Behavior Genetic Frameworks of Causal Reasoning for Personality Psychology

Supplemental Appendices

Appendix A. Some Philosophical Remarks on Causality

In this appendix, we want to elaborate on some of the points treated too briefly in the main text. We have divided the material here into three sections. The first section discusses some of the most abstract metaphysical problems regarding causation. The second section discusses the problem of the reduction basis for the causal relation. The third section discusses two problems of actual causation.

A.1. Some Fundamentals

In the main text, we have assumed that causation is a relation. But there are some reasons for thinking that causation is not a relation at all.[1] To begin to appreciate the question, start with what is probably the orthodox philosophical view today: causation is a relation between events. Add to this the also pretty orthodox view that there are no *negative* events. That is to say, absences and omissions are not events properly so called at all. Now consider the following cases:

[1] Sally's alarm clock didn't go off, and that caused her to be late for work.

[2] Sally contracted flu, and that caused her to be absent from the meeting.

[3] Sally caused a problem at work when she prevented Lauren the security guard from keeping Bob (a known nuisance for the company) out of the building.

[1] For elaboration and discussion of this issue, see Lewis (2004), Beebee (2004), and Livengood and Machery (2007).

In each of these cases, the causal claim looks reasonable at first blush. But if causation is a relation between events and there are no negative events, then none of these claims can be literally true.

Here are three broad lines of response that have been explored in the philosophical literature. First, one might simply deny that causation is a relation. Perhaps it is better understood as a sentential connective or something similar.[2] Second, one might assert that causation is a relation and deny that [1], [2], and [3] are really examples of causation. Perhaps they are merely close relatives of causation. Defending such a view might involve elaborating an error theory (explaining how so many of our causal attributions are false) or adopting a non-literal reading of ordinary causal language. Third, one might argue that absences and omissions are admissible relata for the causal relation after all. For example, one might defend the existence of negative events and assert that negative events can stand in causal relations as causes, effects, or both. Alternatively, one might argue that causation is not a relation *between events* but rather a relation between facts, propositions, properties, or the like and that if so, absences and omissions are unproblematic as relata.

Going forward, let us suppose that causation is a relation. What are its formal properties? The orthodox view seems to be that causation is necessarily an irreflexive, asymmetric, and transitive two-place relation.[3] However, one might have reasons for dissent. Consider the following case derived from an example given by Good (1961) and subsequently discussed and modified by Hitchcock (1995, 2001) and Hall (2004):

[2] For an example of this approach, see Mellor (2004).

[3] To say that a relation R is irreflexive is to say that no individual x is R-related to itself. To say that a relation R is asymmetric is to say that for any individuals x and y, if x is R-related to y, then y is not R-related to x. To say that a relation R is transitive is to say that for any individuals x, y, and z, if x is R-related to y and y is R-related to z, then x is R-related to z.

> While Sherlock Holmes is hiking in the mountains, his enemy Professor Moriarty pushes a boulder so that it rolls down a hill straight at where Holmes is standing. The boulder causes quite a lot of noise as it careens down the hill. Hearing the noise, Holmes looks around and sees that he is in danger. Luckily, he has time to dodge the boulder and successfully does so. Hence, Holmes survives.

In the Boulder case, as it has come to be called, it seems right to say that Moriarty causes the boulder to roll down the hill towards Holmes. It also seems right to say both that the boulder rolling down the hill causes Holmes to dodge out of the way and also that Holmes dodging out of the way causes him to survive. But it doesn't seem right to say that Moriarty causing the boulder to roll down the hill towards Holmes causes Holmes to survive. Hence, causation is not transitive.

Challenges to irreflexivity and asymmetry are less mundane. Suppose, for example, that one takes seriously the possibility of time travel. If time travel is possible, then predestination paradoxes, such as the one featured in the original *Terminator* movie, are possible. If causation is transitive, then it is neither irreflexive nor asymmetric. To see why this is the case, recall that in the *Terminator*, John Connor sends Kyle Reese to the past to protect Sarah Connor from a terminator robot. Sarah falls in love with Kyle, and by the end of the movie, she is pregnant with his child: John Connor. Assuming that causation is transitive, Sarah's falling in love with Kyle causes John to send Kyle to the past, which causes Sarah to fall in love with Kyle. Hence, causation is not asymmetric. And if causation is transitive, we can close the loop even more to say that Sarah's falling in love with Kyle causes Sarah to fall in love with Kyle. Hence, causation is not irreflexive either.

Alternatively, one might give up on transitivity as a way of maintaining that causation is irreflexive and asymmetric while accepting that time travel is possible.[4]

In the main text, we described causation (sometimes called *structural* causation for reasons described in more detail in part three of this appendix) as a two-place relation between properties, which may be represented by random variables. But taking causation in that sense, one might have a very different reason for skepticism about irreflexivity, asymmetry, and transitivity. One might suppose, for example, that being more conscientious causes one to exercise frequently and that exercising frequently causes one to be more conscientious. We might represent this as *Conscientiousness* causing *Exercise* and *Exercise* causing *Conscientiousness*. But if that's right, then causation is not asymmetric. Similarly, suppose that being conscientious early in life causes people to be conscientious later in life as well. We might represent this as *Conscientiousness* causing itself. But if that's right, then causation is not irreflexive. Insofar as *direct* structural causation is what we are talking about, causation is not transitive either. For example, *Conscientiousness* might directly cause *Exercise* and *Exercise* might directly cause *Health* without *Conscientiousness* directly causing *Health*.

Now, the orthodox view in philosophy is that causation is an irreflexive, asymmetric, and transitive two-place relation between *event tokens*. An event token is a particular occurrence, such as Sally's drinking a glass of milk at some specific time. Just how much detail needs to go into specifying an event token—how fine-grained one's description needs to be—is a difficult question (see Kim, 1973; Lewis, 1986; and Paul, 2004). But however much detail we require, we

[4] You may think that time travel is too implausible to take seriously or that while time travel is physically *possible*, there are good reasons to think that it doesn't occur in the actual universe. If so, then what is the point of thinking about time travel cases? The answer has to do with what the target of philosophical investigations into the nature of causation is. Typically, philosophers are interested in the properties that a relation *must have* in order to count as the causal relation. But the "must" there refers not just to the actual world but to all possible worlds. If there is a way things could be where there is a relation that counts as the causal relation and that relation is not transitive, then the causal relation is not transitive.

want to distinguish between a particular occurrence and the *types* or *classes* to which that particular occurrence belongs. Sally's drinking a glass of milk when and how she does it and Jacob's drinking a glass of milk when and how he does it are distinct occurrences, but they belong to the same type: milk drinkings. The parallel idea for properties (or the variables they represent) is the relationship between property *instances* and property *universals*. For both events and properties, maintaining that causation is irreflexive, asymmetric, and transitive is much more plausible for tokens (or instances) than it is for types (or universals).

