

Genetic traits and causal explanation

Abstract

I use a contrastive theory of causal explanation to analyze the notion of a genetic trait. The resulting definition is relational, an implication of which is that no trait is genetic always and everywhere. Rather, every trait may be either genetic or non-genetic, depending on explanatory context. I also outline some other advantages of connecting the debate to the wider causation literature, including how that yields us an account of the distinction between genetic traits and genetic dispositions.

1) Introduction

Many traits and dispositions are labeled *genetic*, and such labeling is both widespread in science and apparently far from arbitrary. Yet it is a truism that every trait is the end-product of a complex developmental process involving both many different genes and a multitude of environmental and other epigenetic factors. How then can some traits usefully be termed genetic and others not?

In this paper, I import influential recent theories of causal explanation to develop an account of genetic causation, and hence of genetic traits and dispositions. My thesis is that the latter two are best seen as *explanatory claims*; in particular, a trait is genetic, roughly speaking, just in case it is explained by genes or it is *not* explained by environment. The greater novelty lies in how this idea is made precise in the form of a relational definition, an implication of which is that no trait is genetic always and everywhere. Rather, every trait may be either genetic or non-genetic, depending on explanatory context.

Precisely because genes are part of the causal history of any trait, merely tracking which traits have genetic causes is insufficient for picking out some rather than others to be ‘genetic’. Some authors have therefore sought to supplement that by defining genetic traits to be those where there is some special kind of connection between gene and trait over and above mere causation. Examples include: that genetic traits are those that are caused by genes especially ‘directly’ (Hull 1981), or that there is a special kind of necessity linking them to genes (Gifford 1990). A stronger tradition has been to analyze the matter in terms of *differences* – this gene rather than that one is the cause of this trait rather than that one (Sterelny and Kitcher 1988). Indeed, a focus on differences lies at the heart of the statistical techniques standard in behavioral genetics, which look for correlation between *variation* in phenotype and *variation* in genes. That still leaves it to be decided though exactly what pattern of correlation or variation is taken to be characteristic of genetic traits in particular. One crude approach, seldom advanced explicitly, would be to look simply for high statistical heritability. More sophisticated definitions can also be framed in terms of such statistical relations though, such as Smith (2007); these likewise focus on differences. (All these previous approaches are discussed in section 5.)

This paper’s approach too focuses on the causal connection between gene and trait, and is also framed in terms of differences. But it is connected explicitly to the wider literature from philosophy of science on causal explanation. Among the benefits resulting from this, a major one is that it yields an analysis of the relation between genetic *dispositions* and traits (section 6). These two concepts are clearly distinct and as a result used differently but, to my knowledge, their relation has not been analyzed previously. It

turns out that a well founded account of genetic dispositions is crucial to understanding whether to call genetic or non-genetic the many traits that have statistical heritabilities around 0.5 in the populations in which they have been studied.

This paper is theoretical in the sense that it is initially couched in terms of theories of causation rather than in terms of biological or medical practice. Nevertheless, there is value in first investigating just what sense can be made of genetic causation or explanation in the abstract, so to speak, before then examining the applicability of whatever version of those notions turns out to be theoretically defensible.

2) Contrastive explanation

The core of this paper will be to analyze genetic explanation. To that end, I shall adopt the leading contemporary theory of causal explanation, which attributes to such explanations a *contrastive* structure – a cause-rather-than-contrast explains an effect-rather-than-contrast. To illustrate, consider the claim ‘Socrates sipping hemlock explains why he died’. That sounds plausible enough, but consider two possible clarifications:

- 1) ‘Socrates sipping *rather than guzzling* hemlock explains why he died.’ (Seems wrong.)
- 2) ‘Socrates sipping hemlock *rather than wine* explains why he died.’ (Seems right again.)

The lesson is that explanatory properties are sensitive to choice of contrast. A similar lesson applies to effects as well as causes. Imagine that a short circuit ignites a mixture of wood and potassium salts, yielding a purple fire. Then:

- 1) The short circuit explains the purple fire rather than *no* fire.

2) But it's the potassium salts that explain the purple fire rather than *yellow* fire.

The contrastive view dates from Dretske (1972). Notable developments of it include van Fraassen (1980), Garfinkel (1981), Achinstein (1983), Hitchcock (1996), and – most influential recently – Woodward (2003). I leave further details and justifications to those works, and here focus instead on how to apply the contrastive apparatus to genes.

Begin with the explanandum, i.e. with the object of a genetic explanation.

Historically, this has variously been taken to be a behavior, trait, disposition, or piece of knowledge. I shall take behaviors and knowledge to be examples of phenotypic traits or dispositions. The analysis to be developed here will therefore concern just the latter two. To begin with I shall focus on traits, before turning in section 6 to dispositions.

Label the particular actual trait of interest T_a . In accordance with the contrastive view, it is crucial also to specify a contrast to T_a , so label that T^* . To see intuitively how choice of T^* matters, consider T_a = my actual two legs, and T^* = I have only one leg. This represents a paradigm case of a genetic trait – what about me could be more genetically influenced than that I have two legs rather than one? Now, as it happens, one of my legs is actually slightly bent due to a childhood accident. This suggests an alternative contrast of T^* = my two legs are both straight.¹ In order to explain why my leg is bent rather than straight, i.e. T_a rather than this new T^* , we would now appeal to my accident and *not* to genes.² That is, for the first choice of T^* we deem genes to be

¹ The different choices of contrast here can be seen as a formal device to pick out different *aspects* of my legs – respectively, their number and their straightness. Choice of T^* can also serve to represent the distinction (Sober 1998, 795) between a trait's initial development and its subsequent modifiability.

² Moreover, in some contexts even being born with a defect may also be explained environmentally – see the discussion of a thalidomide example in section 5 below.

explanatory, while for the second choice we do not.³ Yet T_a is identical in both cases, namely my actual legs. So just specifying T_a alone is insufficient.

