Genetically caused trait is an interactive kind

Riin Kõiv

Department of Philosophy, University of Sydney

Institute of Philosophy and Semiotics, University of Tartu

Email: riin.koiv@sydney.edu.au, riin.koiv@ut.ee

Abstract

In this paper I argue that the extent to which a human trait is genetically caused can causally depend upon whether the trait is categorized within human genetics as genetically caused. This makes the kind *genetically caused trait* an interactive kind. I demonstrate that this thesis is both conceptually coherent and empirically plausible. I outline the core rationale of this thesis and demonstrate its conceptual coherence by drawing upon Waters' (2007) analysis of genetic causation. I add empirical plausibility to the thesis by describing a hypothetical but empirically plausible mechanism by which the fact that obesity is categorized as genetically caused within human genetics increases the extent to which obesity is in fact genetically caused.

Keywords

genetic causation; human genetics; actual difference makers; interactive kinds; looping effect

Acknowledgements and funding information

1. Introduction

Empirical research into the genetic basis of human traits is thriving. There is a growing body of knowledge about the degree to which genes causally influence human psychology, behaviour, social traits, metabolic process, biometric and physiological attributes. For example, we know that height and body mass index are to a high degree genetically caused, that educational attainment is somewhat less genetically caused, and that stomach cancer is genetically caused to a negligible degree (Czene et al., 2002; Silventoinen et al., 2003, 2020). There is also a growing body of knowledge about which specific genes causally contribute to these and other traits. Let's say that if a trait has genetic causes in the sense studied in human genetics research then this trait belongs to the kind *genetically caused trait*. In this paper, I will argue that whether, and to what extent, a trait is in fact genetically caused (or caused by some specific genes) can causally depend upon whether or not the trait is categorized as genetically caused, and known to be so categorized by the carriers of the trait. This makes *genetically caused trait* an interactive kind. "Interactivity" refers to the feedback loop that arises when members of a kind are influenced by classificatory beliefs about the kind in the manner that changes the kind itself, and this in turn calls for change in the classificatory beliefs about the kind. This phenomenon has been extensively discussed in philosophy in relation to the social and psychological sciences as many of the human kinds studied by such sciences have been argued to have this interactive feature (for discussion of interactive kinds see: Allen, 2021; Cooper, 2004; Hacking, 1999, 2007; Hauswald, 2016; Khalidi, 2010; Kuorikoski & Pöyhönen, 2012). For example, individuals given a psychiatric diagnosis, and thereby categorized under a particular mental disease category, might change their self-perception and behaviour in light of the diagnosis so as to not comply (or sometimes comply better) with the diagnosis. In consequence, the theories referring to the kind must be updated in light of such changes. I argue that in a similar manner, the kind *genetically caused trait* is interactive: individuals who learn that a trait they carry has genetic causes might change their attitudes and behaviour towards the trait so that the extent to which the trait in fact has genetic causes changes.

My thesis has two components. First, I aim to show that the idea that *genetically caused trait* is an interactive kind is conceptually coherent. I will draw upon Waters' (2007) influential account according to which a trait is genetically caused in the empirically relevant sense insofar as genes are actual difference making causes of the trait. I explain why and how the thesis that *genetically caused traits* is interactive is consistent with this account.

Secondly, I aim to show that the idea that *genetically cased trait* is interactive is empirically plausible – that under certain circumstances it is likely that a trait's being categorized as genetically caused by human genetic research will change the degree to which the trait is in fact genetically caused in the sense of interest to this research. To show this, I refer to empirical work on lay beliefs regarding genetic causation. Because while the kind that I argue is interactive is the kind tracked by the *scientific* concept of having genetic causes, the (often mistaken) lay beliefs about what it means for a trait to have genetic causes are a relevant component in the mechanism that accounts for the interactivity of this kind. More specifically, I refer to the work of Dar-Nimrod and colleagues on how lay people are likely to interpret, and react to, the empirical finding that a trait is genetically caused (e.g., Dar-Nimrod et al.,

2021; Dar-Nimrod & Heine, 2011). Dar-Nimrod and colleagues argue that lay people hold essentialist attitudes towards traits they believe to be genetically caused and, in consequence, tend to behave fatalistically in relation to such traits. Assuming this framework, I outline a hypothetical but empirically plausible toy example of how categorizing a trait as genetically caused within empirical contexts can lead people bearing the trait to behave so that the extent to which the trait is in fact genetically caused in the relevant population increases. I show this by using obesity as my example trait.

