Studies in History and Philosophy of Science 94 (2022) 87–98

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

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

On the causal interpretation of heritability from a structural causal modeling perspective

Qiaoying Lu [a,b,*], Pierrick Bourrat [c,d]

[a] Institute of Foreign Philosophy, Peking University, PR China
[b] Department of Philosophy, Peking University, No.5 Yiheyuan Road, Haidian District, Beijing, 100871, PR China
[c] Department of Philosophy, Macquarie University, North Ryde, NSW, 2109, Australia
[d] Department of Philosophy and Charles Perkins Centre, The University of Sydney, Camperdown, NSW, 2006, Australia

ARTICLE INFO

Keywords:
Heritability
Causal interpretation
Gene–environment interaction
Gene–environment covariation
Structural causal modeling
Interventionist account of causation

ABSTRACT

Heritability estimated using the analysis of variance (ANOVA) for ascribing causal responsibility to genes for a phenotype has been criticized widely. First, there are problems associated with articulating the exact causal meaning of heritability in the standard model. Second, in conditions of gene–environment interaction or covariation that violate the assumptions made by the standard model, a causal interpretation of heritability is thought to be unwarranted. This paper aims to rethink these ideas and associated disputes from a structural causal modeling (SCM) perspective. Using SCM, we show that, in the standard model, heritability reflects the causal effect of eliminating genotypic differences on the change of phenotypic variance of a population. In the presence of interaction or covariation, heritability is estimated incorrectly using ANOVA. However, SCM can provide the causal effect of genotypes on the phenotypic variance regarding particular interventions. We also show that SCM can identify different types of causal effect and answer individual-level causal questions. We conclude that SCM has the resources to provide a systematic causal interpretation that can supplement traditional heritability estimates via ANOVA and offer a more substantial causal analysis of genetic causation.

1. Introduction

Heritability estimated using the analysis of variance (ANOVA) to measure the causal contribution of genes to traits has been criticized widely. The basic function of the method is to partition the phenotypic variance of a population into two parts: one due to genotypic difference and the other due to environmental difference. Heritability is defined as the ratio of the genotypic component to the total variance. Classically, it is interpreted as the "relative importance" of genotypic causes to the phenotype of the population (Falconer & Mackay, 1996, pp. 122–123). One famous criticism of heritability analysis was formulated by Richard Lewontin in the 1970s. Following Arthur Jensen's inference that from high heritability estimates for IQ, one can assert that intellectual ability is determined predominantly by genes (Jensen 1969, p. 7). Lewontin (2006 [1974]) highlighted three significant limitations of this method: namely, the problems of interaction, locality, and tautology.

The current consensus among philosophers of biology is that heritability analysis has minimal causal implications and is misleading when used to provide causal explanations for individual development (Downes & Matthews, 2020). Nevertheless, several authors have defended the method's explanatory usefulness (Bourrat, 2021b; Oftedal, 2005; Pearson, 2007; Sesardic, 2005; Tal, 2009). Given the lengthy discussions on the causal interpretation of heritability, it is prudent to examine the concept in precise causal terms. For instance, Tal (2012, pp. 234–235) discusses different causal consequences in the presence of gene–environment interaction and covariation in applying probability to heritability analysis. Lynch and Bourrat (2017) investigate heritability estimates in covariation cases by distinguishing causal origins. Bourrat (2021b) interprets heritability in terms of James Woodward's notion of range of causal influence based on information theory. However, to date, there has been no explicit discussion of heritability using systematic formal tools to study causality.

In recent years, causal modeling has reshaped our understanding of causality and the relationship between statistics and causation. Causal models are mathematical objects representing causal relationships that (1) facilitate causal discovery from statistical data, and (2) permit one to infer specific causal claims based on those models. The concept originates in the work of Sewall Wright (1921) and has been addressed by

* Corresponding author: Department of Philosophy, Peking University, No.5 Yiheyuan Road, Haidian District, Beijing, 100871, PR China.
E-mail addresses: qy.lu@pku.edu.cn (Q. Lu), p.bourrat@gmail.com (P. Bourrat).

https://doi.org/10.1016/j.shpsa.2022.05.005

influential contemporary works such as Spirtes et al. (2000) and Pearl (2009). The first thorough application of causal modeling in biology can be found in Shipley (2000). Although developed independently with different aims, the primary concepts of causal modeling have been entangled with Woodward's interventionist account of causation. Both causal modeling and Woodward's interventionist account share the basic idea that causation should be defined in terms of *intervention* because we acquire causal knowledge by actually or hypothetically intervening in the world and then use this knowledge for manipulation and control (see, e.g., Pearl, 2009, pp. 361–362; Woodward, 2003, Chapter 2.1).[1] This paper investigates the disputes surrounding ANOVA heritability estimates using Pearl's structural causal modeling (SCM). It is fair to note that Woodward's and Pearl's frameworks have influenced one another (Pearl, 2003, 2009, p. 239; Woodward, 2003, pp. 38–39).

Crucially, we apply SCM to analyze the causal implications of heritability instead of obtaining an estimation, which is typically the aim of statistical approaches such as ANOVA or potential outcomes (see, e.g., Rubin, 2005). When facing a query about a certain type of causal effect,[2] the answer provided by SCM is essentially indirect: "if you tell me how the world works (by giving me the full causal graph), I can tell you the answer" (Imbens, 2020, p. 1132). In other words, SCM is suitable for investigating theoretical models with possible causal queries about identification. Given a particular causal story, SCM can tell us which and why specific causal quantities are identifiable; for instance, the strength of a causal effect can be identified (reduced) by a unique formula comprising only statistical terms.[3] Pearl (2019) applies SCM to the Match, Oxygen, and Fire example to identify quantities of causal sufficiency. This paper borrows Pearl's example to introduce SCM, but instead focuses on quantities of different types of causal effect. Using SCM's well-established causal semantics, we can articulate causal interpretations of those causal effects.

The paper will run as follows. Section 2 briefly introduces the standard procedure of heritability analysis in situations where there are no gene–environment interactions or covariations, which we refer to as the "standard model." We briefly present the three problems often cited in the literature—interaction (in its vernacular sense), locality, and tautology—and respond to them to delimit the possible scope of a causal interpretation of heritability. Section 3 presents two further obstacles posited for the causal interpretation of heritability: interaction (in its statistical sense) and covariation. Section 4 introduces the SCM framework and applies it to the standard model of heritability analysis. Our results broadly agree with the current consensus; however, SCM explicates the specific type of causal effect reflected by standard heritability. Section 5 illustrates why interaction and covariation render heritability estimates via ANOVA misleading and shows the causal interpretations that SCM can offer. By extending the application of SCM to an individual-level analysis, we conclude that SCM can serve as a powerful tool to identify genetic causation in more sophisticated ways than is possible via traditional ANOVA.

2. The standard model of heritability analysis

ANOVA was first introduced by R. A. Fisher (1918) and has become the primary method for classical quantitative genetics. According to the standard model, for a certain phenotype (e.g., plant height), the phenotypic variance of a population (V_P) can be decomposed into the genotypic variance (V_G) and the environmental variance (V_E). In its simplest form, we have:

$$V_P = V_G + V_E. \tag{1}$$

From there, broad sense heritability (H^2) is defined as:

$$H^2 = \frac{V_G}{V_P}. \tag{2}$$

H^2 measures how much of the total phenotypic variance is due[4] to genotypic differences. In comparison, the narrow sense heritability (h^2) measures the proportion of phenotypic variance due to alleles transmitted from parents to offspring (Falconer & Mackay, 1996, p. 123). h^2 is used primarily in breeding studies and by theorists concerned with making evolutionary projections (Downes & Matthews, 2020). Since we are concerned with genetic causation within one generation, we will use "heritability" to refer to H^2 throughout the rest of this paper. Another notion of heritability relies on parent–offspring regression rather than ANOVA (see Godfrey-Smith, 2007; Jacquard, 1983; Bourrat, 2022; Okasha, 2006). However, we will not discuss it here because of the close links between this latter notion and h^2.[5]

The standard procedure of heritability analysis can be presented using an example. Suppose a population of individual plants with various heights. Assume that the individuals only differ in their genotypes at particular loci ($g1$ and $g2$) and in their environment with different temperatures ($e1$, $e2$, and $e3$), which results in six subsets of $(g,\ e)$ combinations. This means that other factors affecting height are randomized or remain constant. Suppose that each subset has an equal number of individuals and that differences in height within each subset are too small to be significant. The average height values for the six subsets (hypothetically) observed are presented in Table 1. For instance, the value of $(g1,\ e1)$ is 5, meaning that the average height of $g1$ individuals living in $e1$ is 5 units. The heritability estimated from Table 1 represents the heritability of the original population.

Here, $\bar{g}$ represents the phenotypic value for a given genotype averaged over all environments, $\bar{e}$ the value for a given environment averaged over all genotypes, and $\bar{p}$ the grand mean. Hence, V_P is calculated by squaring every number deviation from $\bar{p}$, adding them and dividing the result by 6 (there are six subsets). To obtain V_G, we square the deviations of $\bar{g}$ from $\bar{p}$, add them, and divide by 2 (two genotypes). Similarly, we obtain V_E by squaring the deviations of $\bar{e}$ from $\bar{p}$, adding them, and dividing by 3 (three environments). The results are:

Table 1
A standard example where there is no gene–environment interaction or covariation.