Matters become a little more complicated when we turn to the explanans, i.e. to a trait's causal history. The contrast for the explanans is some alternative version of that history. More precisely, it is some alternative event within that history plus all the causal consequences of that alternative event. I shall be interested only in a trait's causal history since the relevant organism's conception. Of particular importance to us is that we may partition this history into two types of factor, colloquially labeled 'genes' and 'environment'.⁴ More precisely, the partition of the causal history is into two highly asymmetric portions – first, the particular genome at the moment of the particular organism's conception; and second, the entire rest of that organism's developmental history. Corresponding to this division, contrasts associated with the explanans are also of

³ The objects of ascriptions of genetic explanation may form a broader class than 'traits' as that term is customarily used in biology. On some views, for instance, the single event of my currently having my actual legs can, strictly speaking, correspond to more than one trait, depending on whether we are focusing on my legs' number or straightness. In effect, this would be to individuate traits in part via explanatory context. I shall ignore controversies over exactly how traits should be defined, and instead shall use the term rather liberally. Generally, I shall understand the relata of a causal explanation to be two pairs of events, or more precisely, two pairs of an actual and contrast event, and my eventual definitions will be framed accordingly. Therefore I retain the text's specifications of T_a and T^* , even at the cost of occasional conflict with some views of trait individuation. (The substance of this paper's case could be made even given other choices of explanatory relata than pairs of events, and so for ease of exposition I shall sometimes leave it underspecified exactly what kinds these relata are.)

⁴ I ignore the occasionally fuzzy borderline between these two categories because, with regard to genetic traits, that has not been the salient locus of philosophical dispute. Rather, the strengths and weaknesses of various analyses have concerned other matters (see section 5).

two types – those that cite alternative particular genomes at conception, and those that instead cite alternative versions of the rest of the developmental history.⁵

3) A definition

We may now formulate an explicit definition. It is intended to apply to token cases, i.e. to particular traits of particular organisms. Let T be a function that takes an organism's causal history as input, and yields trait values as outputs. Formally, let T_a = the actual trait value; G_a = the actual genome at conception; and E_a = the rest of the actual developmental history. The organism's complete history is therefore $G_a \& E_a$. And let T^* = the salient contrast trait value; G^* = the salient contrast history resulting from the substitution⁶ in of an alternative genome at conception; and E^* = the salient contrast history resulting from the substitution in of some other alternative event in the developmental history.⁷

Some clarifications from the start will be useful. In all cases, the contrast-effect will be a T^* , and the contrast-cause either a G^* or an E^* . (As noted, in any one case the contrast-cause will be of either G^* type or E^* type, never both simultaneously.) T^* , G^* ,

⁵ In principle, of course, one could formulate contrasts that vary *both* these aspects of the explanans. As a matter of fact, such 'combined' contrasts do not ever seem to be salient in actual disputes (see also van Fraassen 1980, 126). But if they were, in my view consideration of them would give us no information regarding genetic explanation.

⁶ 'Substitution' is intended here as a neutral term that may be taken to correspond either to a 'miracle' in Lewis-style possible-worlds semantics, or to an 'intervention' in the semantics of the causal modeling literature. Again, for the purpose of elucidating genetic traits I do not endorse any particular semantics for counterfactuals in general, since that is not the salient locus of philosophical interest.

⁷ In general, T^* , G^* and E^* may refer to *sets* of contrasts. For ease of exposition, I shall often assume them to be singletons. A definition analogous to [GT], given below, can easily be formulated for non-singleton cases too. (See Northcott 2008b for more on non-singleton contrasts.)

and E^* are all counterfactuals. For both G^* and E^* , the only differences from the actual history are, to repeat, the initial substitution itself plus that substitution's causal consequences. Intuitively, this amounts to a *ceteris paribus* provision – we are interested in what would have occurred had G_a or E_a been different but with nothing else changed. For example, what would have happened had a mutated version of a gene been present, keeping the rest of the genome and all epigenetic factors the same? In this way, we would isolate the causal contribution in this particular situation of that gene alone.

For ease of exposition, I shall often denote a G^* or E^* just by the alternative substituted in, for instance 'G* = alternative genome X'. Then, our definition is:

- 1) When a G^* is salient, T_a is *genetic iff*: $T(G^*) = T^*$
- 2) When an E^* is salient, T_a is *genetic iff*: $T(E^*) \neq T^*$ [GT]

$T(G^*)$ denotes the trait that would have resulted from the alternative causal history represented by G^* . On a contrastive view, the conditions for full explanation are: that the actual cause yields the actual effect; *and* that the contrast-cause would have yielded the contrast-effect. The first condition is satisfied automatically here, since by assumption $T(G_a \& E_a) = T_a$.⁸ When a G^* is salient, the second condition is satisfied *iff* $T(G^*) = T^*$; when an E^* is salient, it is satisfied *iff* $T(E^*) = T^*$.

⁸ Strictly speaking, this assumption is reliable only given determinism. For simplicity, I shall consider here only the deterministic case. Evaluation of indeterministic counterfactuals, and of probabilistic causal-explanatory claims more generally, goes beyond the scope of this paper. In practice though, uncertainty due to indeterminism seems only rarely to be the focus of actual disputes about whether a trait is genetic.

In words, definition [GT] amounts to saying that a trait is genetic just in case it is explained by genes or it is *not* explained by environment. If genes made the difference, the trait is genetic; likewise, it is genetic if environment did *not* make the difference. Formally, these correspond respectively to the conditions $T(G^*) = T^*$, and $T(E^*) \neq T^*$.⁹

It is important to be clear from the start on the type-token distinction. Definition [GT] applies only to the token case of a particular trait of a particular organism. Which contrasts are salient will obviously vary with context. Moreover, the function T is essentially a device for representing causal relations, and such relations are also context-dependent. Whether striking a match causes fire, for instance, depends on whether the match is wet, whether there is sufficient oxygen, whether it is windy, etc.

Claims that a trait is genetic are often made at the *type* level – ‘Down’s syndrome is genetic’ or ‘scars are not genetic’. How then can [GT] be applied to them? My answer is that such type claims are implicitly about particular populations of token cases. In particular, usually they tacitly assume ‘typical’ or ‘normal’ populations and explanatory concerns, corresponding to particular (collections of) choices of T^* and G^* or E^* . For each of the token cases within the population, the conditions in [GT] are deemed to be satisfied. Any given case of Down’s syndrome, for example, could not have been avoided by any salient alternative environmental input – or at least that is what is being asserted when that trait is described as ‘genetic’ at a type level.