A further question about the formal character of the causal relation is how many relata it has. The orthodox answer here is that the causal relation has two relata: the cause and the effect. We endorse the orthodoxy, but a strong case has been made that the causal relation has *four* relata. On such a view, causation is contrastive. True causal claims have the following form: *c* as opposed to *d* causes *e* as opposed to *f*. It is only in cases where the relevant contrasts are clear that it would make sense to say simply that *c* causes *e*. Schaffer (2005) argues that a contrastive account of causation on which the causal relation has four relata resolves a collection of stubborn paradoxes of causation.[5]

A.2. Reductionism about Causation

Questions about the formal properties of the causal relation have received relatively little treatment in the philosophical literature. Instead, most work on causation has focused on a *material* question: what exactly *is* the causal relation. As Schaffer (2016) puts it, "What is the metaphysical basis for causal connection? That is, what is the difference between causally related and causally unrelated sequences?" On the material question, philosophers have largely agreed that the causal relation should be reduced to *something*, but there is essentially no agreement as

[5] For a different take on the arguments and further discussion, see Schaffer (2016).

to the correct reduction basis. A few prominent examples should suffice to illustrate.[6] On one venerable reading, Hume (1748) reduces causation to regular succession. Psillos (2009) articulates the Humean account as follows: an event *c* belonging to the type *C* is a cause of an event *e* belonging to the type *E* if and only if (1) events of type *C* are constantly conjoined with events of type *E*, (2) *c* is temporally prior to *e*, and (3) *c* is spatiotemporally contiguous with *e*.

Several other classes of reductive theory may be thought of as broadly Humean in spirit. For example, Davidson (1967) reduces causation to laws of nature by maintaining that an event *c* is a cause of an event *e* if and only if there is a covering law and a matter of fact that together entail that *c* causes *e*. So long as the laws of nature are understood as non-governing descriptions of events (so-called Humean laws), accounts that reduce causation to laws of nature may be thought of as simply clarifying the Humean idea of constant conjunction.[7] Another broadly Humean alternative is to weaken the idea of constant conjunction to some kind of chancy conjunction. Along these lines, Good (1961) and Suppes (1970) propose to reduce causation to facts about probability. The core idea for probability theories is that *c* is a cause of *e* if and only if the probability that *e* occurs given that *c* occurs is greater than the probability that *e* occurs given that *c* does not occur. The basic probabilistic proposal is clearly inadequate, since if there is a common cause *d* of *c* and *e*, the probability of *e* given that *c* occurs will be greater than the probability of *e* given that *c* does not occur. However, the basic idea may be patched up. In order to give readers unfamiliar with the philosophical literature a sense of the complications that need to be added, we will here describe a more sophisticated probabilistic theory of causation

[6] Mostly, we will stick to simple, early versions of the various attempts to reduce causation. Every theory needs to be greatly elaborated (and has been in the philosophical literature) in order to handle various difficulties.
[7] Beebee (2000) contrasts governing and non-governing accounts and defends the claim that laws of nature are non-governing. If one takes laws of nature to be *governing*, then reducing causation to laws of nature will not satisfy Humean scruples. Maudlin (2007) might be read as reducing causation to governing laws. As he puts it, "The 'necessary connexion' that Hume sought at the heart of causation is nomic necessity" (145).

defended by Glynn (2011), which resists obvious counter-examples. Let C be a variable that takes the value 1 if the event *c* occurs and that takes the value 0 if the event *c* does not occur. And let E be a variable similarly defined with respect to the event *e*. Moreover, if $\boldsymbol{V}$ is a set of variables, let $\boldsymbol{V}^*$ be a proposition asserting that the variables in $\boldsymbol{V}$ take their actual values. Stripped down and simplified slightly, according to Glynn's account, an event *c* is a cause of an event *e* if and only if there is a set $\boldsymbol{S}$ of variables such that

(1) $\Pr(E = 1 \mid C = 1, \boldsymbol{S}^*) > \Pr(E = 1 \mid C = 0, \boldsymbol{S}^*)$, and

(2) for all sets $\boldsymbol{T}$, either

(i) there is a set $\boldsymbol{U}$ such that

$\Pr(E = 1 \mid C = 1, \boldsymbol{T}^*, \boldsymbol{U}^*) > \Pr(E = 1 \mid C = 0, \boldsymbol{T}^*, \boldsymbol{U}^*)$, or

(ii) there is a variable D in $\boldsymbol{T}$ such that

(a) there is a set $\boldsymbol{V}$ such that

$\Pr(D = 1 \mid C = 1, \boldsymbol{V}^*) > \Pr(D = 1 \mid C = 0, \boldsymbol{V}^*)$, and

(b) there is a set $\boldsymbol{W}$ such that

$\Pr(E = 1 \mid D = 1, \boldsymbol{W}^*) > \Pr(E = 1 \mid D = 0, \boldsymbol{W}^*)$,

where no variable in the set $\boldsymbol{S}$ may represent any event occurring later than *e* or any event that is not "reasonably natural," which last condition is meant to exclude disjunctive events and other *ad hoc* constructions.

In his *Enquiry*, Hume writes (76), "We may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*." Hume's definition naturally gives rise to the attempts to reduce causation to patterns of regular succession, laws of nature, and probabilities. But immediately after giving his definition, Hume offers the following gloss on what he has just said: "Or in other words *where, if the first*

object had not been, the second never had existed." Moved by Hume's "other words," Lewis (1973) attempted to reduce causation to counterfactual dependence, and many (perhaps most) contemporary metaphysicians have followed Lewis in thinking that causation reduces to counterfactual dependence. The core idea for counterfactual theories of causation is that *e* causally depends on *c* if and only if the following two counterfactual conditionals are both true: (1) if *c* were to occur, then *e* would occur, and (2) if *c* were not to occur, then *e* would not occur. Suzy was born with an extra X chromosome in each of her cells, and she is taller than one would predict by looking at her parents. If Suzy had not been born with an extra X chromosome, she would not have been taller than one would predict by looking at her parents. Hence, Suzy's being taller than expected causally depends on her being born with an extra X chromosome. In order to secure some desirable formal properties for causation, such as being transitive, Lewis defines causation as the ancestral of causal dependence. He writes (563), "Let *c*, *d*, *e*, … be a finite sequence of actual particular events such that *d* depends causally on *c*, *e* on *d*, and so on throughout. Then this sequence is a *causal chain*. Finally, one event is a *cause* of another iff there exists a causal chain leading from the first to the second."[8]

Not all attempts to reduce causation to something else are Humean in spirit. For example, Menzies and Price (1993) reduce causation to agency. They write (with the letters denoting events changed to be consistent with our usage here), "The central thesis of an agency account of causation is something like this: an event *c* is a cause of a distinct event *e* just in case bringing about the occurrence of *c* would be an effective means by which a free agent could bring about the occurrence of *e*" (187). Agency theories are actually reductive only to the extent that the

[8] As with probabilistic theories of causation, the simple account has obvious counter-examples. But it can be and has been patched up. See Lewis (2000) for a significantly updated version of his basic idea. Paul and Hall (2013) discuss the counterfactual theory of causation and its difficulties in great detail.

power of the agent to bring things about is both more primitive than causation and also itself not causal in character.[9] Of course, agents are not the only entities that have dispositions or capacities to act, and agents are not the only entities that have power to bring things about. Generalizing from agents to objects, we come to a *dispositional-powers* theory. The core idea for dispositional-powers theories is that causation is the enactment of a dispositional power belonging to an object. Paraphrasing Harré and Madden (1973, 221), the dispositional-powers theory says that causation is a relation between an object having a dispositional power and the manifestation of that dispositional power in appropriate triggering circumstances. The causal relation obtains when such a dispositional power is expressed.