	$e1$	$e2$	$e3$	$\bar{g}$
$g1$	5	10	15	10
$g2$	15	20	25	20
$\bar{e}$	10	15	20	$\bar{p} = \mathbf{15}$

[1] Gebharter (2017, Chapter 5) also argued that recent developments in causal modeling modify Woodward's original framework in such a way to clarify the notion of a *possible* intervention.

[2] In general, we take a causal effect to refer to a change in an effect variable due to an intervention on a causal variable. Since we can ask different causal questions led by particular interventions on the cause, the answers provided imply specific types of causal effect. A formal and generic notion of causal effect within SCM is introduced in Section 4, from which different types of causal effect can be derived.

[3] Note here that the causal effect can be estimated if data for these terms are provided. However, in practice, scientists often only have partial causal stories; therefore, the requirement for sufficient causal knowledge limits the scope of the applicability of SCM for this purpose.

[4] Although a causal gloss is often given to it, the term "due" here is not causal, but purely statistical.

[5] See Bourrat (2022) for an in-depth analysis of the relationship between the two notions of heritability.

Table 2
An interaction example where there is no covariation.

	$e1$	$e2$	$e3$	$\bar{g}$
$g1$	5	10	15	10
$g2$	27	20	13	20
$\bar{e}$	16	15	14	$\bar{p} = 15$

$$V_{P(S)} = 41.7,\ V_{G(S)} = 25,\ V_{E(S)} = 16.7,$$

where the subscript "S" stands for "standard." With the components of variance obtained, following Equation (2), heritability is:

$$H^2_{(S)} = \frac{V_{G(S)}}{V_{P(S)}} = 0.6.$$

This result is classically interpreted as genes causing 60% of the height variance in the population. Accordingly, the environment causes 40% of the total variance. It provides an answer about the relative importance of genes and the environment in causing phenotypic variance.

There are three general problems invoked to show the limitations of the classical causal interpretation of standard heritability. First, radical interactionists might argue that the decomposition of V_G and V_E is illegitimate because a phenotype is always the product of an interaction between genes and the environment (i.e., when "interaction" is understood in a common sense). This problem can be dispelled by noting the consensus that genes and the environment are both necessary causes for the phenotype (Sesardic, 2005, p. 59; Sterelny & Griffiths, 1999, pp. 15, 99). As Lewontin (2006 [1974], p. 520) claims, we should distinguish two questions concerning causation. The first is to regard genes and the environment as "two alternative and mutually exclusive causes" and ask which causes the phenotype (ibid). The second is to recognize that both are necessary for the phenotype and ask about the relative importance of each. Heritability analysis is supposed to answer the second question rather than the first. In our standard example, genes are a relatively more important cause than the environment.

A second problem is that a heritability estimate is fundamentally *local* since it "depends upon the *actual* distribution of genotypes and environments in the particular population sampled" (Lewontin, 2006 [1974], p. 521, emphasis added). When targeting a particular population, it is inevitable that the observed genes and environmental factors do not include all existing or potential causes. However, it is still legitimate to ask whether the actual variation in genes or environment explains the *observed* phenotypic variation in the population. This idea connects with Kenneth Waters' distinction between actual and potential difference makers (Waters, 2007). An actual difference maker is a variable with actual variation in the value that produces actual variation in the value of the effect variable, whereas a potential difference maker only satisfies the counterfactual patterns specified by the interventionist theory, regardless of whether the causal variable actually varies or causes actual differences. Therefore, it should be noted that heritability estimates amount to measuring the *actual* genotypic influence against the phenotypic variation of a particular population (see also Bourrat, 2021b).

The third problem of heritability analysis is that it is *tautological* (Lewontin, 2006 [1974], p. 521) since one can estimate heritability without knowing exact information about genes (e.g., the DNA sequence) and environment. In the standard example, we do not need to know the precise difference between $g1$ and $g2$; we only require the phenotypic difference due to the difference between them. As we see it, the oddity of tautology arises when heritability analysis is conceived incorrectly as a mechanism-elucidating approach. If we want to know the causal mechanism involved between genes and the environment for producing phenotypes, exact information about each will be required. However, as James Tabery argues (Tabery, 2014, p. 5), heritability analysis represents a different approach that searches for the causes of variation in the population (see also Lynch, 2021). As we argue in Section 5.3, SCM gives the causal effect of genes on phenotypic *variance* based on a theory about causal *mechanisms*, which inherently bridges the gap between these two approaches.

To summarize, as many authors agree, heritability estimates in the conditions of the standard model measure how much of the phenotypic *variation* is determined by the *actual* difference of genotypes, compared to the environment, in a particular population. In contrast with the general claim that genes cause phenotypes, standard heritability reflects a very restricted sense of causal influence (Taylor, 2006). Before using SCM to examine this limited type of causal effect more closely, we present two obstacles to interpreting heritability estimates causally in the following section.

3. Two obstacles for a causal interpretation of heritability

The standard model we presented in the previous section makes two unrealistic assumptions: the effects of genes and the environment are (1) additive and (2) causally independent. When the additivity assumption is violated, there will be gene–environment interactions; when the independence assumption is not met, there will be gene–environment covariations. Each of these consequences leads traditionally defined heritability to lose its causal meaning. We first examine interaction with an example, then turn to covariation.

First, recall the standard example. Label the genotypic, environmental, and phenotypic variables as G, E, and P, respectively. Here, G and E refer to physical variables rather than phenotype units. The statistical function of these three variables can be summarized from the data in Table 1 as follows:

$$P_{(S)} = -10 + 10G + 5E,$$

where, again, the subscript "S" stands for "standard."

Interpreting this function causally, the causal effects of G and E on P are additive. This is so because the total phenotypic effect of a given genotypic change and a given environmental change equals the sum of each effect. This additive relation can be recognized more easily using reaction norms (Sterelny & Griffiths, 1999, p. 15). A reaction norm shows the pattern of phenotypic values of a genotype across a range of environments.[6] If we draw the reaction norms of $g1$ and $g2$ together, we obtain the graphic representation of Fig. 1a where the two lines are parallel. This parallel pattern means that the phenotypic difference made by a given genotypic change (from $g1$ to $g2$, for instance) is the same across all environments, and the phenotypic difference made by a given environmental change is the same for both genotypes. In other words, the phenotypic effect of G (or E) does not depend on the value of E (or G).[7]

Contrast this with a situation where the effects of G and E on P are *nonadditive*. Using reaction norms, this is manifested by the lines being nonparallel or even crossing, such as in Fig. 1b. In this case, the phenotypic effect of G (or E) *does* depend on the value of E (or G). It is said that there is a gene–environment statistical interaction (often denoted as $G \times E$), which posits a fundamental obstacle to interpreting ANOVA heritability estimates causally (Sesardic, 2005, p. 52; Lewontin, 2006 [1974], p. 522; Tal, 2012, p. 227; Downes & Matthews, 2020). An example of interaction (with no covariation) that corresponds to the reaction norms

[6] The reaction norm is superior to heritability in the sense that, in principle, one could derive a heritability estimate from reaction norms. However, heritability summarizes some information that is not present in reaction norms, such as the particular frequencies of genotypic alleles.

[7] The standard example represents linear relations between G and P, and between E and P. However, linearity is not a requirement for the additivity of the effects of G and E. Suppose the statistical functions are $P = E^2$ for $g1$, and $P = E^2 + 5$ for $g2$, for which there are nonlinear relations between P and E. So long as the curves of the reaction norms are parallel, the additivity assumption is met.

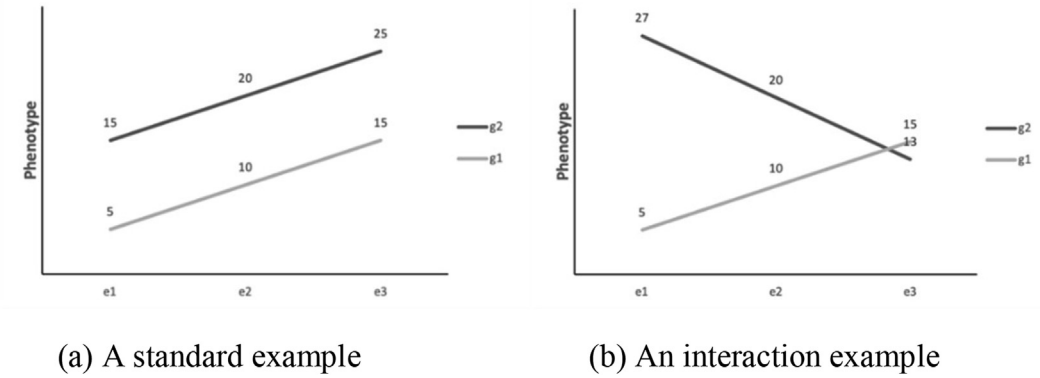

Fig. 1. Reaction norms for the additive and nonadditive examples. (a) the reaction norms for the standard example in Table 1; (b) the reaction norms for the interaction example in Table 2.

in Fig. 1b is presented in Table 2. Here, the phenotypic effect of the change of genotypes (from $g1$ to $g2$, for instance) does depend on the environment in which individuals live: it is 22 units in $e1$, 10 units in $e2$, and -2 units in $e3$.

If one performs an ANOVA from Table 2, the results are as follows:

$$V_{P(I)} = 49.7,\ V_{G(I)} = 25,\ V_{E(I)} = 0.7,$$

where the subscript "I" stands for "interaction."