A common worry is whether it makes sense to talk of contrasting either genes or environment in isolation, given the two factors’ obvious continual interaction. Both gene-

⁹ Given that $T(G_a \& E_a) = T_a$, claims that a trait is genetic thus boil down to evaluations of particular counterfactuals. Of course, as with all counterfactuals, these ones may be vague or indeterminate. In so far as they are, then so likewise I claim are the associated genetic claims.

environment interaction and correlation are standard difficulties in population-level analyses, but at the token individual level the worry is misplaced. For example, suppose I had been born with greater athletic talent.¹⁰ Possession of, so to speak, a more athletic genome would presumably have led to a different environment too relative to the one I actually did experience – perhaps it would have led me to seek, and to have been given, more intensive athletic training.¹¹ But that does not render this counterfactual any harder to evaluate than counterfactuals in general. For any genetic contrast G^* (or environmental one E^*), the history need be held fixed only up to the time of the antecedent, i.e. up to the time of the relevant substitution. The subsequent history can then unfold as it may. In causal modeling terms, there is no problem if the subsequent history changes too, so long as it does so only as a causal consequence of the initial intervention (Woodward 2003).

As is well known, the very same genome may yield dramatically different phenotypic outcomes depending on environment. This is just gene-environment interaction at work. One consequence is that the phenotypic significance of a particular genetic difference may also vary with environment. For instance, anticipating an example below, even though genes may explain why I am taller than my father, it is also possible that in a very nutrient-poor environment both of us alike would have grown only to five

¹⁰ On some views of personal identity this more athletic creature would no longer be ‘me’ at all. For our purposes, the difficulty is not a deep one. If desired, the discussion can be re-phrased without loss as whether a counterpart with the particular alternative genome at conception would have ended up with the same traits as me.

¹¹ This would be an instance of gene-environment *correlation* – the more athletic genotype is correlated with a more intensive training environment. In addition, more intensive training presumably itself causes better athletic performance. If this extra effect on performance is greater in the case of the more athletic genotype than the less athletic one, it would be an instance too of gene-environment *interaction* – i.e. when the impact of a particular environment varies depending on the genotype.

feet. In other words, our genetic difference might yield a height difference only in some environments not others.

Formally, such context-dependence is already incorporated into [GT], via the sensitivity of the T function to environmental conditions. Nevertheless, the issue can still lead, intuitively at least, to problems for any causation-based analysis of genetic traits. A familiar example, discussed for instance in Sesardic (2005), illustrates how: suppose that all red-haired children are discriminated against because of their hair color, receiving no formal education and thus scoring poorly on scholastic tests. Suppose also (plausibly) that red hair is caused by the presence of certain genes (or, strictly, by the presence of certain alleles of those genes rather than others). It follows that, given the discriminatory social environment, those genes will also cause reduced scholastic achievement. Yet presumably, given normal education red-haired children would have done just as well as other children. Thus we naturally explain the low scholastic scores by appeal to the social discrimination, not to the ‘red-hair genes’. Yet, the objection concludes, formula [GT] will count the trait as explained by genes, and thus (sometimes) as genetic, nevertheless.

In reply, first note that this is an issue for everyone (as Sesardic’s discussion makes clear). Any definition based on statistical correlation, for instance, will likewise endorse a large genetic role here. Next, there is no dispute that in such a scenario the ‘red-hair genes’ *are* causes of low scholastic achievement, via the intermediate cause of the discrimination that they induce. The problem is thus a version of Mill’s classic one of causal selection: there is no ontological preference for the ‘good’ cause (social discrimination) over the ‘awkward’ one (i.e. the genes). The conventional wisdom is that distinguishing between the two causes can therefore only be done by appeal to pragmatic

factors. The issue here is that our common emphasis on the discrimination rather than the genes is presumably the result of a moral judgment, perhaps intertwined with awareness of possibilities for intervention.

It may well be that, as a result, our notion of ‘genetic trait’ is infused with these extra pragmatic considerations, in which case no purely logical analysis of the term will ever be fully satisfactory. On the other hand, a recent strand in the causation literature addresses this very issue. Backed by empirical evidence, it suggests that a ‘cause’ is, by definition, roughly speaking that particular counterfactual dependency that is rendered salient by normative considerations (Hitchcock and Knobe 2009). So understood, we might, as desired, pick out as a cause in this example the dependency on social discrimination but not the dependency on red-hair genes. A causal definition of a genetic trait would therefore be endorsed again. Once full resolution of these theoretical issues is reached, the formulation in [GT] might be adjusted appropriately.

4) Objectivity and context

It follows from definition [GT] that being a genetic trait is a relational property, of the form T_a is genetic relative to T^* and G^* (or to T^* and E^* , as the case may be). There is no absolute fact of the matter, independent of explanatory context. More formally, whenever we ask whether some trait is genetic, on my view a presupposition of the question is a particular specification of contrasts. That these relativizations are often not explicit does not show that they are not present, only that they are tacit. The intuition against the thought that the same trait could be both genetic and not genetic, is explained as being the result of a violation of pragmatic maxims dictating relevance to our

conversational presuppositions (in particular, to the presupposition fixing only a particular specification of contrasts as salient). Similarly, explanatory claims in general often have a non-relational surface form even though really they are relational.

Such a relational property is not arbitrary. In particular, once (but only once) given a specification of contrasts, the truth of whether a trait is genetic is clearly objective – or anyway as objective as the evaluations of the relevant counterfactuals. It is thus certainly possible to rule out some claims as erroneous. For example, assertions that some individual's criminality is genetic may, in context, be assertions that given the same upbringing or schooling an individual with a different genome would not have committed crimes. That claim is an objective causal hypothesis.