Two further accounts try to reduce causation to something obviously scientifically respectable. The first of these tries to reduce causation to the operations of a mechanism. As Glennan (1996) puts it, "Two events are causally connected when and only when there is a mechanism connecting them" (64). Precisely what makes something a mechanism is not especially clear, but as a placeholder, let us suppose that a mechanism is an organized collection of parts that behave in a regular way.[10] Then according to the mechanist theory of causation, a cause and an effect are two arrangements of a mechanism such that if the mechanism is in the cause arrangement at some time, then it will—by behaving in its regular way—come to be in the effect arrangement at some later time. The operation of the mechanism in transitioning from the cause configuration to the effect configuration is causation. The mechanism account was developed by reflecting on how explanation works in the life sciences, and it has its natural home

[9] We think there is something appealing about the agency theory as a story about how we learn about causal relations. That is, we do seem to have a relatively primitive sense of the difference between observing some event and bringing it about. (Though the psychological details are complicated by, among other things, facts about causal perception, as opposed to causal inference. For an introduction, see Danks, 2009.) But we are not convinced that the agency theory succeeds as a reductive metaphysical account.

[10] In addition to Glennan (1996), see Bechtel and Abrahamsen (2005), Machamer, Darden, and Craver (2000), and Tabery (2004) for further discussion of mechanism.

there. By contrast, Dowe (1995) and Salmon (1997) propose reducing causation in a way that is better suited to physics. Specifically, they try to reduce causation to the transfer of a conserved quantity. Salmon (1997) modifies his earlier account and defends the following three definitions:

> DEFINITION 1: A causal interaction is an intersection of world-lines that involves exchange of a conserved quantity. (468)
>
> DEFINITION 2: A causal process is the world-line of an object that transmits a non-zero amount of a conserved quantity at each moment of its history (each spacetime point of its trajectory).
>
> DEFINITION 3: A process transmits a conserved quantity between A and B ($A \neq B$) if and only if it possesses [a fixed amount of] this quantity at A and at B and at every stage of the process between A and B without any interactions in the open interval (A, B) that involve an exchange of that particular conserved quantity. (462)

Both the mechanist theory and the conserved quantity theory have much to recommend them, but both have serious weaknesses with respect to the relationship between causation and time. Each theory is either silent about the temporal direction of the causal relation or has to build in the relation by hand.

How should one react to this whimsical cavalcade of reductive theories? Some philosophers have responded with skepticism about the metaphysical or scientific credentials of causation. Perhaps causation has a useful role to play in a folk theory, but it does not belong in scientific ontology.[11] Other philosophers have argued that there is simply no good reduction basis for the causal relation. Taylor (1966, 40, quoted in Carroll 2009, 284) puts it elegantly when he writes, "To say of anything, then, that it was the cause of something else, means simply and solely that it was the cause of the thing in question, and there is absolutely no other conceptually clearer way of putting the matter except by the introduction of near synonyms for causation." More recently, several philosophers and allied researchers have offered non-

[11] See Skyrms (1984) and Norton (2003) for two relatively recent examples of this brand of skepticism about causation.

reductive accounts of causation that nonetheless place informative constraints on how causation is related to probability, intervention, counterfactual dependence, and the like. The most important early example of this approach is Cartwright (1979), who defends the following thesis: "Causal principles cannot be reduced to laws of association; but they cannot be done away with" (419). She goes on to develop a non-reductive but informative relationship between causation and probability. Cartwright's key idea is that a cause raises the probability of its effects in every situation that is (otherwise) causally homogeneous with respect to the effect. Let $\{c_i\}$, $1 \leq i \leq n$, be the set of events that are causally relevant to some event e. In other words, each c_i is such that either it causes e or it causes $\neg e$. Let k_j be the j^{th} conjunction of length n formed by conjoining distinct elements of $\{c_i\}$ or their negations, e.g. $k_1 = c_1 \wedge c_2 \wedge \ldots \wedge c_n$, $k_2 = \neg c_1 \wedge c_2 \wedge \ldots \wedge c_n$, $k_3 = c_1 \wedge \neg c_2 \wedge \ldots \wedge c_n$, and so on until all 2^n conjunctions of length n have been listed. Cartwright then sets out necessary and sufficient conditions for the causal relation to hold (423). An event c is a cause of an event e, denoted $c \rightarrow e$, iff $\Pr(e \mid c \wedge k_j) > \Pr(e \mid k_j)$ for all k_j over $\{c_i\}$, where $\{c_i\}$ satisfies the following four conditions:

(i) If c_i is an element of $\{c_i\}$, then either $c_i \rightarrow e$ or $c_i \rightarrow \neg e$.

(ii) The event c is not in $\{c_i\}$.

(iii) For all d, if $d \rightarrow e$ or $d \rightarrow \neg e$, then either $d = c$ or d is in $\{c_i\}$.

(iv) If c_i is in $\{c_i\}$, then it is not the case that $c \rightarrow c_i$.

As Cartwright observes, the necessary and sufficient conditions she sets out do not provide an analysis of causation, since the causal schema $c \rightarrow e$ appears on both sides of the equivalence. "It does, however, impose mutual constraints, so that given sets of causal and associational laws cannot be arbitrarily conjoined" (424).