The sum of $V_{G(I)}$ and $V_{E(I)}$ is smaller than $V_{P(I)}$; therefore, Equation (1) must include a third term we denote as $V_{G\times E}$, which represents the remaining variance due to statistical interaction. Thus, we obtain:

$$V_{P(I)} = V_{G(I)} + V_{E(I)} + V_{G\times E(I)}. \tag{3}$$

In our example, we can calculate that $V_{G\times E(I)} = 24$, which is relatively large compared to $V_{G(I)}$ and $V_{E(I)}$. If we apply the heritability formula anyway, we obtain $H^2_{(I)} = 0.503$.

While $H^2_{(I)}$ is statistically correct, its causal interpretation is problematic. The detection of statistical interaction indicates that the two causes are interdependent in influencing P. This challenges the decomposition of V_P into V_G and V_E because, as argued by Tal (2012, p. 234), "attempting to separate the effects of genes and environment under substantial $G \times E$ is futile, since the interaction component is ultimately some *unknown combination* of G and E" (original emphasis). In practice, if the interaction term is relatively small, one can ignore it to provide adequate heritability estimates. However, there is no guarantee that, in general, the interaction is not substantial in nature. For example, there is emerging evidence of substantial interaction in psychiatric disorders; therefore, deliberate testing of interaction hypotheses involving meta-analysis has been suggested (Moffitt et al., 2005). It has been common knowledge that—without evidence that substantial interaction is, in fact, rare—additivity cannot be assumed *a priori*.

The obstacle posited by interaction can also be presented in an interventionist way (we revisit the notion of intervention in Section 4). In the 1930s, Lancelot Hogben wrote to Fisher regarding a problem. When the lines of reaction norms are nonparallel, given heritability estimate x, it does not follow that "the variance would be reduced by x percent if there were no genetic difference" (Hogben, cited in Tabery, 2014, p. 29). To make Hogben's point concrete, let us examine our interaction example in Table 2. Given two values of G, we have two ways to intervene on G to eliminate genetic difference: either by changing $g2$ individuals in the population into $g1$ (resulting in a population only composed of $g1$ individuals) or changing $g1$ individuals into $g2$ (resulting in a population of $g2$ individuals). The remaining phenotypic variances of these two interventions can be calculated as:

$$V^{ALL-g1}_{P(I)} = 32.7,\ V^{ALL-g2}_{P(I)} = 16.7.$$

Given $V_{P(I)} = 49.7$, the reduced percentages of phenotypic variance are:

$$\Delta V^{ALL-g1}_{P(I)} = 0.342,\ \Delta V^{ALL-g2}_{P(I)} = 0.664.$$

If the heritability estimate $H^2_{(I)} = 0.503$ were to measure how much of the phenotypic variance is caused by genotypic difference, the variance should be reduced by 50.3% when the genotypic difference is eliminated. However, neither of the above two estimates is equal to $H^2_{(I)}$.[8] Hence, in the presence of interaction, it would be mistaken to interpret $H^2_{(I)}$ causally as representing the actual causal effect of genes on phenotype.

The discrepancy between $\Delta V^{ALL-g1}_{P(I)}$ and $\Delta V^{ALL-g2}_{P(I)}$ is due to the effects of G and E interacting nonadditively.[9] Recall the standard example where the additivity assumption is met. We revisit this point in more detail in Section 4; however, suffice to say here that if one were to perform the two interventions in the standard example, the remaining phenotypic variances would be equal, resulting in equal reduced percentages of variance justifying a causal interpretation of standard heritability. Our analysis in the interaction example using SCM in Section 5.1 represents a step further toward understanding the causal implication of nonadditivity and addresses recent calls in behavioral genetics and epidemiology to investigate the notion of interaction further (e.g., Baye et al., 2011; Darling et al., 2016).

The standard model of heritability analysis also assumes that G and E are either causally independent in nature or made independent in experimental settings. Each subset has an equal size (or statistically insignificant differences) in the standard example, meaning that genotypes are distributed randomly in environments. This randomization can be achieved in careful experimental designs even if G and E are causally dependent in nature. A randomized design eliminates the causal influences between G and E; hence, an absence of covariation between the two variables can be assumed.[10] However, in observational or natural studies, there is often a significant covariance between G and E, which

[8] However, the average of $\Delta V^{ALL-g1}_{P(I)}$ and $\Delta V^{ALL-g2}_{P(I)}$ is equal to $H^2_{(I)}$. Section 5.1 addresses this point in further detail.

[9] Later, Fisher replied: "dear Hogben, I think I see your point now. You are on the question of non-linear interaction of environment and heredity. The analysis of variance and covariance is only a quadratic analysis and as such only considers *additive* effects" (cited in Tabery, 2014, p. 33, emphasis added).

[10] Following Pearl (2009), randomization can be seen as an *intervention* in the sense that "subjects are 'forced' to take one treatment or another in accordance with the experimental protocol, regardless of their natural inclination" (p. 332).

Table 3
A covariation example where there is no interaction.

	$e1$	$e2$	$e3$	$\bar{g}$
$g1$	5 (10)	10 (15)	15 (20)	11.11
$g2$	15 (15)	20 (20)	25 (20)	20.45
$\bar{e}$	11	15.71	20	$\bar{p} = \mathbf{16.25}$

indicates an *unknown* (direct or indirect) causal relation between them.[11] A covariation example (with no interaction) where the genotypes are distributed nonrandomly is presented in Table 3.[12] The size for each subset is given after the values of phenotype. For instance, the size for the subset $(g1,\ e1)$ is 10.

In this case, the variances estimated by ANOVA should be weighted by the relative sizes. The results are:

$$V_{P(C)} = 37.2,\ V_{G(C)} = 21.6,\ V_{E(C)} = 12.6,$$

where the subscript "C" stands for "covariation."

The sum of $V_{G(C)}$ and $V_{E(C)}$ is, again, smaller than $V_{P(C)}$. Since there is no interaction in our covariation example, Equation (1) must include another term we denote as $2cov_{GE}$,[13] which represents the remaining variance due to some unknown dependence between G and E. Thus, we have:

$$V_{P(C)} = V_{G(C)} + V_{E(C)} + 2cov_{GE(C)}. \quad (4)$$

Here, $2cov_{GE}$ represents the component of phenotypic variance reflected by the covariance between G and E, indicating an unknown causal relation. In our case, $2cov_{GE} = 3$. Applying the heritability equation, we get $H^2_{(C)} = 0.581$.

It has long been debated whether this kind of heritability should include $2cov_{GE}$ (e.g., Block & Dworkin, 1974; Jencks, 1980; Sober, 2001). Lynch and Bourrat (2017) conclude that if the origin of covariance can be traced to genotypic causes, $2cov_{GE}$ should be incorporated in measures of heritability. Since an unobserved causal dependence might confound the causal interpretation of statistical results, such as in cases of Simpson's paradox (Pearl & Mackenzie, 2018, Chapter 5), there is broad consensus that without information on the causal relation between G and E, the presence of covariation posits a fundamental obstacle for causally interpreting ANOVA heritability.

A complete model combining interaction and covariation can be presented as:

$$V_P = V_G + V_E + V_{G\times E} + 2cov_{GE}. \quad (5)$$

To conclude, the presence of $V_{G\times E}$ and $2cov_{GE}$ raises the question of whether we should include more terms than V_G in heritability for it to be interpretable causally. Put differently, when the actual causal story is more complicated than that assumed by the standard model, heritability estimated via ANOVA can no longer reflect the relative importance of genotypes for the phenotypic variance. A promising approach is to appeal to SCM, which does not assume additivity or independence and offers structural functions and causal graphs to present complex causal relationships.

4. Standard heritability estimates using SCM

Consider the following example, modified from Pearl (2019). A match is struck, igniting a fire in a room. We pick out two causes for the fire—striking the match and the presence of oxygen—and ask the question: "which one caused the fire?" The usual answer would be "striking the match." Note, however, that both factors are necessary because the fire would not have occurred in the absence of either. Why, then, do we consider striking the match to be a more responsible cause for the fire? We first build a causal model for this example.

Following the SCM framework, a causal model $M = < U,\ V,\ F >$ represents causal relationships among endogenous[14] variables V via a set of functions F (Pearl, 2009, p. 203). U is a set of background variables (exogenous variables) determined by factors outside the model. The value of each endogenous variable is determined causally by its direct causes (or parents) $PA_i \subseteq V$ and $U_i \subseteq U$ (see also the Causal Markov Condition in Pearl, 2009, p. 30), which is presented by each function. If we can obtain the joint probability distribution of U, we can define a probability distribution over V (Pearl, 2009, p. 205). Consider FI ($FI = 1$ for the presence of a fire, $FI = 0$ for the absence of it), MA ($MA = 1$ for striking the match, $MA = 0$ for not striking it), and OX ($OX = 1$ for the presence of oxygen, $OX = 0$ for the absence of it) to be endogenous variables. Consider U_{OX} and U_{MA} to be background variables causally determining OX and MA, respectively. Assume that the fire is determined solely by the match and oxygen. Using lower case letters to denote a variable's value, the structural functions of M for this example are:

$$MA = f_{MA}(u_{MA});\ OX = f_{OX}(u_{OX});\ FI = f_{FI}(ma,\ ox).$$

$$f_{FI}:\ FI = MA{*}OX.$$

An SCM corresponds to a causal graph G where each node corresponds to a variable and the edges point from causal parent(s) to causal offspring. Any sequence of consecutive edges in a causal graph, regardless of its directionality, is called a path. For instance, an edge pointing from MA to FI represents a causal path from MA to FI. The corresponding causal graph $G(M)$ is as in Fig. 2a.