Often, contrasts are not specified explicitly. Therefore we still require an account of how, in those cases, they are determined implicitly. The answer, as already mentioned and, as for any conversational presupposition, is conversational context. Can we flesh that out in more detail? One general constraint we have already seen – contrasts are counterfactual, i.e. $T^* \neq T_a$, and neither G^* nor E^* replicates the organism's actual developmental history $G_a \& E_a$. A further general constraint is that contrasts must, so to speak, genuinely contrast. I propose that T_a and $G_a \& E_a$ must each be nomologically incompatible with their associated contrasts. (This ensures that T_a and $G_a \& E_a$ constrain the definition at all.) However, it also remains true that, beyond this, there exists no generally agreed-upon formal procedure for nailing down in every case exactly how circumstances do specify choices of contrast.¹² How then can we ever be confident what the intended contrasts are, and so how can [GT] ever be tested? The best way, it seems to

¹² See Schaffer (2005), Maslen (2004) and van Fraassen (1980) for further discussion.

me, and therefore the best evidence for [GT], is deliberately to manipulate contrasts and then to track whether our judgment that a trait is genetic indeed consequently varies in the manner predicted. Obviously, no exhaustive catalogue of cases is possible so I can only appeal to particular illustrative examples.

To this end, consider T_a = my height, and $G_a \& E_a$ = my own life history. Suppose I ask the explanatory question, why am I taller than a neighbor of mine who grew up in a poor family? This immediately suggests the contrast T^* = my neighbor's height, and, roughly speaking, either G^* = my neighbor raised in my environment, or E^* = me raised in my neighbor's environment. If we judge that the neighbor would not have grown any taller even given my wealthier upbringing, i.e. $T(G^*) = T^*$, then sure enough we correspondingly judge that my greater height is indeed genetic – because explained by our different genes. And likewise if, in the E^* case, we judge that if raised in his environment I too would have attained only my neighbor's height, i.e. $T(E^*) = T^*$, then, as per [GT], we would *not* judge my greater height to be genetic – because explained instead by my upbringing.

Next, suppose I set, for the same actual trait T_a = my height, a different conversational context by asking, why am I taller than my father? Now the contrasts are naturally T^* = my father's height, and, roughly speaking, either G^* = my father raised in my environment, or E^* = me raised in my father's environment. If I judge that my father would likely have reached my height if he'd had my nutritionally superior upbringing, i.e. that $T(G^*) \neq T^*$, then our height difference is not explained by genes – and sure enough is not judged genetic. Similarly, if I believe that I would have attained only my father's height given his upbringing, i.e. that $T(E^*) = T^*$, then our height difference is explained

by environment and once more we accordingly judge it not genetic. Again, judgment tracks the predictions of [GT].

Notice how the very same trait, namely my height, is thus deemed genetic in some explanatory contexts but not in others. Indeed, according to [GT], a similar fate awaits *any* trait. We already saw this also for $T_a = \text{my legs}$ (section 2).

Does this not all leave too much weight resting on which contrasts are deemed ‘salient’? No, and it is important to appreciate in exactly what way that issue matters – and in what way it doesn’t. For sure, according to this paper’s account whether a given trait is deemed genetic depends critically on choice of contrasts. But in order to *test* that account, what is relevant is that judgments of whether a trait is genetic track changes in contrasts in the manner claimed. As noted, we may test that in turn by manipulating choice of contrast independently and then seeing whether our judgment follows as predicted. Therefore it is not fatal that we have no foolproof algorithm for generating choice of contrast in every context. That merely implies that there may be no fact of the matter regarding whether a trait is genetic before contrasts are specified – which is exactly what a relational definition such as [GT] is claiming anyway. The demand for an objective determination of contrasts perhaps betrays an unspoken but incorrect assumption that there should always be some fact of the matter regarding whether a trait is genetic. But if context leaves choice of contrast indeterminate, the existence of such a contrast-independent fact of the matter is precisely what [GT] denies.

5) Relation to previous literature

Many of the conclusions from this paper are familiar. Sterelny and Kitcher (1988) long ago argued that genetic causation should be understood context-specifically. We can speak of a gene causing a phenotypic trait in a particular case even though that gene may not do so in every circumstance. In this paper, that idea is generalized and formalized by defining the explanans of a genetic explanation to be $G_a \& E_a$ -rather-than- G^* , i.e. one genome rather than another. The sensitivity to extra-genomic environment is represented via the sensitivity of the trait function T to E_a . Sterelny and Kitcher's explication of the 'gene for' locution can be represented similarly. Generally, the common emphasis on genetic *differences* explaining trait *differences* is captured naturally by a contrastive apparatus. This paper's account of genetic explanation shares features with other accounts too. It shares the emphasis on pragmatic relativization in Gannett (1999), for instance, and the emphasis on the token case of Waters' (1994) 'Difference Principle'. The point here is to demonstrate how these conclusions can be expressed in terms of, and endorsed by, the wider theory of causal explanation.

As noted earlier, one thing that has proved elusive in philosophy generally is an objective algorithm for determining what contrasts are implied by any given context. Nevertheless, that does not mean that we cannot examine which pragmatic factors tend to influence choice of T^* , G^* and E^* , or, more particularly, that influence whether it's a genetic cause that is picked out and thus a trait declared 'genetic'. Indeed, in effect much of the literature has focused on exactly this issue. Often, what makes a certain cause salient is the research program or goals of the particular scientist. Some may focus only on those causes that actually vary in the population for instance (Waters 2007), others on those that do not actually vary but that are nevertheless potentially the target of

efficacious interventions. But, according to [GT], these disputes regarding goals are not disputes regarding the *definition* of whether a trait is genetic.

Smith (2007), focusing particularly on the example of genetic diseases, surveys many causal selection criteria that have been used to justify picking out traits as genetic, such as those mentioned in section 1: unusually ‘direct’ causation by genes; and genes being the only ‘abnormal’ factor in a trait’s causal history. In this paper’s terms, these criteria can be seen as criteria for selecting some rather than other contrasts. Smith argues convincingly that none of them is adequate in the face of actual biological complexity. One common weakness is their lack of relationality – that is to say, a given trait will be declared either genetic or non-genetic *simpliciter*, with no allowance for variation with context.