Our approach to causation is similar to that taken by Pearl (2000), Spirtes, Glymour, and Scheines (2000), Woodward (2003), and other researchers working in the graphical causal modeling tradition. We take the term "causation" to refer a relation between two property universals (such as *Height* and *Weight*). We use random variables to represent property universals, and we use arrows to represent direct structural causal relations.[12] Hence, a directed graph over a set of random variables represents a causal structure with respect to that collection of property universals. Since we take structural causation to be a primitive relation, we do not think that it can be reduced to probability or counterfactual dependence or the like. But like Cartwright, we think there is an interesting non-reductive account to be given. We all have a naïve, pre-theoretical grasp of the difference between *seeing* that something is the case and *making* it to be the case. Following Pearl (2000), we represent the difference mathematically using the $do(\cdot)$ operator. When we merely *observe* that some individual u has the specific value x for property X, we write $X(\mathrm{u}) = x$. When we *act* in order to *set* the value of X to x for individual u, we write $do(X(\mathrm{u}) = x)$. If $\boldsymbol{Z}$ is an ordered set of variables, $\boldsymbol{z}$ is an ordered set of values, and the length of $\boldsymbol{Z}$ is the same as the length of $\boldsymbol{z}$, then $do(\boldsymbol{Z}(\mathrm{u}) = \boldsymbol{z})$ represents acting to set the values of the variables in $\boldsymbol{Z}$ to the corresponding values in $\boldsymbol{z}$. Acting on an individual to set the values of variables representing one or more of its properties is to *intervene on* the individual.[13] With the $do(\cdot)$ operator in hand, we can offer a non-reductive definition of structural causation for random variables and by extension for the properties they represent. Specifically, say that X is a *direct structural cause* of Y relative to some background $\boldsymbol{Z}$ if and only if there exists a test pair of

[12] While we prefer taking property universals as the relata of the structural causal relation, it might turn out that the same mathematical machinery may be used on the assumption that the relata are event types. If so, then one's choice of relata is mostly an aid to imagination: like Maxwell's (1855, 1861, and 1864) physical models of the ether or what Hardy (1929, 18) called *gas*. What ultimately matters is the mathematical machinery and the work that machinery does for us.

[13] As a convenient shorthand, we sometimes say that we intervene on a variable.

values $\langle x_1, x_2 \rangle$ for X such that $x_1 \neq x_2$ and a collection of values $\boldsymbol{z}$ for the variables in $\boldsymbol{Z}$ such that the following inequality holds:

$$\Pr(Y = y \mid do(X = x_1, \boldsymbol{Z} = \boldsymbol{z})) \neq \Pr(Y = y \mid do(X = x_2, \boldsymbol{Z} = \boldsymbol{z})).$$[14]

To be clear, our definition is not reductive, since we take the do(·) operator to have causal content. Hence, we are not offering a probabilistic agency account of causation. The causal relation is primitive. Probability raising, counterfactual dependence, manipulability, and the like are *marks* or *signs* of the presence of causal relations.

A.3. The Problems of Actual Causation

As we remarked at the beginning of this appendix, according to the orthodox view, causation is a relation between events. But on our account, causation is a relation between property universals. We think this is all to the good, since we take it that scientific inquiry typically aims to discover relations between property universals, as opposed to relations between particular property instances or particular events of property instantiation. That is, scientific inquiry typically aims at finding out, for example, whether smoking causes cancer, not whether Joe's smoking caused Joe's cancer. However, in many practical, everyday cases—in cases where we want to explain what happened or in cases where we want to assign blame—we do want to know something about how particular property instances are related.

Let's introduce some more terminology. Whereas structural causation is a relation between two property universals, *actual causation* is a relation between two property instances (such as *Joe's height* and *Joe's weight*). We use evaluated random variables to mathematically represent property instances. Actual causation has no special graphical representation. Features

[14] To be absolutely clear, we are not here offering an account like that defended by Menzies and Price (1993), which attempts to reduce causation to agency. Rather, we take our definition to be non-reductive, since the idea of the *do*(·) operator has causal content. As in Cartwright's account, causal notions appear on both sides of the equivalence.

of the graph representing the structural causes over a set of random variables represent some constraints on what actual causal relations obtain, but it is important to notice that the actual causes in a given case are not always equivalent to the structural causes in that case.[15] Here is an example. Diet is a structural cause of intelligence, since in the case where one has the PAH gene and suffers from phenylketonuria (PKU), the probability of having normal-range intelligence is much greater when one has a diet that excludes phenylalanine than when one has a diet that includes phenylalanine. But for a person who does *not* have the PAH gene but has some other genetic cause of lower-than-normal-range intelligence, such as Fragile X syndrome, eating a diet rich in phenylalanine is not an *actual* cause of having lower-than-normal-range intelligence.

Some philosophers thinking about actual causation have argued that we should distinguish between causes properly so called and background conditions.[16] These philosophers point out that although every effect depends on many different things, only some of those things count as actual causes. Often only *one* thing counts as an actual cause of a given effect. For example, having the PAH gene is (plausibly) *the* actual cause of having PKU, even though one would not develop PKU without many other genes doing the normal things that they do—producing skin and bones and heart tissue and so on. The PAH gene is the actual cause of PKU. The other genes are not actual causes of PKU: they are mere background conditions.[17] Philosophers who think that the PAH gene is the actual cause of PKU face the problem, called the *selection problem*, of distinguishing between things to promote to the status of actual causes and things to relegate to the background. Other philosophers have argued that we ought to give

[15] See Hausman (2005) for interesting discussion of the relationship between actual causation and structural causation. See Halpern (2016) for extended discussion of actual causation in the graphical causal modeling tradition.
[16] See Mackie (1965).
[17] Hitchcock (2007), Hitchcock and Knobe (2009), and Halpern and Hitchcock (2015) attempt to distinguish between causes and background conditions by appealing to judgments about normality. Waters (2007) attempts to draw the distinction by focusing on what he calls *actual difference making*.

an egalitarian account of causation according to which there is no theoretically interesting distinction to be drawn between causes and background conditions. We might draw such a distinction for practical reasons—and clearly, such a distinction is drawn in ordinary language. But a *philosophical* account of causation should not pay attention to such things.

Hence, there are two related problems, which might both legitimately be called problems of actual causation. The first problem is to identify the actual causes according to an egalitarian theory that does not distinguish between causes and background conditions. Call this the *actual causation learning problem.* As we have seen, actual causes are different than structural causes. If $X = x$ and $Y = y$ and X is a structural cause of Y, it still might be the case that $X = x$ is not an actual cause of $Y = y$. However, some philosophers have attempted to provide theories of actual causation that are in an important sense structural. They are structural in the sense that the actual causes of a given property instance may be determined by looking at the relevant causal structure together with facts about the results of various possible interventions on the structure. To illustrate, consider Woodward's (2003) account, which says that $C = c$ is an actual cause of $E = e$ if and only if the following two conditions both hold:

(W1) The actual value of C is c, and the actual value of E is e.

(W2) There exists a path P from C to E, and there exist manipulations $do(C = c^*)$ for $c^* \neq c$ and $do(\boldsymbol{W}=\boldsymbol{w})$ for $\boldsymbol{w}$ in the redundancy range of P such that if the variables in $\boldsymbol{W}$ were set to the values $\boldsymbol{w}$ and C were set to the value c^*, then E would not have the value e.

Where $\boldsymbol{w}$ is in the redundancy range of a path P if and only if carrying out the manipulations denoted by $do(\boldsymbol{W} = \boldsymbol{w})$ leaves all of the variables on P at their actual values.