The "*do*-operator" or "$do(.)$" in the SCM framework represents the operation of an intervention. Following the interventionist theories of causation, a variable X is a cause of a variable Y, if an intervention on X produces a change in the value of Y (Woodward, 2003). Intervention is different from observation. Suppose there is a common cause Z of both X and Y. When the value of Z changes, X and Y will both change. Hence, if we only *observe* a change in X followed by a change in Y, we cannot conclude safely that Y's change is caused by X's change. In contrast, an

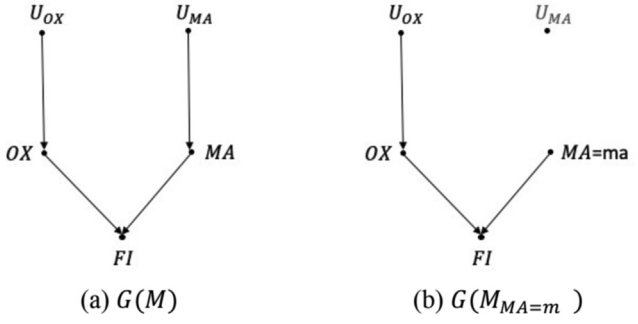

Fig. 2. Causal graphs for the Oxygen, Match, and Fire example, modified from Pearl (2019, Figs. 1 and 3). (a) G(M) is the corresponding causal graph for the pre-intervention causal model M; (b) $G(M_{MA=ma})$ is the corresponding causal graph for the post-intervention causal model $M_{MA=ma}$.

[11] Here, we do not consider the situation that a significant covariation between G and E is due to a sample selection bias (a biased selection of P might yield a statistical correlation between G and E).

[12] Reaction norms can be plotted based on Table 3, which will be the same as in Fig. 1a. However, since the two genotypes are not equally frequent in the whole population, the resulting reaction norms do not correspond to intervening in an actual population.

[13] The reason that the added term is $2cov_{GE}$ rather than cov_{GE} stems from the properties of the variance of a sum of two correlated variables. For two variables X and Y, $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$.

[14] Endogenous variables are those variables an experimenter wishes to include in the model.

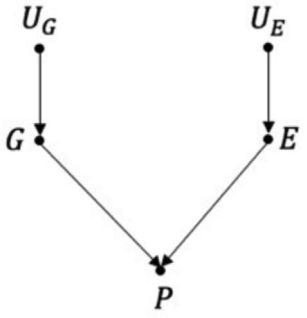

Fig. 3. Causal graph $G(M_{(S)})$ for the causal model $M_{(S)}$.

intervention $do(X = x)$ means to disconnect X from its former causal parents and set its value to x. If this operation produces a change in Y's value, we can establish that there is a causal relationship between X and Y. In this sense, $do(.)$ emulates a virtual intervention on the world from which the data are collected. It operates via replacing the structural function of the intervened-upon variable with the value specified by the intervention. The new model represents the causal story of the post-intervention world.

Let us return to our example. We can obtain $do(MA = ma)$ by replacing the former f_{MA} with $MA = ma$, and obtain a new model $M_{MA=ma}$. The corresponding causal graph $G(M_{MA=ma})$ is represented in Fig. 2b where the edge from U_{MA} to MA is deleted since, according to the new structural function, MA is now held constant (by intervention) at the value of ma, rather than being affected by U_{MA}.

A *formal* notion of the causal effect of X on Y is then defined as a function from $do(X = x)$ to the probability distribution of Y in the new model, denoted as $P(y|do(x))$ (Pearl, 2009, p. 70). Note here that x and y indicate that every value of X and Y should be computed. In other words, this definition defines a relationship between *variables*. Since a variable is constructed from each of its values, the causal effect of specific values (or events) is also defined. For instance, $P(FI = 1|do(MA = 1))$ represents the causal effect of the match-striking event (in the interventionist sense) on the event of the presence of a fire in the room.[15] Traditionally, judgments of causal strength (or causal importance) concern the relationship between events rather than variables (as defined by the formal notion).[16] Based on the formal causal effect, specific causal effects for different events can be compared against the same background to answer questions about relative importance.

In our case, a comparison of causal importance can be made in terms of how the probability of the presence of fire $P(FI = 1)$ would change given interventions on each cause from absent to present. This is because, intuitively speaking, if the change of a cause from absent to present makes a significant difference to the probability of the presence of the effect, we would think that the cause is an important one. In formalism, we can compare the following two formulas: $[P(FI = 1|do(MA = 1)) - P(FI = 1|do(MA = 0))]$ and $[P(FI = 1|do(OX = 1)) - P(FI = 1|do(OX = 0))]$. However, these two formulas cannot be compared directly unless we have conducted exact randomized controlled experiments. In practice, we often have partial interventional information or only observational data.

To address this problem, Pearl (2009, p. 92) introduced a graphical test of *identification* with a set of "identifying models" for formulas including the *do*-operator, such as $P(y|do(x))$. Given a certain SCM with a corresponding causal graph, if that causal graph is identical to (or can be transformed into) one of the identifying models, the quantity of $P(y|do(x))$ is identifiable. That is, it can be reduced to a unique formula comprising only observational probability distributions. When data can be provided, the quantity is estimated by the reduction formula. Different identifying models require different reduction formulas (Section 5.2 will discuss this further). In our example in Fig. 2a, where there is no confounding between MA and FI and between OX and FI, we can apply Equation (6) as the reduction formula:

$$P(y|do(x)) = P(y|x). \tag{6}$$

Hence, we have[17]:

$$P(FI = 1|do(MA = 1)) - P(FI = 1|do(MA = 0)) = P(FI = 1|MA = 1) - P(FI = 1|MA = 0) = P(FI = 1|MA = 1) = P(OX = 1).$$

and

$$P(FI = 1|do(OX = 1)) - P(FI = 1|do(OX = 0)) = P(FI = 1|OX = 1) - P(FI = 1|OX = 0) = P(FI = 1|OX = 1) = P(MA = 1).$$

In a terrestrial context on Earth, match-lighting is a much rarer event than the presence of oxygen, such that $P(MA = 1) \ll P(OX = 1) \approx 1$. This means striking the match is a far more efficient cause for the fire. These results precisely reflect our intuition of the relative importance of match-lighting. Note that the generality of Equation (6) and other reduction formulas mean that they can be applied for any probability distribution of certain events in different contexts.

We now apply SCM to the standard model of heritability analysis. Recall the standard example of Table 1. Since other relevant factors are randomized or remain constant, the phenotypic difference is entirely caused by G and E. Assign U_G and U_E to stand for background factors determining the values of G and E. Since we also supposed that genotypes are distributed randomly in environments, G and E are causally independent. Thus, we have the causal model $M_{(S)}$:

$$G = f_{(S)G}(u_G);\ E = f_{(S)E}(u_E);\ P = f_{(S)P}(g, e),$$

and

$$f_{(S)P} : P = -10 + 10G + 5E,$$

where "S" stands for "standard." Fig. 3 represents the corresponding causal graph $G(M_{(S)})$.

Three remarks should be made about $M_{(S)}$. First, $M_{(S)}$ combines all the heritable causes of the phenotype into one variable G, and all the environmental causes into one variable E. A more detailed SCM, with more fine-grained variables for genotypic and environmental causes, could be given for other applications. Second and related, the direct cause is a *relative* term. A causal graph shows a direct cause with an arrow edge from a cause variable to its effect variable. This does not exclude the possibility of intermediate variables in other SCMs or the real world. Third, the values of G and E are not the binary values of "presence" and "absence" in the Oxygen, Match, and Fire example, but alternative

[15] Here, the term "event" is used in a commonsense way that can be represented by the values of a variable.

[16] Causal importance has been related to necessity and sufficiency. A cause is necessary for an effect if the former's absence leads to the latter's absence; a cause is sufficient for an effect if the former's presence leads to the latter's presence. This distinction goes back to J. S. Mill (1843). Semi-formal explications have been provided in Good (1961), Mackie (1965), and Rothman (1976).

[17] Pearl (2009) explores causal necessity and sufficiency in terms of SCM, resulting in a measure of causal strength with the notions of the "probability of sufficiency" (*PS*), "probability of necessity" (*PN*), and "probability of necessity and sufficiency" (*PNS*). If a change of the cause from absent to present cannot, under any circumstance, make the effect change from present to absent, following Pearl (2009, p. 291), $PNS = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0))$, which is the same as the comparison we made here.

genotypes and alternative environments. Having assumed that the presence of a genotype and an environment is necessary for the presence of a phenotype, $M_{(S)}$ concerns which kinds of genotype and environment will cause which type of phenotype. Hence, an intervention on G does not mean bringing about (or removing) a genotype but fixing it at one particular genotypic value.