Standard methods in behavioral genetics and other sciences can be used to define statistical heritability – roughly speaking, the proportion of phenotypic variance in a given population that is ‘due to’ genetic variance. Glossing over many details, this essentially tracks statistical correlation between genetic and phenotypic variance. One attractive feature of statistical heritability in this context is that it incorporates a focus on differences, and thus is able to allot different scores to different traits despite the fact that all traits alike have genetic causes. Moreover, because it is population-specific, statistical heritability also incorporates a certain relationality. Nevertheless, notwithstanding its possible usefulness for other purposes, high statistical heritability is a poor candidate for a definition of whether a trait is ‘genetic’. Notoriously, a trait commonly thought genetic, such as number of legs, will often score very low for heritability simply because it is almost universal and what little variation there is in the population is due to

environmental factors such as accidents. As it were, the relationality here is of the wrong sort – it is not sensitive to the appropriate counterfactuals, but rather to other members of an actual population. As a result, it is much disputed whether any causal inferences follow from heritability scores at all.¹³ Moreover, population-level statistics such as heritability are inapplicable to individual-level cases.¹⁴ Thus it is necessarily beyond its purview whether my own eye color, for instance, is genetic.

Smith (2007) offers his own proposed analysis of genetic traits, in particular of genetic diseases, also based on population frequencies but much more sophisticated than simple statistical heritability. Nevertheless, the same basic worries apply, namely the lack of connection to the wider causation literature and the inapplicability to individual-level cases.

Another proposal might be to define genetic traits in terms of evolution, perhaps to be those that make a difference to evolutionary outcomes.¹⁵ After all, evolution is often defined as change in gene frequencies within a population, and traditionally one of the necessary conditions for a trait to evolve is that that trait be heritable. But a closer look gives pause. First, every trait is influenced by genes and thus potentially makes a difference to evolutionary outcomes, so that alone is insufficient to distinguish between traits in a principled way. Further, some genetic diseases, such as Klinefelter's syndrome, are not heritable and thus presumably do *not* make a difference to evolutionary outcomes

¹³ For a sample of such attacks, see Lewontin (1974), Shipley (2000), Spirtes et al. (2000), and Northcott (2008a). In the philosophical literature, Sesardic (2005) mounts the most vigorous defense of heritability against this consensus.

¹⁴ Sober (1988) argues otherwise, although see Northcott (2006) in response.

¹⁵ I thank an anonymous referee for this suggestion.

in the way envisioned.¹⁶ Moreover, finally, much contemporary theory emphasizes the importance of non-genetic mechanisms in evolution anyway (Sterelny 2003, Oyama et al. 2001).¹⁷

I believe it is an open question whether a purely logical analysis exists that does successfully track our every usage of ‘genetic’ with respect to traits. There is certainly no guarantee that any should. (Indeed the red-haired example in section 3 suggests, consistent with [GT], that some pragmatic relativization is unavoidable.) But the point of this paper is something different, namely to formulate a connection to wider philosophical literature. How exactly the contrasts in [GT] should be filled in is a separate matter, about which [GT] is not directly concerned since it remains neutral on where contrasts come from. As noted earlier, the ‘test’ of [GT] is, rather, that once contrasts have been determined, matters then track [GT] as predicted. Embedding our analysis into the theory of explanation, meanwhile, offers several benefits of its own, to which I turn now.

First, consider why we should even care whether a trait is genetic or not. What normative punch could ever result from such a claim? This paper’s account, by way of its connection to the causation literature, offers an answer – the counterfactuals that, according to [GT], comprise such claims are also exactly those that license *interventions*. Declaring my eye color to be genetic, for instance, asserts that – in context – no salient

¹⁶ Klinefelter’s syndrome describes XXY individuals, i.e. males with a third sex chromosome. Its cause is an error during meiosis rather than a father with the condition. (Indeed, before modern reproductive technology one of the condition’s symptoms was infertility.)

¹⁷ An alternative proposal is that genetic traits are those that are evolutionary adaptations. But this again seems to track actual usage badly. For example, many genetic diseases, such as cystic fibrosis or Down’s syndrome, are hardly adaptive.

environmental intervention could have altered it in the past, or could alter it now. Calling a trait genetic thus serves to pick out which interventions are (or would have been) efficacious, and which not. Implications follow immediately for explanation and blame. For example, deeming a disease to be genetic implies that it cannot be blamed on any (salient) aspect of parental care.

As ever, the emphasis on salience is critical here. Take the disease phenylketonuria, or PKU. As is well known, this was traditionally thought a ‘genetic disease’ in that it resulted from a particular genome, possession of which made the onset of the disease almost inevitable. That is, for any normal genome G^* , $T(G^*) = T^*$, where $T^* =$ a normal child, in contrast to a child suffering from PKU. Moreover, it was not thought that any environmental intervention could avert this outcome, hence $T(E^*) \neq T^*$. Famously though, it is now known that a special diet can alleviate (most of) the impact of the disease. In this sense, it is no longer a ‘genetic disease’, and the status of PKU has become ambiguous, as given the contrasts that are likely salient now, cases of it may be explained by either genes or environment. Depending on which particular contrast is involved, the disease thus may or may not be appropriately described as genetic. (Or, in particular circumstances, it may now be best described as a genetic *disposition* – see section 6.)

An explicitly causal scheme also highlights an ambiguity that has tended to be overlooked – namely, whether we are talking about genetic *causation* or merely genetic *explanation*. In the metaphysics literature, a contextual-contrastive view is relatively standard now with regard to explanation, but with regard to causation itself it is much more controversial (e.g. Davidson [1967] 1980). In particular, relativization to contrasts

introduces a pragmatic element that has traditionally been thought characteristic only of explanation. On the other hand, there is a growing segment of the literature that does endorse an explicitly contrastive view of causation itself (Northcott 2008b, Schaffer 2005, Maslen 2004).¹⁸ This view is also endorsed by the contemporary Bayes net and causal modeling literatures (Pearl 2000, Spirtes et al. 2000), as well as, as already noted, by Woodward (2003). On such a view, choices of contrast influence not just whether genes *explain* some trait but also whether they *cause* it. The point is that much talk of genetic causation so far has carried an (unacknowledged) metaphysical commitment to such causal contrastivism. It is better to be open about this.