The second problem is the selection problem. After the egalitarian actual causation learning problem is solved, we often want to identify one or a small number of the (egalitarian)

actual causes as *the* actual cause. In other words, we want to sort the actual causes into at least two groups: genuine causes and background conditions. Solving the selection problem has enormous practical significance. Let's illustrate with a personality example. Suppose Christina has a personality that everyone would regard as being in the normal range. One day, Christina finds herself in an environment that is very unusual: she is abducted by terrorists who threaten to harm her family if she does not cooperate with them. She cooperates and helps to carry out a terrorist attack. If the environment had been typical, Christina would not have helped to carry out a terrorist attack. Moreover, if Christina had been unusually stubborn or unusually loyal or unusually self-sacrificing or something similar, then she would not have cooperated. Hence, we have reason to think that both the unusual environment and Christina's normal personality are (egalitarian) actual causes of her helping to carry out a terrorist attack. However, we predict that if we were to put the question to naïve participants, they would say that Christina's environment, but not her personality, was the cause of her helping to carry out a terrorist attack.[18] By contrast, suppose that Felicity has a personality that everyone would regard as being outside the normal range. For concreteness, suppose she is extremely generous and gullible. One day, Felicity finds herself in an environment that is not very unusual: she is approached by a homeless person as she walks around in a busy city. Felicity responds to the homeless person by taking him with her to the bank and transferring ownership of her house. If Felicity's personality had been normal, she would not have given a homeless person her house. And if the environment had been less typical in specific ways, she would not have given away her house. Again, we predict that if we were to put the question to naïve participants, they would say that Felicity's personality, but not her

[18] Our prediction here is based on a large body of research on the role of norms and responsibility in causal attribution. See Hitchcock and Knobe (2009), Sytsma, Livengood, and Rose (2012), Zultan, Gerstenberg, and Lagnado (2012), Gerstenberg, Halpern, and Tenenbaum (2015), Kominsky et al. (2015), Livengood and Rose (2015), Livengood, Sytsma, and Rose (2017), and Icard, Kominsky, and Knobe (2017).

environment, was the cause of her giving away her house. These judgments are relative to various facts about what is normal or what bears responsibility for the outcomes of interest or something similar.

One might worry that normality, responsibility, and the like are too subjective or too sensitive to our goals or context to admit any principled characterization. Hence, one might worry that solving the selection problem has no value. However, the selection problem is the primary causal reasoning problem for attorneys and judges, politicians, historians, and diagnosticians of all sorts. So, solving the selection problem would seem to have very significant practical value. One might still wonder, though, what exactly diagnosticians gain from solving the selection problem. What is the point of sorting the (egalitarian) actual causes into causes and conditions? Our suggestion is that solving the selection problem earns its keep in settings where we want to intervene in big, complicated, messy causal systems that we do not thoroughly understand. Since we know how things work when everything is "normal" (acting as it typically does or as it is supposed to do), when something breaks, we identify things that are "abnormal" and try to set or reset them to be normal so that we don't accidentally break something else in virtue of fixing what has broken. Insofar as one is setting things back to normal, one does not need to know how the system works in order to be confident that one's intervention is safe.

Appendix A.4. Nuances of Causal Effect Learning

In a causal effect learning problem, we want to know what values a response variable would (or will) take on if some predictor variable were (to be set) one way as opposed to another. For example, we want to know whether it is true that if an individual were more conscientious, she would also be more health conscious. Notice that the modal claim that "If an individual were more conscientious, then she would be more health conscious" is distinct from the non-modal claim that "If an individual is more conscientious, then she is more health conscious." The latter is

a straightforward associational claim. The former may be inferred from the latter in some, but not all, cases. Causal effect learning problems are contrastive in character. For philosophical discussion of causation as contrastive, see Schaffer (2005). One natural way to understand that contrastive character is in terms of experimental treatment and control.

Many researchers trained in the potential outcomes tradition not only emphasize random assignment as the best method for establishing causation, they think that if no experimental intervention is possible, causal claims are simply nonsense.[19] If we cannot, even in principle, assign individuals to have high or low conscientiousness, then researchers who share Holland's philosophy will deny that conscientiousness causes health behavior.[20] Or anything else for that matter. But if intervening on conscientiousness makes sense in principle (even if it will never be practically feasible), then we might, for example, try to solve the causal effect learning problem by using propensity scores to create matched groups, as in Rosenbaum and Rubin (1985). However, unless one creates a propensity score using full genotypic and family environmental data (i.e., identical twins), the matching will be incomplete, reducing confidence in the causal claim.

Appendix B. Technical Details of Behavior Genetic Methodology

In this appendix, we provide additional technical details concerning behavior genetic methodology. These details, although important and critical for properly implementing the statistical approaches, are less germane to the main arguments of the manuscript. Given the mismatch between the availability of genetically informative data and the relative prevalence in personality psychology compared to non-genetically informative studies, we encourage readers

[19] See Section 7 in Holland (1986) for discussion of this point. See Glymour and Glymour (2014) for a recent critical appraisal of the typical attitude among researchers in the potential outcomes framework.
[20] To be clear, we agree with Pearl (2001), Glymour (2004), and Glymour and Glymour (2014) in rejecting Holland's philosophy.

to pursue these methods. In particular, forming collaborative relationships with current behavior geneticists may prove the most beneficial in wading through the underlying assumptions and somewhat daunting data management issues in behavior genetics – problems that are not overwhelming with a little guidance.

B.1. Behavior Genetic Model Assumptions

Twin and family models may be biased by violations of the equal environments assumption and the presence of assortative mating. First, the models assume that there is no assortative mating for the phenotypes of interest, meaning individuals in the population produce offspring with partners not selected for the variables under study. If there is assortative mating, then parents possess correlated genotypes and therefore transmit correlated alleles. Fraternal twins would then be more genetically similar than expected, in particular at loci that are relevant for the phenotype of interest. Assortative mating is common for most psychological phenotypes, ranging from quite small associations for personality ($r \sim .1$), to moderate associations for physical or medical phenotypes ($r \sim .3$) to quite large for social attitudes or educational attainment ($r > .6$; D'Onofrio et al., 1999; Eaves et al., 1999). However, these phenotypic estimates of spousal similarity may or may not be reflected at the genotypic level (Robinson et al., 2017), and correlation at the genotypic level is the primary concern. The consequence of assortative mating is that heritability is underestimated because fraternal twins are actually more genetically related (and therefore phenotypically similar) than specified (i.e., 50% related), although the amount of bias tends to be relatively small (Keller et al., 2010; Keller et al., 2013). If the level of assortative mating is known (e.g., Conley et al., 2016), then the assumed genetic relatedness between siblings can be corrected to produce unbiased estimates. For designs other than the classical twin design or similar sibling-based designs, the interpretation is different. For

example, assortative mating would lead to increased estimates of heritability in parent-offspring studies because the pair share more than 50% of segregating genetic material. Correlations between adoptive parents or siblings and a focal child should not be affected by assortative mating.