By examining $M_{(S)}$ and $G(M_{(S)})$, we can see the assumptions of additivity and independence made by the standard model. First, $G(M_{(S)})$ shows that there is no causal path between G and E, indicating causal independence between G and E. Although this remark may appear insignificant in the context of the standard model, a departure from this assumption (to which we return in Section 5.2) will prove crucial. Second, although $G(M_{(S)})$ has the same causal structure as $G(M)$ in the Oxygen, Match, and Fire example, the structural functions are notably different. In the case of $f_{(S)P}$, the phenotypic effects of G and E are additive, indicating no gene–environment interaction. In contrast, f_{FI} tells us that there is an interaction between *MA* and *OX* concerning *FI*. That said, the causal effect of a change in *MA* will depend on the value at which *OX* is set. For instance, when oxygen is absent, changing the value of *MA* from 0 to 1 will lead to no change in *FI*. The fact that SCM can capture the Oxygen, Match, and Fire example gives us confidence that SCM can also provide legitimate causal analyses in interaction cases involving genetic causation (we revisit this in Section 5.1).

The most relevant feature of SCM for our purpose is the conceptual distinction between causality and statistics. Although $f_{(S)P}$ has the same format as the statistical function among G, E, and P given in Section 3, the difference between them is that a structural function represents an invariant mechanism that determines the values of the left variable from those of the right variables.[18] Hence, an SCM comprising structural functions represents the causal theory covering a set of causal worlds (actual or possible); each causal world has a particular probability distribution of exogenous variables (Pearl, 2009, p. 207). The usefulness of the distinction between causality and statistics, as Pearl sees it—and we agree—"lies primarily in helping investigators trace the assumptions that are needed to support various types of scientific claims" (2009, p. 334).

One might notice that causal graphs resemble path diagrams in path analysis in evolutionary genetics (e.g., Otsuka, 2014). This should not come as a surprise since path analysis is an ancestor of SCM. As such, it can be used to estimate the structural parameters inan SCM (e.g., the numbers "10" and "5" in $f_{(S)P}$) by regression coefficients in linear systems. However, a structural function, in itself, does not depend on the method used to produce the estimation. SCM can accommodate a broader range of assumptions, such as nonlinearity or even unspecified parameters (for further details, see Pearl, 2009, p. 367). It should be noted that modern regression techniques have been designed to devise innovative experimental tools such as instrumental variables and regression discontinuity designs. This last point is important in the context where it remains uncertain whether SCM is a more powerful approach than regression approaches for empirical work (see Imbens, 2020, who argues that regression approaches are preferable in economics). As we see it, SCM and regression techniques are complementary tools for causal analysis; each has relative merits for different fields and different tasks.

If we return to the standard example, according to Pearl's graphical test of identification, we can reduce $P_{(S)}(p|do(g))$ by applying Equation (6). $P_{(S)}(p|g)$ can be obtained directly from observational data in Table 1. Thus, we obtain the result shown in Table 4. The values of P are shown in the columns of p, and the values of E for every p are shown in the brackets.

The post-intervention probability distribution of P shown in Table 4 presents the formal causal effect of G on P. However, heritability is estimated by the ratio of phenotypic *variance*. To link the generic causal effect from SCM to heritability, we derive variance from the probability distribution. According to Table 4, although the post-intervention phenotypic values are different for $do(g1)$ (the left two columns in Table 4) and $do(g2)$ (the right two columns in Table 4), the post-intervention phenotypic variances are equal such that:

$$V_{P(s)}^{do(g1)} = V_{P(s)}^{do(g2)} = 16.7.$$

Given pre-intervention variance $V_{P(S)} = 41.7$, the reduced percentages of phenotypic variance are:

$$\Delta V_{P(s)}^{do(g1)} = \Delta V_{P(s)}^{do(g2)} = (41.7 - 16.7)\Big/41.7 = 0.60.$$

Compare the above two estimates to the ANOVA heritability estimate in Section 2 ($H_{(S)}^2 = 0.60$). The three results are all equal. The estimates of the reduced percentages of variance, precisely speaking, concern a particular type of causal effect—that is, the causal effect of *eliminating the difference in G (by intervention) on the reduction in the variance for P*. Accordingly, the causal effect of eliminating the difference in E (by intervention) on the reduction in the variance for P is 0.4. Therefore, in comparing the causal effects of eliminating the difference in each cause on the reduction in the variance for P, ANOVA heritability estimates are said to reflect the relative importance of genetic cause on the phenotype.

Two remarks should be made about this type of causal effect—one on the measure of the effect and the other on the specific intervention on the cause. For the former, since the variance of P is a measurement summarizing all the values of P, it depends on the specific population targeted by experimenters. That said, this type of causal effect depends on the values of G and E *actually* manifested in the population; therefore, it might differ for another population with different realizations of G and E values. This explains why heritability estimates are fundamentally *local*. For the latter, the two estimates, which depend on two interventions, should be seen as two separated results. According to $M_{(S)}$, the variable G has two mutually exclusive values, reflecting our assumption that $g1$ and $g2$ are genotypic variations, and an individual cannot simultaneously have both $g1$ and $g2$. It follows that any intervention $do(g)$, by definition of the *do*-operator, can only be realized as either $do(g1)$ or $do(g2)$. In contrast with $do(g1, e1)$, where G is intervened on to become $g1$, and E is intervened on simultaneously to become $e1$, $do(g1, g2)$ is meaningless.[19]

Examining this more closely, $\Delta V_{P(s)}^{do(g1)}$ presents the causal effect of eliminating genotypic differences by fixing the whole population at $g1$ on the reduction in the total phenotypic variance; $\Delta V_{P(s)}^{do(g2)}$ presents the causal effect of eliminating genotypic differences at $g2$ on reducing phenotypic variance. Since the two estimates are equal, either $do(g1)$ or $do(g2)$ in eliminating genotypic differences would reduce the phenotypic variance by 60%.[20] We can infer that the causal effect of eliminating genotypic differences by setting the whole population *at any genotypic*

Table 4
The causal effect of G on P in the standard example presented in Table 1.

$do(g1)$		$do(g2)$	
p	$P(p)$	p	$P(p)$
5 ($e1$)	1/3	15 ($e1$)	1/3
10 ($e2$)	1/3	20 ($e2$)	1/3
15 ($e3$)	1/3	25 ($e3$)	1/3

[18] Here, "invariant" means that the causal mechanism represented by a structural function is modular and stable (for details, see Woodward, 2003, Chapter 6). This notion of invariance differs from but is compatible with the formal measure of causal invariance proposed by Bourrat (2021a).

[19] The same occurs when there are three genotypes (or more): $g1$, $g2$, and $g3$. It makes no sense to perform $do(g1, g2, g3)$, or any combination of two of these.

[20] When there are three or more genotypes exhibited in the population, assuming no covariation or interaction, the causal effects of the elimination of difference in G at any genotypes on the change in the variance for P would all be equal.

value on the change of phenotypic variance is 60%. Therefore, $H^2_{(S)}$, which is equal to $\Delta V^{do(g1)}_{P(s)}$ and $\Delta V^{do(g2)}_{P(s)}$, is a summary of the causal effects of eliminating genotypic differences on the change of phenotypic variance. However, when the causal effect of eliminating genotypic differences by setting the whole population *at a given genotypic value* is different from that *at another genotypic value* (as is the case in interaction cases), we can no longer infer a single type of causal effect across all the values of G. In this case, a substantial causal analysis should involve information regarding the precise interventions on G with different genotypic values.

To conclude, the application of SCM to the standard model grounds standard heritability estimates in causal terms. This is consistent with Bourrat (2021b), Lynch and Bourrat (2017), Sesardic (2005), and Tal (2012). However, SCM provides a more explicable causal interpretation of genetic causation than the classical interpretation given by the standard model of heritability analysis.

5. Interaction and covariation: an SCM approach

This section demonstrates why interaction and covariation posit fundamental obstacles for a causal interpretation of ANOVA heritability estimates within SCM. Further, it illustrates the kind of causal analysis that can be derived from SCM. Section 5.1 addresses interaction, Section 5.2 addresses covariation, and Section 5.3 extends the application of SCM to additional types of causal effect that can provide answers to individual-level questions.

5.1. Interaction

Recall the interaction example. To apply SCM, we first build a causal model $M_{(I)}$, where the subscript"I" stands for "interaction." We have the following structural functions for $M_{(I)}$:

$$G = f_{(I)G}(u_G);\ E = f_{(I)E}(u_E);\ P = f_{(I)P}(g,\ e).$$

In the presence of interaction, the genotypic effect and the environmental effect are nonadditive, and $f_{(I)P}$ does not have a similar regression format as $f_{(S)P}$ in $M_{(S)}$. Since G and E are causally independent in this example, the corresponding causal graph $G(M_{(I)})$ is the same as $G(M_{(S)})$ in Fig. 3. The graphical identification depends solely on the causal graph; therefore, we can also apply Equation (6) in this example, as shown in Table 5:

We then calculate the post-intervention phenotypic variances for $do(g1)$ and $do(g2)$:

$$V^{do(g1)}_{P(I)} = 16.7,\ V^{do(g2)}_{P(I)} = 32.7.$$

Given $V_{P(I)} = 49.7$, the reduced percentages of phenotypic variance are:

$$\Delta V^{do(g1)}_{P(I)} = 0.664,\ \Delta V^{do(g2)}_{P(I)} = 0.342.$$[21]

Hence, the causal effect of eliminating genotypic difference by fixing the population at $g1$ on the reduction of phenotypic variance is 66.4%, and that at $g2$ is 34.2%. In contrast with the standard example, the two estimates are now unequal.[22] Here, SCM provides an explanation that is consistent with Hogben's problem.