Next, definition [GT] easily handles borderline cases. For example, male bellies tend to expand with age. This phenomenon is genetic in the sense of occurring in a wide range of human environments, and because it is often explained by having a male rather than female genome. But it is also not genetic in the sense that it could be avoided by eating less. Formally, for T_a = my middle-aged male cousin's large belly, and T^* = a smaller belly, it is true that for G^* = a typical female genome, $T(G^*) = T^*$, i.e. a woman with the same lifestyle as my cousin would have had a smaller belly. Moreover, for E^* = many alternative environments than the one my cousin has experienced, such as that in many other cultures, $T(E^*) = T_a$, i.e. the large belly is not explained by my cousin's particular environment. So the trait seems genetic. But for *other* choices of E^* , such as E^* = he ate less, now $T(E^*) = T^*$, i.e. now environment *is* explanatory, and so the trait no longer seems genetic. The point is that either choice of E^* will likely be salient quite often. As a result, frequently the size of the belly will seem genetic but frequently also it

¹⁸ And also of probabilistic causation (Hitchcock 1996).

will not. The larger point is that, as with many traits, I have no clear intuition as to whether expanding male bellies *simpliciter* 'are genetic'. Rather, my judgment only *becomes* clear once given a particular explanatory context, just in the manner that [GT] predicts.

Perhaps we can now also see a connection between a trait being genetic and it being *innate*. In particular, might the concept of innateness be equivalent to genetic explanation and lack of environmental explanation? To investigate that would take us beyond this paper and into the huge literature on innateness and the many complications to measuring 'nature versus nurture'. The point here is merely to flag the possibility that innateness might be elucidated by the literature on causation, and perhaps also the relation between 'innate' and 'genetic'.

Lastly, a final advantage of analyzing genetic traits via an explicit theory of causation is that this also yields us an analysis of the relation between such traits and genetic *dispositions*. I turn to that final advantage now.

6) Traits versus dispositions

The conditions of adult membership to the Philharmonic Academy in Bologna required a candidate to write an elaborate motet in six parts, founded upon a melody assigned from the Roman Antiphonarium, the work to conform to the strictest rules, with double counterpoint and fugue. In the summer of 1770, the Academy was visited by a 14-year-old boy who tried the test. In less than three-quarters of an hour he rapped at his door and asked to be let out. The authorities sent him word not to be discouraged, but to keep on trying, as he had yet three hours, and might accomplish it. They were greatly astonished on finding that he had already finished, having produced a complete master work,

abundantly up to all requirements, the whole written in a peculiarly neat and accurate manner. The 14-year-old boy was the young Mozart.¹⁹

Was Mozart's dazzling skill a genetic trait? It is commonly supposed so. The thought behind this judgment is easily captured formally: set T_a = Mozart's feat of writing the motet in 45 minutes, compared to T^* = needing longer as is typical for the rest of us, and let G^* = some typical non-Mozartian human genome. Then it seems clear that $T(G^*) = T^*$, in other words that Mozart's feat is explained genetically. But there's a catch. Like any trait, T_a was the product of both genetic and environmental inputs. And this particular T_a required not just normal environmental inputs of nutrients, physical nurture and so on but also something much less commonplace, namely that almost from infancy Mozart was hot-housed as a prodigy by his musician father.²⁰ So for many salient E^* , this T_a is explained environmentally – without the hot-housing, Mozart could not have pulled off his dazzling feat in Bologna. Intuitively, and according to [GT], that renders T_a not genetic after all. This variability of verdict with explanatory context is not unique, indeed we have seen that it is true of all traits. What is unusual about the Mozart case is the salience of environmental explanation even though Mozart is often perceived in nativist terms as a unique 'genius'.

A telling detail is that Mozart is more usually described as having a God-given *talent* than God-given *traits*. Reading 'God-given' as 'genetic', the solution, I shall argue, lies in the notion of genetic *dispositions*. In particular, talents must be explicated as

¹⁹ Adapted from Mathews (1891, 295-296).

²⁰ Later in life, Mozart himself commented on the necessity of an intensive environmental input: "People make a mistake who think that my art has come easily to me. Nobody has devoted so much time and thought to composition as I. There is not a famous master whose music I have not studied over and over."

dispositions rather than traits, for like dispositions they are only potentials and so might never be realized. (To be clear on terminology here: I shall describe the trait of the 14-year-old Mozart's musical skill in Rome by his 'feat', and his initial disposition to be able eventually to reach that level of skill by 'talent'.) Whereas my having two legs is normally described as a genetic trait rather than a disposition, in the case of Mozart's musical ability it is the other way round. Why? And what, if anything, of interest does the distinction mark?

Begin with the analysis of dispositions generally. Paradigmatically, a given vase is fragile *iff* it shatters when struck. What is the biological analogue? Label the relevant disposition here, corresponding to fragility, $D = \text{Mozart's musical talent}$. Corresponding to the shattering, i.e. to the realization of the disposition, we have the trait $T_a = \text{Mozart's feat of writing the motet in 45 minutes}$. Corresponding to the striking of the vase, i.e. to the relevant environment, we have $E_a = \text{Mozart's actual childhood hot-housing}$. Then, filling in the definition of a disposition, Mozart has the disposition D *iff* T_a when E_a .²¹

So far, that just expresses that D is a disposition. What does it mean to say that it is a *genetic* one? I propose to define the latter to be a dispositional property of a genome at conception.²² Thus genetic dispositions, unlike most genetic traits, are things we do possess literally from conception. Of course, because genes feature in the causal history of any trait, it now follows that *all* traits – genetic and non-genetic alike – are realizations

²¹ I stay neutral here on the vexed metaphysical issue of dispositional realism. Thus I take no position on whether a trait is also *explained* by an underlying disposition in addition merely to realizing it.

²² Environments too have dispositions, of course. For instance, a particular kind of schooling may tend to produce particular kinds of graduates. Such dispositions do not seem relevant here though.

of some genetic disposition or other.²³ So on the face of it, appeal to such dispositions seems rather vacuous. Why then should such appeals ever be made? Because, it turns out, in particular circumstances they *can* be informative after all. Tracing exactly how will prove to be intricate work.

Begin on the genetic side. Continuing with the Mozart case for illustration, recall that D is a property of Mozart's actual genome G_a , that T_a = his musical skill, and E_a = his hot-housing upbringing. What of the salient contrast genome G^* ? There are two possibilities: either G^* has D too, or it does not. Only in the latter case, it turns out, will invocation of a genetic disposition be useful and hence conversationally apt. Assume to start with that it is indeed the latter case. Because therefore G^* does not have D whereas G does, it follows from the definition of a disposition that:

- 1) G_a implies T_a when E_a ; and
- 2) G^* implies not- T_a when E_a .