Second, the equal environments assumption may be violated if identical twins are treated more similarly than fraternal twins simply due to their zygosity status (i.e., being an identical rather than fraternal twin), but only if such treatment alters the phenotype of interest. This problem would emerge if parents systematically instilled similarity in their identical twins because of beliefs about what it means to be an identical twin, as one example. Parents may dress identical twins more similarly, but unless dress in childhood impacts the phenotype of interest, the assumption is not violated. Good empirical evidence supports the equal environment assumption (Conley, Rauscher, Dawes, Magnusson, & Siegal, 2013; Derks, Dolan, & Boomsma, 2006; but see also Marceau et al., 2016). A similar complication could arise due to sibling contrast effects, whereby the behavior of one sibling affects the other (Carey, 1986). Opposite-sex fraternal twins may share less psychological similarity due to sex roles. For this reason, participant sex (and several other variables) should almost always be treated as a covariate (McGue & Bouchard, 1984).

In the molecular genetic context, population stratification refers to a confound whereby any cultural difference across groups of individuals is necessarily associated with any genetic variants that differ in frequency across groups. For example, Hamer and Sirota (2000) warned of the “chopstick gene.” If one were to run a GWAS on chopstick ability including participants from cultures that used chopsticks and those that did not, then any variant that differs in frequency across these populations (for example, those related to hair and eye color) would

appear as a potential causal factor in chopstick use. Of course, the difference is cultural, not genetic.

Solutions to population stratification rely on narrowing the sampling frame and robustness checks. GWAS are almost exclusively run on ancestrally homogenous groups, most commonly White individuals of European ancestry. Even when subsetting the sample based solely on individuals that share some ancestral background, as many as 20 principal components are extracted from the genetic data to use as control variables. Confounding due to genetic drift is apparent even within samples that would traditionally be considered homogeneous (Novembre et al., 2008). A more stringent robustness check is to identify whether genetic associations hold within a family. For example, Okbay, Beauchamp et al. (2016) used this approach to identify causal effects on educational attainment. Within a family, it is random whether one sibling receives an allele that is found to be positively associated with educational attainment in a GWAS compared to one that is negatively associated. Thus, if the sibling that was randomly assigned at conception to have the increasing allele has higher educational attainment compared to his or her co-sibling, then this is evidence for a causal effect of the allele on educational attainment. Because population stratification and all other sorts of between-family confounds are controlled and the direction of causation is extremely unlikely to flow in the opposite direction (i.e., attaining a college degree causes a genomic mutation), the identified association is most likely causal.

B.2. Examples of Gene-Environment Interplay

The impact of different forms of gene-environment interplay on family-based behavior genetic model parameters is predictable (Purcell, 2002). When genes and shared environments are correlated (i.e., passive gene-environment correlation), all family members will increase in

psychological similarity. Because all family members increase in similarity, shared environmental variance is observed. When genes and nonshared environments are correlated (i.e., evocative or active gene-environment correlation), siblings with greater genetic relatedness are more likely to experience similar nonshared environments compared to siblings with less genetic relatedness. At least two patterns are possible. Genetically related individuals may select similar environments which magnifies psychological similarity. Or, genetically less related individuals may select dissimilar environments, which in turn magnifies psychological dissimilarity. The net result of either of these situations is that genetically more related individuals are more psychologically similar than genetically less related individuals. As heritability is estimated via this difference, heritability is increased when genes and nonshared environments are correlated.

In molecular genetic designs, each form of gene-environment correlation magnifies SNP effect sizes due to reinforcement from the environment. For example, individuals with a certain allele may have elevated levels of extraversion and actively seek out or evoke socially stimulating experiences, which in turn increases extraversion. The result is a stronger personality-SNP association due to the environmental feedback effect. Similarly, parents may pass on a certain allele that confers elevated levels of depression, and because the parents also suffer from depression, the environment reinforces this phenotype (i.e., passive gene-environment correlation). This process results in a stronger personality-SNP association due to an environmental effect. Note that the implication of passive gene-environment correlation differs across twin and family studies (i.e., shared environmental variance) and molecular genetic studies (i.e., genetic variance). Kong et al. (2018) demonstrated a novel form of gene-environment correlation, genetic nurture. Using parental and offspring genotypes, they found that

even the non-transmitted parental alleles were associated with child development. This source of environmental variation (from the child's perspective) would also magnify personality-SNP associations (Koellinger & Harden, 2018).

Moving on to Gene × Environment interaction, similar implications can be drawn for family-based designs. When genes and a shared environment interact, more genetically related pairs will respond more similarly to the environment, producing genetic variance. There is some evidence, particularly in the United States, that genetic influences on cognitive ability differ across levels of family socioeconomic status (Tucker-Drob & Bates, 2016). Children from relatively more well-off families display greater heritability compared to their less well-off peers. If not explicitly specified, this developmental process would contribute genetic variance to cognitive ability. Identical twins will respond to the shared environment in a similar manner due to their identical genotypes, but fraternal twins will respond differently, exaggerating psychological dissimilarity.

When genes and a nonshared environment interact, phenotypic similarity for even genetically identical pairs is reduced, producing nonshared environmental variance. For example, Hicks et al. (2009) demonstrated that environmental adversity moderated genetic influences on externalizing. In risky environments, such as affiliating with anti-social peers, genetic influences on externalizing were substantially larger. Again, the important point here is not the specifics of this interaction, but rather the implications for estimating genetic and environmental influences on externalizing without this more detailed information. Under this circumstance, there are likely plenty of pairs of identical twins that are concordant for exposure to anti-social peers, meaning their similarity is not affected by the interaction. However, some pairs of identical twins will be

discordant for the environmental exposure and respond to their different environments as a function of their genotype, producing dissimilarity even in genetically identical individuals.

Molecular genetic designs push these sorts of aggregate questions forward by specifying which genetic variant interacts with the environment. It may be the case that the experience of psychological stress has a greater impact on depression for carriers of a certain allele. However, these sorts of candidate Gene × Environment interaction hypotheses are almost all unconfirmed due to low power (Duncan & Keller, 2011). For this reason, molecular genetic designs typically estimate the additive effect of each genetic variant. When Gene × Environment interaction occurs, regression parameters are affected by both the allele frequency and form of the interaction. If a perfect cross-over interaction occurs, such that the marginal effect of each allele is flat, then no main effect of the SNP will be observed in a standard main-effect GWAS (although a relatively large regression weight would be estimated for the interaction term, if it were included). In every other case of interaction, a marginal main effect will be observed in the main-effect GWAS in the absence of the interaction term being included in the model. Given the extreme number of potential interaction terms that can be tested, techniques have been introduced that balance considerations of potential forms of interactions with relative statistical power to prioritize SNPs that display a main effect (Hsu et al., 2012). In terms of GCTA, these techniques typically assume that all genetic effects are additive, although recent developments have begun to implement computational approaches to formally test moderation of heritability by exposure to different environments (Yang et al., 2011).