As mentioned before, a SCM with a particular distribution of exogenous variables represents a causal world. In this example, given the phenotypic data (shown in Table 2) (hypothetically) collected from nature or an experiment, which fixes the distribution of exogenous variables, we now possess complete causal information about the causal world. Denote this world with interaction as "w^I." To measure the causal effect of G on the variance of P in w^I, we virtually intervene on G: namely, $do(g1)$ resulting in w^{g1} with all $g1$ individuals, and $do(g2)$ resulting in w^{g2} with all $g2$ individuals. As we can see from Table 2, the $g1$ group (the first row in Table 2, which corresponds to w^{g1}) and the $g2$ group (the second row in Table 2, which corresponds to w^{g2}) exhibit different degrees of phenotypic spread in specific environments. In particular, 5 units per environmental change for w^{g1} and 7 units for w^{g2}, which leads to unequal phenotypic variances in w^{g1} and w^{g2}, and further leads to the inequality of $\Delta V^{do(g1)}_{P(I)}$ and $\Delta V^{do(g2)}_{P(I)}$. In contrast, in the standard example (Table 1), the degrees of phenotypic spread in environments are the same for the $g1$ and $g2$ groups (5 units per environmental change); hence, $V^{do(g1)}_{P(s)}$ and $V^{do(g2)}_{P(s)}$ are equal.

Table 5
The causal effect of G on P in the interaction example presented in Table 2.

$do(g1)$		$do(g2)$	
p	$P(p)$	p	$P(p)$
5 ($e1$)	1/3	27 ($e1$)	1/3
10 ($e2$)	1/3	20 ($e2$)	1/3
15 ($e3$)	1/3	13 ($e3$)	1/3

Now, a possible move is to average over $\Delta V^{do(g1)}_{P(I)}$ and $\Delta V^{do(g2)}_{P(I)}$, which amounts to 0.503, the heritability estimate obtained from ANOVA ($H^2_{(I)}$). Does this average warrant a causal interpretation for $H^2_{(I)}$? As we see it, the answer is twofold. First, it is not easy to see valuable applications of this average in causal practice. As mentioned above, the two estimates represent two *separate* results. We can manipulate the two numbers to compute an average and assert that it means the *average* causal effect of eliminating genotypic differences on the change of variance of P. However, we doubt that this sense of causal effect would be useful in causal reasoning because it could lead to some serious misunderstandings. Second, regardless of the application problem, this average does not answer the question supposed to be answered by the heritability concept. The fact that the two estimates are unequal means that there are two different types of effects corresponding to two interventions on G. In this case, we cannot infer a single type of causal effect of eliminating genotypic differences on the reduction of phenotypic variance. Hence, ANOVA heritability estimates in the presence of interaction (such as $H^2_{(I)}$) can no longer be used to answer the relative importance question.

To conclude, consistent with our analysis in Section 3, without equal estimates for the causal effect of different genotypes on the change of phenotypic variance, an ANOVA heritability estimate in the presence of interaction cannot be interpreted as the relative importance of genetic causation. Nevertheless, a legitimate causal analysis with the information of each possible intervention on the genotypes can be specified by SCM.

5.2. Covariation

The standard model of heritability analysis also assumes that G and E are causally independent. However, the presence of $2cov_{GE}$, such as in our covariation example, indicates some unknown causal relation between G and E. Three kinds of causal structures are compatible with the presence of $2cov_{GE}$, as shown in Fig. 4.

First, there is a direct cause from G to E, as presented in Fig. 4a. This can occur when the genotype causally influences the individual's choice of different environments; that is, individuals with specific genotypes are inclined to stay in particular environments. Second, there is a direct cause from E to G, as presented in Fig. 4b. Recent developments in epigenetics lead some authors to argue for a pluralistic view of heredity or inclusive inheritance (Laland et al., 2015), including heritable epigenetic marks as inherited material (Bourrat & Lu, 2017; Lu & Bourrat, 2018). Although this might be rare in nature, it can occur when some heritable epigenetic marks

[21] These results are consistent with those from Hogben's method of analysis (see Section 3).

[22] In the presence of interaction, when there are three or more genotypes, the reduced percentages of phenotypic variance by intervening on all $g1$, $g2$, $g3$ (etc.) individuals, will be different.

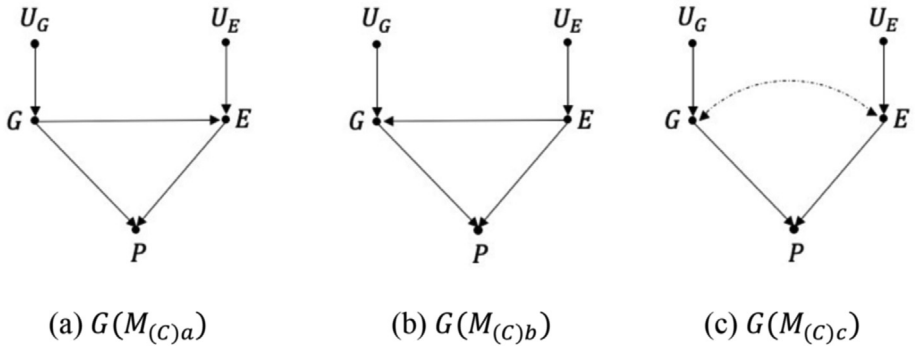

Fig. 4. Possible causal graphs for the covariation example. (a) $G(M_{(C)a})$ is the corresponding causal graph for $M_{(C)a}$ with a direct causal relationship from G to E (b) $G(M_{(C)b})$ is the corresponding causal graph for $M_{(C)b}$ with a direct causal relationship from E to G; (c) $G(M_{(C)c})$ is the corresponding causal graph for with an unobservable common cause for G and E, symbolised by the dashed arc.

Table 6
The causal effect of G on P in the covariation example of Fig. 4a.

$do(g1)$		$do(g2)$	
p	$P(p)$	p	$P(p)$
5 ($e1$)	2/9	15 ($e1$)	3/11
10 ($e2$)	3/9	20 ($e2$)	4/11
15 ($e3$)	4/9	25 ($e3$)	4/11

that have a causal influence on the phenotype are also influenced by environmental factors during development[23] (see also Tal et al., 2010). Third, there is an unobservable common cause for G and E, as presented in Fig. 4c. This occurs when an unknown (genotypic or environmental) cause that experimenters do not include into G or E in the experimental design is present.

According to Pearl's graphical identification test, we have $P_{(C)a}(p|do(g))$ for $M_{(C)a}$ reduced by Equation (6), as shown in Table 6:

The post-intervention phenotypic variances can be calculated as:

$$V_{P(C)a}^{do(g1)} = 15.4,\ V_{P(C)a}^{do(g2)} = 15.7.$$

Given the pre-intervention phenotypic variance $V_{P(C)} = 37.2$, we obtain the reduced percentages of variance as:

$$\Delta V_{P(C)a}^{do(g1)} = 0.586,\ \Delta V_{P(C)a}^{do(g2)} = 0.578.$$

Since the two estimates are unequal, we can only legitimately infer two different types of causal effect of eliminating genotypic differences by fixing the population at a particular genotypic value on the change in phenotypic variance. In particular, given $M_{(C)a}$, the causal effect of eliminating genotypic differences at $g1$ on the reduction of phenotypic variance is 58.6%, and that at $g2$ is 57.8%. As we demonstrated in Section 5.1 with the interaction example, averaging these two values would not adequately represent a single type of causal effect of eliminating genotypic differences on the change of phenotypic variance.

It should be noted that the causal effect identified here is the *total* effect, which includes both direct and indirect effects. Conversely, in the standard and interaction examples, there are no indirect effects, only direct ones. A "direct effect" is defined in the SCM as a causal effect that is not mediated by other variables. In Fig. 3, the effect of G on P is not mediated by E. Hence, the total effect of eliminating the differences in G on the change in the variance of P equals its direct effect. However, here

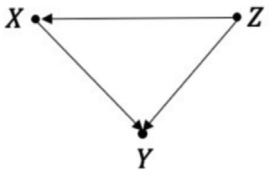

Fig. 5. A simple example of the back-door path. Considering the causal relationship from X to Y, $X \leftarrow Z \rightarrow Y$ represents a back-door path.

in Fig. 4a, G also indirectly influences P mediated by E; thus, the total effects do not equal the direct effects. To disentangle the different contributions of G to P (direct and indirect effects), to the extent this is possible, requires some extra tools. Presenting such tools would exceed the scope of this paper; see Pearl (2009, Chapter 4.5) for further details.

As for the causal structures of Fig. 4b and c, Equation (6) cannot be applied because they include a confounding pattern. When measuring the causal effect of X on Y, one should prevent the influence of any path with a spurious correlation between X and Y. This kind of path is called a "back-door path" and is defined as an unblocked path between X and Y with arrows pointing at X.[24] A common cause is a typical example of a back-door path, as presented in Fig. 5. Apart from the direct path from X to Y, there is an unblocked path between X and Y with an arrow pointing at X ($X \leftarrow Z \rightarrow Y$). The influence of this latter path would confound the measurement of causal effect from X to Y.[25]

According to Pearl, it takes two steps to eliminate the influence of back-door paths. First, a back-door path can be blocked by conditioning certain variables to stop the back-door flow of statistical dependence between X and Y. According to the definition of d-separation, in the example of Fig. 5, via conditioning on Z, the originally unblocked back-door path is now blocked. If there are multiple back-door paths from X to Y, we should find a set of variables Z' that blocks all of them (see the Back-Door criterion in Pearl, 2009, p. 78). Second, the causal effect from X to Y can be given by adjusting for Z'. Adjustment amounts to partitioning the population into homogeneous groups relative to Z', assessing the causal effect of X on Y in each homogenous group, and then averaging[26] the results. Hence, in the case with back-door paths, the formal

[23] Here is a classic example of epigenetic inheritance (Morgan et al., 1999). Mice with the same genotype display a range of colors of their fur, which are due to a difference in DNA methylation levels on the promoter of the dominant *agouti* gene. This epigenetic pattern can be inherited through generations. Moreover, a diet rich in methyl donors might induce a phenotypic change via the change of inherited epigenetic patterns.

