155532

Griffiths, P. E., & Neumann-Held, E. M. (1999). The many faces of the gene. *BioScience, 49*, 656–662.

Griffiths, P. E., Pocheville, A., Calcott, B., Stotz, K., Kim, H., & Knight, R. (2015). Measuring causal specificity. *Philosophy of Science, 82*(4), 529–555. https://doi.org/10.1086/682914

Griffiths, P. E., & Stotz, K. (2006). Genes in the postgenomic era. *Theoretical Medicine and Bioethics, 27*(6), 499. https://doi.org/10.1007/s11017-006-9020-y

Griffiths, P. E., & Stotz, K. (2013). *Genetics and philosophy: An introduction*. New York: Cambridge University Press.

Jacquard, A. (1983). Heritability: One word, three concepts. *Biometrics, 39*, 465–477.

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). *Behavioral genetics* (7th ed. edn.). New York: Worth Publishers.

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G. A., Hoekstra, H. E., Futuyma, D. J., Lenski, R. E., Mackay, T. F. C., Schluter, D., & Strassmann, J. E. (2014). Does evolutionary theory need a rethink? *Nature News, 514*(7521), 161. https://doi.org/10.1038/514161a

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics, 1*(1), 1–18.

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *The American Journal of Human Genetics, 26*, 400.

Lloyd, E. A., & Wade, M. J. (2019). Criteria for holobionts from community genetics. *Biological Theory, 14*(3), 151–170. https://doi.org/10.1007/s13752-019-00322-w

Lu, Q., & Bourrat, P. (2018). The evolutionary gene and the extended evolutionary synthesis. *The British Journal for the Philosophy of Science, 69*(3), 775–800. https://doi.org/10.1093/bjps/axw035

Lynch, K. E. (2021). The meaning of "cause" in genetics. *Cold Spring Harbor Perspectives in Medicine*, a040519. https://doi.org/10.1101/cshperspect.a040519

Lynch, K. E., & Bourrat, P. (2017). Interpreting heritability causally. *Philosophy of Science, 84*(1), 14–34. https://doi.org/10.1086/688933

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits, 1*. Sinauer Sunderland, MA.

Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature News, 456*(7218), 18–21.

Mameli, M. (2004). Nongenetic selection and nongenetic inheritance. *The British Journal for the Philosophy of Science, 55*, 35–71.

Marshall, D. (2016). The varieties of intrinsicality. *Philosophy and Phenomenological Research, 92*(2), 237–263. https://doi.org/10.1111/phpr.12156

Matthews, L. J., & Turkheimer, E. (2021). Across the great divide: Pluralism and the hunt for missing heritability. *Synthese, 198*, 2297–2311.

Moran, N. A., & Sloan, D. B. (2015). The hologenome concept: Helpful or hollow? *PLoS Biology, 13*(12), Article e1002311. https://doi.org/10.1371/journal.pbio.1002311

Moss, L. (2003). *What genes can't do* (new ed edn.). Cambridge, Mass: The MIT Press.

Northcott, R., & Piccinini, G. (2018). Conceived this way: Innateness defended. *Philosophers' Imprint, 18*(18), 16.

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution. Monographs in population biology, No. 37*. Princeton: University Press.

Oftedal, G. (2005). Heritability and genetic causation. *Philosophy of Science, 72*(5), 699–709. https://doi.org/10.1086/508126

Okasha, S. (2006). *Evolution and the levels of selection*. Oxford; New York: Clarendon Press; Oxford University Press.

Okasha, S. (2010). Replies to my critics. *Biology and Philosophy, 25*(3), 425–431. https://doi.org/10.1007/s10539-009-9158-x

Olson, M. E. (2019). Plant evolutionary ecology in the age of the extended evolutionary synthesis. *Integrative and Comparative Biology, 59*(3), 493–502. https://doi.org/10.1093/icb/icz042

van Opstal, E. J., & Bordenstein, S. R. (2015). Rethinking heritability of the microbiome. *Science, 349*(6253), 1172–1173. https://doi.org/10.1126/science.aab3958

Otsuka, J. (2016). Causal foundations of evolutionary genetics. *The British Journal for the Philosophy of Science, 67*(1), 247–269. https://doi.org/10.1093/bjps/axu039

Oyama, S., Griffiths, P. E., & Gray, R. D. (2003). *Cycles of contingency: Developmental systems and evolution*. Mit Press.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York: Cambridge University Press.