Labeling as usual by T^* the salient contrast to T_a , and assuming the actual environment E_a ²⁴ in both cases, these just amount to:

- 1) $T(G_a) = T_a$; and
- 2) $T(G^*) = T^*$.

But we already know that $T(G_a) = T_a$, since that is just the actual case. So the pragmatic import of the claim that D is genetic boils down to claim 2 regarding a counterfactual.

And claim 2 is of course just one half of our definition of a genetic *trait*. In other words,

²³ Therefore disposition is the broader category here: all genetic traits are realizations of genetic dispositions, but genetic dispositions – when realized at all – may be realized by traits that are either genetic or non-genetic.

²⁴ Modulo any changes in environment that are causally downstream of the substitution of G^* for G.

given that the salient contrast for the explanans is genetic, asserting that a trait T_a is genetic and asserting that the underlying disposition D that it realizes is genetic, *amount to exactly the same claim* – namely, that $T(G^*) = T^*$.

All this, recall, was on the assumption that G^* does not possess D . Now suppose by contrast that G^* *does* possess D . It would follow that G^* implies T_a when E_a , and hence that $T(G^*) \neq T^*$. In other words, with respect to the explanandum T_a -rather-than- T^* , having G_a rather than G^* no longer makes a difference and thus appealing to genes is no longer explanatory. In such circumstances, the assertion that D is genetic no longer serves any pragmatic purpose, as D is a property of G and G^* alike. For that reason, such assertions are only made in the earlier case, i.e. when $T(G^*) = T^*$.

Turn next to the second half of the story, so to speak, to when the salient contrast class for the explanans is environmental, i.e. some E^* . The key point is that a genome either will or will not have a given disposition – *regardless* of environment. Intuitively, for instance, Mozart would still have been conceived with his musical talent regardless of whether his upbringing subsequently enabled him actually to fulfill it. Thus G_a will still have disposition D given either E_a or E^* . Recall, the definition of D is: T_a when E_a . To say that D holds given E^* is therefore merely to assert the conditional ‘ T_a when E_a ’ when that conditional’s antecedent is false, i.e. when E^* . Accordingly, this assertion will be (vacuously) true for *any* E^* . It follows that possession of D in itself implies nothing about what would have happened given E^* . In particular, it is therefore left open whether or not $T(E^*) = T^*$.

To appreciate the significance of that, recall in turn that when an E^* is salient, [GT] tells us that asserting a trait to be genetic amounts to the claim $T(E^*) \neq T^*$. But that

latter claim is not true for Mozart, for instance. As we saw, given that Mozart's musical feats depended on his childhood hot-housing, for a typical E^* we find on the contrary that $T(E^*) = T^*$. To capture what is genetic about Mozart's talent we must contrast only alternative genomes, *not* alternative environments. Formally, his case satisfies only one half of our definition of a genetic trait, namely the first half. But we see now that it satisfies *all* of the definition of a genetic *disposition*, because the latter is conveniently silent about $T(E^*)$. And that is why, when expressing the genetic aspect of Mozart's talent, we appeal to disposition rather than trait.

The pay-off from this, finally, is that it now shows us exactly when and why appealing to a genetic disposition is useful. In particular, we assert a disposition rather than a trait to be genetic *iff* the following pragmatic (1 and 4) and metaphysical (2 and 3) conditions are satisfied:

- 1) The explanatory context does not make it clear that G^* rather than E^* is the salient contrast
- 2) G^* does not possess D , i.e. $T(G^*) = T^*$
- 3) $T(E^*) = T^*$
- 4) We wish only to assert $T(G^*) = T^*$ [GD]

Here is the reasoning. First, often we shall want to assert $T(G^*) = T^*$ (assuming it is true) as doing so may have scientific value, i.e. often conditions 2 and 4 hold. And often we achieve this simply by asserting the *trait* T_a to be genetic since, when a G^* is salient, that is just to assert that $T(G^*) = T^*$. But given conditions 3 and 1, this usual strategy fails, since asserting a trait to be genetic is also to assert that $T(E^*) \neq T^*$ in E^* cases, and this latter implication now becomes both false and potentially salient. In such

circumstances we may still assert $T(G^*) = T^*$, but now only by asserting instead that the underlying *disposition* is genetic. Given condition 2, asserting the disposition to be genetic implies that $T(G^*) = T^*$. In G^* cases, this merely replicates the implication of asserting the trait to be genetic. But given also conditions 3 and 1, the situation changes, as now the silence of genetic dispositions regarding $T(E^*)$ – which distinguishes them from genetic traits – becomes decisively useful.

Thus the metaphysical implications of an appeal to genetic dispositions are derivable only indirectly, via pragmatic considerations. Intuitively, *we appeal to dispositions when a trait is explained (in that context) by both genes and environment but we want to focus attention just on the genes side.*

The best evidence for this account is again examination of how actual usage tracks explanatory context in just the way predicted. With Mozart, often $T(G^*) = T^*$ and $T(E^*) = T^*$, i.e. conditions 2 and 3 hold. In many contexts, such as when discussing his hot-housing, it will be unclear that G^* rather than E^* is salient, i.e. condition 1 holds too. And Mozart's uniqueness, finally, is captured by $T(G^*) = T^*$, where G^* = other human genomes. If we wish to assert that uniqueness then we are satisfying condition 4, and hence now all four conditions. Thus it is that we naturally appeal to Mozart's 'innate genius' or 'God-given talent', i.e. appeal now to a genetic disposition rather than a trait.