[24] A path is "unblocked" when it is not d-separated (see Pearl, 2009, pp. 16–17, for the definition of d-separation).

[25] There are no back-door paths from G to P in Figs. 3 and 4a; therefore, the causal effect of G on P can be reduced directly to observational probabilities using Equation (6).

[26] This sense of averaging should be distinguished from the sense mentioned in Section 5.1. The latter means to average the causal effects of two separated interventions, with each intervention operating on the total population. Here, it means averaging the causal effects for subpopulations within one total population, resulting in a single type of effect for this total population.

Table 7
The causal effect of G on P in the covariation example of Fig. 4b and c.

$do(g1)$		$do(g2)$	
p	$P(p)$	p	$P(p)$
5 ($e1$)	5/20	15 ($e1$)	5/20
10 ($e2$)	7/20	20 ($e2$)	7/20
15 ($e3$)	8/20	25 ($e3$)	8/20

causal effect can be reduced by Equation (7) (Pearl, 2009, p. 79):

$$P(y|do(x)) = \sum_{Z} P(y|x,z')P(z'). \quad (7)$$

Since the variable E blocks the back-door path from G to P both in Fig. 4b ($G \leftarrow E \rightarrow P$) and in Fig. 4c ($G \leftrightarrow E \rightarrow P$), we can apply Equation (7) to both cases such that:

$$P_{(C)b,c}(p|do(g)) = \sum_{E} P(p|g,e)P(e).$$

The causal effect of G on P for $M_{(C)b}$ and $M_{(C)c}$ is given in Table 7:

Since the degrees of phenotypic spread in environments are the same for $g1$ and $g2$ groups (5 units per environmental change), the post-intervention phenotypic variances are equal for $do(g1)$ and $do(g2)$:

$$V_{P(C)b,c}^{do(g1)} = V_{P(C)b,c}^{do(g2)} = 15.7.$$

Given the pre-intervention variance $V_{P(C)} = 37.2$, the reduced percentages of phenotypic variance are also equal:

$$V_{P(C)b,c}^{do(g1)} = V_{P(C)b,c}^{do(g2)} = 0.578.$$

This means that the reduced percentage of phenotypic variance by eliminating genotypic differences by fixing the population *at either genotypic value* is 57.8%. It follows that the causal effect of eliminating genotypic differences on the change of phenotypic variance is 57.8%. Hence, given $M_{(C)b}$ or $M_{(C)c}$, the causal magnitude of the phenotypic variance contributed by genotypic differences is 57.8%. Accordingly, the contribution of the environmental differences is 42.2%. Thus, these figures answer the question regarding the relative importance of genes and the environment in causing the phenotypic variance in the population.

However, the estimates are not the same as the heritability estimated by ANOVA ($H_{(C)}^2 = 0.581$). The difference occurs precisely because the standard model assumes G and E to be causally independent. Given $M_{(C)b}$ or $M_{(C)c}$, the influence of G on P derived from the back-door path should be eliminated in estimating the causal effect from G to P. The SCM framework does this by implementing Equation (7). However, ANOVA heritability analysis implicitly assumes independence without eliminating the influence of back-door paths giving an overestimated result.

For $M_{(C)b}$ and $M_{(C)c}$, a single type of causal effect of eliminating genotypic differences on the change of phenotypic variance can be given; whereas, for $M_{(C)a}$, we can only infer the causal effects *at a given genotypic value*. This means that different causal theories will produce different measurements of the causal effect. Hence, in the presence of covariation, information about the causal relationship between G and E is required to measure the causal effect of G on P accurately. Others have already raised this point (see, e.g., Lynch & Bourrat, 2017; Tal, 2012). However, here, SCM is helpful to clarify the correspondence between causal theories and the identification of types of causal effects. When there is a direct cause from G to E, the ANOVA heritability estimate is misleading because it does not reflect the type of causal effect that can be derived from the standard model. When there is a direct cause from E to G, or there is an unobservable common cause for both G and E, ANOVA heritability is misleading because the causal effects are overestimated via including spurious influences.

Table 8
An interaction and covariation example.

	$e1$	$e2$	$e3$	$\bar{g}$
$g1$	5 (10)	10 (15)	15 (20)	11.11
$g2$	27 (15)	20 (20)	13 (20)	19.36
$\bar{e}$	18.2	15.71	14	$\bar{p} = \mathbf{15.65}$

5.3. Extending the types of causal effects using SCM

Having investigated the presence of interaction and covariation using SCM, in this section, we first consider an example that combines interaction and covariation. Then, we propose that SCM can help extend the identification of causal effects of the genes on phenotype in more ways than traditional heritability analysis.

Consider the example presented in Table 8.

In this example, both interaction and covariation occur. If we apply ANOVA, we have:

$$V_{P(IC)} = 40.7,\ V_{G(IC)} = 16.9,\ V_{E(IC)} = 2.7,$$

where the subscript "IC" stands for "interaction *and* covariation."

The sum of $V_{G(IC)}$ and $V_{E(IC)}$ is smaller than $V_{P(IC)}$. To apply the heritability equation anyway, we have $H_{(IC)}^2 = 0.415$.

Without further information about the causal relationship between G and E, there may be three possible causal structures, as shown in Fig. 4. Build $M_{(IC)a}$, $M_{(IC)b}$, and $M_{(IC)c}$, respectively. The causal effect of G on P can be given by $P(p|do(g))$. The results are given in Table 9.

We first calculate post-intervention phenotypic variances:

$$V_{P(IC)a}^{do(g1)} = 15.4,\ V_{P(IC)a}^{do(g2)} = 30.7,\ V_{P(IC)b,c}^{do(g1)} = 15.7,\ V_{P(IC)b,c}^{do(g2)} = 30.7.$$

Given $V_{P(IC)} = 40.7$, the reduced percentages of phenotypic variance are:

$$\Delta V_{P(IC)a}^{do(g1)} = 0.622,\ \Delta V_{P(IC)a}^{do(g2)} = 0.246,\ \Delta V_{P(IC)b,c}^{do(g1)} = 0.514,\ \Delta V_{P(IC)b,c}^{do(g2)} = 0.246.$$

Following SCM, we can say that when there is a direct cause from G to E, the causal effect of eliminating genotypic difference by fixing the population at $g1$ on the reduction of phenotypic variance is 62.2%, and that at $g2$ is 24.6%. When there is a direct cause from E to G or an unknown common cause of G and E, the causal effect at $g1$ is 51.4%, and that at $g2$ is 24.6%. Since none of these four estimates equals $H_{(IC)}^2$ (0.415), the ANOVA heritability estimate is misleading in both interaction and covariation situations.

Apart from phenotypic *variance*, more types of causal effect can be derived within SCM to extend our understanding of genetic causation. One kind of causal query narrows down the scope of targeting specific individuals that live in particular environments. To take $M_{(IC)a}$ as an example, we can measure the causal effect for $g2$ individuals who live in $e2$:

$$P_{(IC)a}(P = 10|do(g1), e2) = 1.$$

It follows that, given $M_{(IC)a}$ for a $g2$ individual who lives in $e2$, if it had $g1$ instead, the probability of its height being 10 units is 1. Given that this individual's height is 20 units in the original data, we can predict that, with everything else being unchanged, if it were mutated to have $g1$, its height would decrease by 10 units.[27] In this sense, we can say that its

[27] The reasoning when there are three genotypes in the population is as follows. Given $P(P = 10|do(g1), e2) = 1$, this means that for a $g2$ or $g3$ individual who lives in $e2$, the probability of its height being 10 units would be 1 if this individual was $g1$ instead. Given $g2$ and $g3$ individuals' original height $p2$ and $p3$, respectively, if they were mutated to have $g1$, the resulting height change would be $10 - p2$ and $10 - p3$ units, respectively. The same reasoning applies with more genotypes.

Table 9
The causal effect of G on P in the interaction and covariation example presented in Table 8.

$P_{(IC)a}(p\|do(g))$				$P_{(IC)b,c}(p\|do(g))$			
$do(g1)$		$do(g2)$		$do(g1)$		$do(g2)$	
p	$P(p)$	p	$P(p)$	p	$P(p)$	p	$P(p)$
5 ($e1$)	2/9	27 ($e1$)	3/11	5 ($e1$)	5/20	27 ($e1$)	5/20
10 ($e2$)	3/9	20 ($e2$)	4/11	10 ($e2$)	7/20	20 ($e2$)	7/20
15 ($e3$)	4/9	13 ($e3$)	4/11	15 ($e3$)	8/20	13 ($e3$)	8/20

genotype determines 50% of its height.[28] Although we can infer similar conclusions from the collected data in Table 9, legitimate causal analysis on the *individual* level is thought not to be straightforwardly available in traditional heritability estimates.[29] However, with the help of the SCM framework, it is at least possible to ask causal questions on the individual and population levels systematically.