Appendix B.3. Common Co-Twin Control Interpretations of Null Results

In this section, we highlight some common examples of researchers interpreting null results from a co-twin control design as evidence for no causal relation. We note these examples

for their typicality in this area, not as special cases. For example, it was argued that a "failure to observe an effect within discordant [identical twin] pairs is inconsistent with causality" (McGue, Osler, & Christensen, 2010, pg. 549). Jackson et al. (2016) assessed whether the identical twin that smoked marijuana had slowed cognitive development compared to his or her non-smoking co-twin. Across two studies, cognitive development did not differ across discordant pairs. The authors concluded that this finding "implicates either genetic or family-wide environmental factors—not marijuana use—as the potential driver of differences in cognitive ability" (Jackson et al., 2016, pg. E506). As another example, Mosing, Madison, Pedersen, and Ullen (2016) tested whether the identical twin that practiced music more also had higher cognitive ability, but they did not find a difference. Their conclusion was that the "findings strongly suggest that associations between music practice and IQ in the general population are non-causal in nature" (pg. 504). Similarly, Eney, Tsang, Delaney, Turkheimer, and Duncan (2017) assessed whether the twin that reported drinking more sugary soda (over the past 4 weeks) had higher Body Mass Index (BMI). Although the effect was present in the full sample, the association was "greatly reduced and no longer significant within twin pairs, and the lack of association was due to genetic confounding" (pg. 8).

Appendix B.4. Promising Mendelian Randomization Extensions

Zhu et al. (2018) formalized Pickrell et al.'s (2016) insight that scatterplots of SNP effect sizes could inform causal direction and termed the approach generalized summary data-based Mendelian randomization. Their approach makes several improvements, such as increased power, a method to eliminate SNPs with pleiotropic effects (which violate the assumptions of the instrumental variable analysis), and a method to incorporate covariates to test for conditional causal relations. A major benefit of this approach compared to early examples of Mendelian

randomization is that individual-level data on the phenotypes are not required. GWAS summary statistics, which tend to be publicly available, are all that is required, in addition to a reference genotype panel, which again are commonly available.

Minică et al. (2017) proposed a merged form of the twin direction-of-causation models with Mendelian randomization. There are three main advantages of this approach. First, they propose using polygenic risk scores as the instrument in order to maximize the association between the instrument and the exposure. Polygenic risk scores sum up alleles across the genome based on GWAS summary statistics. If a specific allele is positively associated with a phenotype, then individuals with that allele would receive a positive score. This summation is carried out across the genome to produce a continuous index. Second, carrying out this design in the context of twins allows flexibility in the assumptions of Mendelian randomization. Typically, the instrument (e.g., the polygenic risk score) cannot have an effect on the outcome except through the exposure variable. By summing up effects across the genome, pleiotropic effects are almost guaranteed. Minică et al. (2017) demonstrate through simulation that their approach is unbiased in the presence of pleiotropy, but standard Mendelian randomization approaches are biased. In contrast, Zhu et al. (2018) remove any SNPs that have pleiotropic effects. Third, Minică et al. (2017) boast of much higher power than standard Mendelian randomization. Under some conditions, 2000 twin pairs are equivalent to over 50,000 unrelated individuals. However, there are some important assumptions. Most importantly, one must assume that nonshared environments that influence the exposure and the outcome are uncorrelated, except for the direct effect of the exposure on the outcome. Put differently, there can be no nonshared environmental common causes. The model does not have to make this assumption, but power is substantially

diminished without. Because this model requires different assumptions from standard Mendelian randomization, it offers an interesting complement.

References

Bechtel, W. & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36,* 421-441.

Beebee, H. (2000). The non-governing conception of laws of nature. *Philosophy and Phenomenological Research, 61*(3), 571-594.

Beebee, H. (2004). Causing and nothingness. In *Causation and counterfactuals,* 291-308.

Beebee, H., Hitchcock, C., & Menzies, P. (Eds.). (2009). *The Oxford handbook of causation*. Oxford University Press.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics, 16,* 319-341.

Carroll, J. (2009). Anti-reductionism. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 279-298). Oxford, Oxford University Press.

Cartwright, N. (1979). Causal laws and effective strategies. *Nous,* 419-437.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research, 3*, 507-554.

Collins, J. D., Hall, E. J., & Paul, L. A. (Eds.). (2004). *Causation and counterfactuals*. MIT Press.

Conley, D., Laidley, T., Belsky, D. W., Fletcher, J. M., Boardman, J. D., & Dominque, B. W. (2016). Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proceedings of the National Academy of Sciences, 113,* 6647-6652.

Conley, D., Rauscher, E., Dawes, C., Magnusson, P. K. E., & Siegal, M. L. (2013). Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins. *Behavior Genetics, 43,* 415-426.

D'Onofrio, B. M., Eaves, L. J., Murelle, L., Maes, H. H., & Spilka, B. (1999). Understanding biological and social influences on religious affiliation, attitudes, and behaviors: A behavior genetic perspective. *Journal of Personality, 67,* 953-984.

Danks, D. (2009). The psychology of causal perception and reasoning. *The Oxford handbook of causation,* 447-470.

Davidson, D. (1967). Causal relations. *The Journal of Philosophy, 64,* 691-703.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9,* 403-411.

Dowe, P. (1995). Causality and conserved quantities: A reply to Salmon. *Philosophy of Science, 62,* 321-333.

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry, 168,* 1041-1049.

Eaves, L., Heath, A., Martin, N., Maes, H., Neale, M., … & Corey, L. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Research, 2,* 62-80.

Eney, A. E., Tsang, S., Delaney, J. A., Turkheimer, E., & Duncan, G. E. (2017). Cross-sectional association between soda consumption and body mass index in a community-based sample of twins. *Nutrition Journal, 16,* 1-10.

Gerstenberg, T., Halpern, J. Y., & Tenenbaum, J. B. (2015). Responsibility judgments in voting scenarios. In *CogSci*.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*, 49-71.

Glymour, C. (2004). Review of Woodward's *Making Things Happen*. *The British Journal for the Philosophy of Science*, *55*, 779-790.

Glymour, C. (2010). What is right with 'Bayes net methods' and what is wrong with 'hunting causes and using them'? *The British Journal for the Philosophy of Science*, *61*, 161-211.

Glymour, C., & Glymour, M. R. (2014). Commentary: race and sex are causes. *Epidemiology*, *25*, 488-490.

Glynn, L. (2011). A probabilistic analysis of causation. *The British Journal for the Philosophy of Science*, *62*, 343-392.

Good, I. J. (1961). A causal calculus (I). *The British journal for the philosophy of science*, *11*(44), 305-318.

Good, I. J. (1961). A causal calculus (II). *The British journal for the philosophy of science*, *12*(45), 43-51.

Hall, N. (2004). Two concepts of causation. In *Causation and counterfactuals*, 225-276.

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*, *66*(2), 413-457.

Halpern, J. (2016). *Actual Causality*. MIT Press.