Here is another, more biological example. We would typically say that my having two feet is a genetic trait. But if it is pointed out that thalidomide in the womb during early pregnancy may lead to no feet developing, we revise our claim to saying only that I have a natural *tendency* to develop two feet, i.e. now appealing to a genetic disposition. Why this switch? Because, on this paper's view, in the typical conversational context we

do not have in mind those unusual environments in which two feet will not develop. Rather, we have in mind some E^* such that $T(E^*) \neq T^*$ (for $T^* = I$ do not have two feet). But in a thalidomide context, now $E^* =$ thalidomide in the womb, and so $T(E^*) = T^*$, i.e. the environment becomes explanatory. In both cases, genes are explanatory, i.e. $T(G^*) = T^*$ (say, for $G^* =$ a non-human genome). Formally, the thalidomide context, but not the typical context, fulfills [GD]'s conditions 2 and 3. Mere mention of thalidomide presumably may suggest E^* rather than G^* to be salient, thus fulfilling condition 1. Therefore if condition 4 also holds, i.e. if we wish to assert that my having a human genome explains my two feet, we appeal to the tendency/disposition. Thus our account successfully explains the original datum, namely that when switching from the typical to the thalidomide conversational context we switch from trait to disposition.²⁵

One more example, briefly: Tiger Woods hits the ball further than most other professional golfers. Announcers often mention his 'natural length', i.e. a genetic trait. But when comparing the constantly training adult Woods to his ten-year-old childhood self, thus switching to a context in which an environmental input (i.e. his training) is now explanatory too, the comment becomes instead how he has fulfilled his 'natural talent', i.e. a genetic disposition.

In distinguishing between genetic traits and dispositions, this final section has staked out virgin territory. Yet how else to explain why when regarding Mozart we invoke genetic dispositions, but when regarding my legs we usually invoke genetic traits? Our account explains why appeal to dispositions is likely when $T(E^*) = T^*$ and when G^* does not possess D. Prime generators of such contexts are human traits that are non-

²⁵ We are also licensed to say that, in contrast to some hypothetical thalidomide-resistant creature, humans have the unfortunate genetic disposition to develop *no* feet in a thalidomide-bathed prenatal environment.

universal and environmentally sensitive. And such traits often turn out to be precisely the subjects of famous disputes. Besides musical ability, examples include homosexuality, alcoholism, schizophrenia, high scores on IQ tests, athletic ability, and many cancers. A good analysis of genetic dispositions is therefore relevant to precisely those hot-button controversies that are one of the main things we want an account of genetic traits *for*.²⁶

Acknowledgments

I would like to thank an anonymous referee and especially Katie Plaisance for useful comments. I would also like to thank Katie, and Thomas Reydon, for organizing the stimulating conference at which this paper was presented.

References

- Achinstein, P. (1983). *The Nature of Explanation*. (Oxford: Oxford University Press.)
- Davidson, D. ([1967] 1980). 'Causal relations', in *Essays on Actions and Events*, 149-62.
- Dretske, F. (1972), 'Contrastive Statements', *Philosophical Review* 81, 411-437.
- Gannett, L. (1999), 'What's in a Cause?: the pragmatic dimensions of genetic explanations', *Biology and Philosophy* 14, 349-374.
- Garfinkel, A. (1981), *Forms of Explanation*. (Yale University Press.)
- Gifford, F. (1990). 'Genetic Traits', in *Biology and Philosophy* 5, 327-347.
- Hitchcock, C. (1996), 'The Role of Contrast in Causal and Explanatory Claims', *Synthese* 107, 395-419.
- Hitchcock, C., and J. Knobe (2009). 'Cause and norm', *Journal of Philosophy* 106.
- Hull, D. (1981). 'Units of Evolution: A Metaphysical Essay' in U.L. Jensen and R. Harre (eds.) *The Philosophy of Evolution*. (Brighton: Harvester Press.)
- Lewontin, R. (1974). 'Analysis of variance and analysis of causes', *American Journal of Human Genetics*, 400-411.
- Maslen, C. (2004), 'Causes, Contrasts and the Nontransitivity of Causation'. (In J. Collins, N. Hall, and L.A. Paul (Eds.), *Causation and Counterfactuals* (pp341-57). Massachusetts: MIT Press.)
- Mathews, W.S.B. (1891). *A popular history of the art of music, from the earliest times until the present*. (Text available on Google books.)
- Northcott, R. (2006). 'Causal efficacy and the analysis of variance', *Biology and Philosophy* 21, 253-276.

²⁶ For better or worse, moral charge is attached to some environmental explanations but rarely to genetic ones. This paper's analysis explains why there is therefore motivation (for some) to describe homosexuality or alcoholism as genetic dispositions rather than genetic traits even in advance of the full scientific story – because we are thereby leaving open the possibility of an environmental explanation.

- Northcott, R. (2008a). 'Can ANOVA measure causal strength?', *Quarterly Review of Biology* 83, 47-55.
- Northcott, R. (2008b). 'Causation and contrast classes', *Philosophical Studies* 39, 111-123.
- Oyama, S., P. Griffiths, and R.D. Gray (2001). *Cycles of Contingency. Developmental Systems and Evolution*. (Cambridge, MA: MIT Press.)
- Pearl, J. (2000), *Causality*. (New York: Cambridge University Press.)
- Schaffer, J., (2005). 'Contrastive Causation', *Philosophical Review* 114.3, 297-328.
- Sesardic, N. (2005), *Making Sense of Heritability*. (Cambridge: Cambridge University Press.)
- Shipley, B. (2000). *Cause and Correlation in Biology*. (Cambridge University Press, Cambridge.)
- Smith, K. (2007), 'Towards an Adequate Account of Genetic Disease'. (In Kincaid & McKittrick (Eds.), *Establishing Medical Reality: Essays in the Metaphysics and Epistemology of Medicine* (pp. 83-110). Dordrecht: Springer.)
- Sober, E. (1988), 'Apportioning causal responsibility', *Journal of Philosophy* 85: 303-18.
- Sober, E. (1998), 'Innate Knowledge'. (In E. Craig and L. Floridi (eds), *Routledge Encyclopedia of Philosophy* (pp794-7). Cambridge: Routledge.)
- Spirtes, P., C. Glymour, and R. Scheines (2000), *Causation, Prediction, and Search* (2nd edn). (Cambridge, MA: MIT Press.)
- Sterelny, K., and P. Kitcher (1988), 'The Return of the Gene', *Journal of Philosophy* 85, 339-361.
- Sterelny, K. (2003). *Thought in a Hostile World*. (Oxford: Blackwell.)
- van Fraassen, B. (1980), *The Scientific Image*. (Oxford: Oxford University Press.)
- Waters, C. K. (1994), 'Genes Made Molecular', *Philosophy of Science* 61, 163-85.
- Waters, C. K. (2007), 'Causes that make a difference', *Journal of Philosophy* 104, 551-579.
- Woodward, J. (2003). *Making Things Happen*. (Oxford: Oxford University Press.)