In accordance with the analysis presented in the previous two sections, ANOVA heritability estimates in cases where there is substantial interaction or covariation (or both) are misleading. To the extent that they are identifiable in SCM, the causal effects of genes on a phenotype or phenotypic variance or more specific causal claims can be provided.

6. Conclusions

This paper has clarified the disputes regarding heritability estimated by ANOVA in two respects. First, there exist problems of defining the exact causal meaning of heritability estimated under the assumptions of the standard model. Applying the SCM framework helps clarify the type of causal effect reflected by standard heritability and establishes that standard heritability can be interpreted causally more robustly. Second, there is a consensus that a causal interpretation of heritability is only warranted where there is no interaction or covariation. Applying SCM also produces results generally consistent with this consensus. Although we agree that heritability by itself does not warrant a causal interpretation in the presence of interaction and covariation, we show that specific types of causal effect (typically requiring more than a single number) can be devised to establish genetic effects on the phenotypic variance. In particular, we articulate the two assumptions under the standard model with clear notations: the specific nonadditive structural function and the causal graphs where independence is explicit. These notations provide the formal language for connecting interaction and covariation to standard heritability. Thus, we can reach a coherent understanding of genetic causation in cases with and without covariance and interaction.

Further, we show that SCM permits us to distinguish different types of causal effect by answering individual-level questions regarding genetic causation. This extension of the SCM application can supplement traditional heritability analysis by providing a more substantial causal analysis of genetic causation. As mentioned briefly in Section 2, the debate about the causal interpretation of heritability has been advanced to distinguish two approaches: the variation-partitioning approach and the mechanism-elucidating approach (Tabery, 2014). The former is a population-level approach, while the latter is an individual-level one. It has been posited that the variation-partitioning approach, of which heritability analysis is one instance, has no implication for the mechanism-elucidating approach (see Tabery, 2014, p. 99; Waters, 2007, p. 26). However, as our SCM application shows, identifying the causal effect on phenotype *variance* requires at hand specific structural functions, which represent invariant causal *mechanisms*, indicating that the variation-partitioning approach cannot be separated from the elucidating-mechanism approach. Empirically speaking, new technologies (e.g., genome-wide association study) have located high numbers of DNA sequences as actual difference makers that influence certain traits (Bourrat et al., 2017; Bourrat & Lu, 2017; Frazer et al., 2009; Read & Sharma, 2021; Visscher et al., 2012), and molecular biology continues to offer new information regarding the causal mechanisms of trait production (McGue & Carey, 2017). With more empirical data coming in, SCM will be a promising tool to assist with integrating these two approaches.

[28] When reaction norms are plotted in the interventionistic sense rather than the observational sense, they can provide similar causal analysis at the individual level. However, reaction norms alone cannot provide causal analysis of variance for a population with unequal frequent genotypes.

[29] According to Lee and Chow (2013), the causal meaning of Fisher's average effect of individual alleles corresponds to this kind of causal analysis, but with no formal language to distinguish causation from correlation.

Acknowledgments

We are thankful to Liqian Zhou and Xiaoan Wu for their comments on an earlier version of this paper. We are also grateful to Jun Otsuka, Wesley Fang, and Paul Griffiths for valuable advice and discussions during the 9th Asia-Pacific Conference on Philosophy of Science and a series lecture for Macquarie University. Finally, we thank three anonymous reviewers and the editors for feedback. QL's research was supported under the 2021 National Social Science Foundation of China (Grant No. 21CZX024). PB's research was supported under the Australian Research Council's Discovery Projects funding scheme (Project Number DE210100303).

References

Baye, T. M., Abebe, T., & Wilke, R. A. (2011). Genotype–environment interactions and their translational implications. *Personalized Medicine, 8*(1), 59–70. https://doi.org/10.2217/pme.10.75

Block, N. J., & Dworkin, G. (1974). IQ, heritability and inequality, part 2. *Philosophy & Public Affairs, 4*(1), 40–99.

Bourrat, P. (2021a). Measuring causal invariance formally. *Entropy, 23*(6), 690.

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

Bourrat, P. (2022). Unifying heritability in evolutionary theory. Studies in History and. *Philosophy of Science, 91*, 201–210. https://doi.org/10.1016/j.shpsa.2021.10.019

Bourrat, P., & Lu, Q. (2017). Dissolving the missing heritability problem. *Philosophy of Science, 84*(5), 1055–1067.

Bourrat, P., Lu, Q., & Jablonka, E. (2017). Why the missing heritability might not be in the DNA. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology, 39*(7). https://doi.org/10.1002/bies.201700067

Darling, K. W., Ackerman, S. L., Hiatt, R. H., Lee, S. S.-J., & Shim, J. K. (2016). Enacting the molecular imperative: How gene-environment interaction research links bodies and environments in the post-genomic age. *Social Science & Medicine, 155*, 51–60.

Downes, S. M., & Matthews, L. (2020). Heritability. In *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/spr2020/entries/heredity/.

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

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.

Frazer, K. A., Murray, S. S., Schork, N. J., & Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics, 10*(4), 241–251.

Gebharter, A. (2017). *Causal nets, interventionism, and mechanisms: Philosophical foundations and applications*. Springer.

Godfrey-Smith, P. (2007). Conditions for evolution by natural selection. *The Journal of Philosophy, 104*(10), 489–516.

Good, I. J. (1961). A causal calculus (I). *The British Journal for the Philosophy of Science, 11*(44), 305–318.

Imbens, G. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature, 58*(4), 1129–1179.

Jacquard, A. (1983). Heritability: One word, three concepts. *Biometrics, 39*(2), 465–477.

Jencks, C. (1980). Heredity, environment, and public policy reconsidered. *American Sociological Review, 45*(5), 723–736.
Jensen, A. (1969). How much can we boost IQ and scholastic achievement. *Harvard Educational Review, 39*(1), 1–123.
Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., & Odling-Smee, J. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. *Proceedings of the Royal Society B, 282*(1813), 20151019.
Lee, J. J., & Chow, C. C. (2013). The causal meaning of Fisher's average effect. *Genetics Research, 95*, 89–109. https://doi.org/10.1017/S0016672313000074
Lewontin, R. C. (2006). The analysis of variance and the analysis of causes. *International Journal of Epidemiology, 35*(3), 520–525.
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.* , Article a040519. https://doi.org/10.1101/cshperspect.a040519
Lynch, K. E., & Bourrat, P. (2017). Interpreting heritability causally. *Philosophy of Science, 84*(1), 14–34.
Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly, 2*(4), 245–264.
McGue, M., & Carey, B. E. (2017). Gene–environment interaction in the behavioral sciences: Findings, challenges, and prospects. In P. H. Tolan, & B. L. Leventhal (Eds.), *Gene-environment transactions in developmental psychopathology* (pp. 35–57). Springer.
Mill, J. S. (1843). *A System of Logic: Ratiocinative and inductive* (Vol. 1). John W. Parker (Ed.).
Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry, 62*(5), 473–481. https://doi.org/10.1001/archpsyc.62.5.473
Morgan, H. D., Sutherland, H. G., Martin, D. I., & Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics, 23*(3), 314–318.
Oftedal, G. (2005). Heritability and genetic causation. *Philosophy of Science, 72*(5), 699–709.
Okasha, S. (2006). *Evolution and the levels of selection*. Oxford University Press.
Otsuka, J. (2014). Causal foundations of evolutionary genetics. *The British Journal for the Philosophy of Science, 67*(1), 247–269.
Pearl, J. (2003). Reply to Woodward. *Economics and Philosophy, 19*(2), 341–344. https://doi.org/10.1017/s0266267103001196
Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
Pearl, J. (2019). Sufficient causes: On oxygen, matches, and fires. *Journal of Causal Inference, 7*(2), 20190026.
Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.
Pearson, C. H. (2007). Is heritability explanatorily useful? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 38*(1), 270–288.
Read, J., & Sharma, S. (2021). Hypothesis-driven science in large-scale studies: The case of GWAS. *Biology and Philosophy, 36*(5), 46. https://doi.org/10.1007/s10539-021-09823-0
Rothman, K. J. (1976). Causes. *American Journal of Epidemiology, 104*(6), 587–592.
Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association, 100*(469), 322–331.
Sesardic, N. (2005). *Making sense of heritability*. Cambridge University Press.
Shipley, B. (2000). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference*. Cambridge University Press.
Sober, E. (2001). Separating nature and nurture. In D. Wasserman, & R. Wachbroit (Eds.), *Genetics and criminal behavior* (pp. 47–78). Cambridge University Press.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). The MIT Press.
Sterelny, K., & Griffiths, P. E. (1999). *Sex and death: An introduction to philosophy of biology*. University of Chicago Press.
Tabery, J. (2014). *Beyond versus: The struggle to define the interaction of nature and nurture*. The 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.
Tal, O., Kisdi, E., & Jablonka, E. (2010). Epigenetic contribution to covariance between relatives. *Genetics, 184*(4), 1037–1050.
Taylor, P. (2006). Commentary: The analysis of variance is an analysis of causes (of a very circumscribed kind). *International Journal of Epidemiology, 35*(3), 527–531.
Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics, 90*(1), 7–24.
Waters, C. K. (2007). Causes that make a difference. *The Journal of Philosophy, 104*(11), 551–579.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*(7), 557–585.