Hamer, D., & Sirota, L. (2000). Beware the chopsticks gene. *Molecular Psychiatry*, *5*, 11-13.

Hardy, G. H. (1929). Mathematical proof. *Mind*, *38*(149), 1-25.

Harré, R., & Madden, E. (1973). Natural powers and powerful natures. *Philosophy*, *48*(185), 209-230.

Hausman, D. M. (2005). Causal relata: Tokens, types, or variables? *Erkenntnis*, *63*(1), 33-54.

Hesslow, G. (1976). Two notes on the probabilistic approach to causality. *Philosophy of science*, *43*(2), 290-292.

Hitchcock, C. R. (1995). The mishap at Reichenbach fall: Singular vs. general causation. *Philosophical studies*, *78*(3), 257-291.

Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy*, *98*(6), 273-299.

Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *The Philosophical Review*, *116*(4), 495-532.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy*, *106*(11), 587-612.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945-960.

Hsu, L., Jiao, S., Dai, J. Y., Hutter, C., Peters, U., & Kooperberg, C. (2012). Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genetic Epidemiology*, *36*(3), 183-194.

Hume, D. (1748). *An enquiry concerning human understanding*.

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161, 80-93.

Jackson, N. J., Isen, J. D., Khoddam, R., Irons, D., Tuvblad, C., Iacono, W. G., … & Baker, L. A. (2016). Impact of adolescent marijuana use on intelligence: Results from two longitudinal studies. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, E500-E508.

Keller, M. C., Garver-Apgar, C. E., Wright, M. J., Martin, N. G., Corley, R. P., Stallings, M. C., Hewitt, J. K., & Zietsch, B. P. (2013). The genetic correlation between height and IQ: Shared genes or assortative mating? *PLOS Genetics, 9,* e1003451.

Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics, 40,* 377-393.

Koellinger, P. D., & Harden, K. P. (2018). Using nature to understand nurture. *Science, 359,* 386-387.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, 137, 196-209.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., … & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science, 359,* 424-428.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*, 556-567.

Lewis, D. (2000). Causation as influence. *The Journal of Philosophy*, *97*, 182-197.

Lewis, D. (2004). Void and object. In *Causation and counterfactuals*, 277-290.

Livengood, J., & Machery, E. (2007). The folk probably don't think what you think they think: Experiments on causation by absence. *Midwest Studies in Philosophy*, *31*(1), 107-127.

Livengood, J., & Rose, D. (2015). Experimental philosophy and causal attribution. *A companion to experimental philosophy*, 434-449.

Livengood, J., Sytsma, J., & Rose, D. (2017). Following the FAD: Folk attributions and theories of actual causation. *Review of Philosophy and Psychology*, 8(2), 273-294.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science, 67*(1), 1-25.

Mackie, J. L. (1965). Causes and conditions. *American philosophical quarterly*, *2*(4), 245-264.

Marceau, K., McMaster, M. T. B., Smith, T. F., Daams, J. G., van Beijsterveldt, C. E. M., Boomsma, D. I., & Knopik, V. S. (2016). The prenatal environment in twin studies: A review of chorionicity. *Behavior Genetics, 46,* 286-303.

Maudlin, T. (2007). *The metaphysics within physics*. Oxford University Press.

Maxwell, J. C. (1855). On Faraday's lines of force. In *The Scientific Papers of James Clerk Maxwell*, vol. 1, W.D. Niven (ed.), Cambridge University Press, 155-229.

Maxwell, J. C. (1861). On physical lines of force. In *The Scientific Papers of James Clerk Maxwell*, vol. 1, W.D. Niven (ed.), Cambridge University Press, 451-513.

Maxwell, J. C. (1864). A dynamical theory of the electromagnetic field. In *The Scientific Papers of James Clerk Maxwell*, vol. 1, W.D. Niven (ed.), Cambridge University Press, 526-597.

McGue, M., & Bouchard, T. J. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14,* 325-343.

McGue, M., Osler, M., & Christensen, K. (2010). Causal inference and observational research: The utility of twins. *Perspectives on Psychological Science, 5,* 546-556.

Menzies, P., & Price, H. (1993). Causation as a secondary quality. *The British Journal for the Philosophy of Science, 44*, 187-203.

Mellor, D. H. (2004). For facts as causes and effects. In *Causation and Counterfactuals*, 309-324.

Minică, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E., & Neale, M. C. (2017). Extending causality tests with genetic instruments: An integration of Mendelian randomization and the classical twin design. *Preprint on bioRxiv*. doi: https://doi.org/10.1101/134585

Mosing, M. A., Madison, G., Pedersen, N. L., & Ullèn, F. (2016). Investigating cognitive transfer within the framework of music practice: Genetic pleiotropy rather than causality. *Developmental Science, 19,* 504-512.

Norton, J. (2003). Causation and folk science. *Philosophers' Imprint*, *3*(4). Retrieved from www.philosophersimprint.org/003004/

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., ... & Stephens, M. (2008). Genes mirror geography within Europe. *Nature, 456*, 98-101.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … & Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539-542.

Paul, L. A., & Hall, E. J. (2013). *Causation: A user's guide*. Oxford University Press.

Pearl, J. (2000). *Causality*. Cambridge, UK: Cambridge University Press.

Pearl, J. (2001). Bayesianism and causality, or, why I am only a half-Bayesian. In D. Corfield & J. Williamson (Eds.), *Foundations of bayesianism* (pp. 19-36). Dordecht, Netherlands: Springer.

Pearl, J. (2010). An introduction to causal inference. *The International Journal of Biostatistics*, *6*, Article 7.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: a primer*. New York, NY: John Wiley & Sons.

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics, 48,* 709-717.

Psillos, S. (2009). Regularity theories. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 131-157). Oxford, Oxford University Press.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research and Human Genetics, 5,* 554-571.

Robinson, M. R., Kleinman, A., Graff, M., Vinkhuyzen, A. A. E., Couper, D., … & Visscher, P. M. (2017). Genetic evidence of assortative mating in humans. *Nature Human Behavior, 1,* 0016.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39,* 33-38.

Salmon, W. C. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science*, *64*, 461-477.

Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, *114*, 327-358.

Skyrms, B. (1984). EPR: Lessons for metaphysics. *Midwest Studies in Philosophy*, *9,* 245-255.

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

Spirtes, P. (2010). Introduction to causal inference. *Journal of Machine Learning Research*, *11*(May), 1643-1662.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland Publishing Company.

Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(4), 814-820.

Tabery, J. G. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of science*, *71*(1), 1-15.

Taylor, R. (1966). *Action and purpose*. Englewood Cliffs, NJ: Prentice-Hall.

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science, 27,* 138-149.

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

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

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics, 88,* 76-82.

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowsi, M., … & Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications, 9,* 224.

Zultan, R. I., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: causality and counterfactuals in group attributions. *Cognition*, 125(3), 429-440.