Pocheville, A., Griffiths, P. E., & Stotz, K. (2017). Comparing causes – an information-theoretic approach to specificity, proportionality and stability. In H. Leitgeb, I. Niiniluoto, E. Sober, & P. Seppälä (Eds.), *Proceedings of the 15th congress of logic, methodology and philosophy of science* (pp. 260–275). London: College Publications.

Price, G. R. (1970). Selection and covariance. *Nature, 227*, 520–521.

Rice, S. H. (2004). *Evolutionary theory: Mathematical and conceptual foundations*. Sunderland, Mass., USA: Sinauer Associates.

Rosenberg, E., & Zilber-Rosenberg, I. (2016). Microbes drive evolution of animals and plants: The hologenome concept. *mBio, 7*(2). https://doi.org/10.1128/mBio.01395-15

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., … Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature, 555*(7695), 210–215. https://doi.org/10.1038/nature25973

Roughgarden, J., Gilbert, S. F., Rosenberg, E., Zilber-Rosenberg, I., & Lloyd, E. A. (2017). Holobionts as units of selection and a model of their population dynamics and evolution. *Biological Theory*, 1–22. https://doi.org/10.1007/s13752-017-0287-1

Sandoval-Motta, S., Aldana, M., Martínez-Romero, E., & Frank, A. (2017). The human microbiome and the missing heritability problem. *Frontiers in Genetics, 8*. https://doi.org/10.3389/fgene.2017.00080

Schweitzer, J. A., Van Nuland, M., & Bailey, J. K. (2018). Intraspecific plant–soil feedbacks link ecosystem ecology and evolutionary biology. In T. Ohgushi, S. Wurst, & S. N. Johnson (Eds.), *Aboveground–belowground community ecology, ecological studies* (pp. 69–84). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91614-9_4.

Sesardic, N. (1993). Heritability and causality. *Philosophy of Science, 60*(3), 396–418.

Sesardic, N. (2005). *Making sense of heritability*. Cambridge: Cambridge University Press.

Skillings, D. (2016). Holobionts and the ecology of organisms: Multi-species communities or integrated individuals? *Biology and Philosophy, 31*(6), 875–892. https://doi.org/10.1007/s10539-016-9544-0

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search* (Vol. 81). MIT press.

Sterelny, K., & Griffiths, P. E. (1999). *Sex and death: An introduction to philosophy of biology*. Chicago: University of Chicago press.

Tabery, J. (2014). *Beyond versus: The struggle to define the interaction of nature and nurture*. Cambridge, MA; London: MIT Press.

Tal, O. (2009). From heritability to probability. *Biology and Philosophy, 24*(1), 81–105. https://doi.org/10.1007/s10539-008-9129-7

Tal, O. (2012). The impact of gene–environment interaction and correlation on the interpretation of heritability. *Acta Biotheoretica, 60*(3), 225–237. https://doi.org/10.1007/s10441-011-9139-8

Taylor, P. (2006). Commentary: The analysis of variance is an analysis of causes (of a very circumscribed kind). *International Journal of Epidemiology, 35*, 527–531.

Taylor, P. (2010). Three puzzles and eight gaps: What heritability studies and critical commentaries have not paid enough attention to. *Biology and Philosophy, 25*(1), 1–31. https://doi.org/10.1007/s10539-009-9174-x

Taylor, P. J. (2012). A gene-free formulation of classical quantitative genetics used to examine results and interpretations under three standard assumptions. *Acta Biotheoretica, 60*(4), 357–378. https://doi.org/10.1007/s10441-012-9164-2

Turkheimer, E. (2011). Still missing. *Research in Human Development, 8*(3–4), 227–241.

Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*. New York: Oxford University Press.

Waters, C. K. (2007). Causes that make a difference. *The Journal of Philosophy, 104*(11), 551–579.

Weatherson, B., & Marshall, D. (2017). Intrinsic vs. Extrinsic properties. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy, fall 2017 edn, metaphysics research lab*. Stanford University.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy, 25*(3), 287–318.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*(7), 557–585.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics, 42*(7), 565–569.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences, 109*(4), 1193–1198. https://doi.org/10.1073/pnas.1119675109