What it means for a trait to have genetic causes in the context of human genetic research has been thoroughly studied by philosophers and many agree with at least the essentials of Waters' description (e.g. Lynch 2021, Bourrat 2020). Likewise, there is ample empirical research on how lay people respond to the information that a trait has genetic causes, where much of this research is consistent with the genetic essentialism framework. Yet the implication of these two bodies of research – that being a genetically caused trait can be subject to the feedback loop characteristic of human kinds targeted within various social and psychological sciences – has not explicitly been addressed. It is relevant to do so. That the kinds studied within human sciences can interact with categories and theories about these kinds is thought to be important for various reasons. First of all, it is thought by some to undermine the objectivity and generalizability of the corresponding scientific categories and, consequently, of the theories that employ these categories (see e.g. Allen 2021). Secondly, it has normative consequences for scientific practice. That a kind is interactive implies that facts about the kind can be created by the very theories that represent these facts. This means that such theories and categories are not only subject to various epistemic norms but also answerable for creating certain facts, some of which might not be desirable. If *genetically caused trait* is interactive, these same implications will pertain to theories that appeal to the genetic causes of human traits. Thirdly, in the empirical literature it is well known that the extent to which a trait has genetic causes can vary from population to population and change in time. Which factors impact such variation is subject to ongoing research. The argument presented in this paper identifies a novel factor that might account for such variation. This said, the relevance and further consequences of my central thesis is not the topic of this paper. The main aim of this paper is to outline the general idea behind the thesis and thus pave the way for future work on the various implications of this idea.

I begin in section 2 by describing the concept of an interactive kind as it is discussed in the context of the philosophy of human sciences, and how it might apply to genetics. In section 3, I explain Waters` (2007) account of causation as "actual difference making" to provide a framework for thinking about genetic causation. This is needed to articulate what it is for a trait to be genetically caused, i.e., what constitutes the kind that I argue is interactive. I will also outline the core rationale of the thesis that *genetically caused trait* is an interactive kind. In sections 4 and 5, I flesh out this core rationale by providing an example of how the feedback-loop characteristic of interactive kinds might, and is empirically likely to, emerge in the case of *genetically caused trait.* In section 4, I introduce Dar-Nimrod's work on lay interpretations of claims about genetic causation. Assuming this work, in section 5 I describe a mechanism by which the fact that obesity is categorized as a genetically caused trait in human genetic research increases the extent to which obesity is, as a matter of fact, genetically caused. In section 6, I respond to two objections.

2. Categories, kinds, interaction

In this section I clarify the concept of an interactive kind and specify how my thesis relates to traditional discussions on interactive kinds. Let's distinguish between kinds and categories, as is typically done in the literature on interactive kinds. Categories (sometimes used interchangeably with "concepts") are devices that we, in our attempts to represent the world, use to categorize things in the world as being of the same kind, as belonging together in virtue of sharing some relevant features. Categories specify the criteria that an entity must meet in order to be of a given kind, typically by listing the features that the entity must have to be of the kind. Kinds are the things in the world that our categories refer to and are about.[1] For instance, psychiatrists use the disease category "multiple personality disorder". This category specifies that someone has multiple personality if she exhibits certain symptoms, e.g. is delusional, hallucinates, speaks in a disorganized manner etc. The disease itself that this criterion – having certain symptoms – picks out is the corresponding kind *multiple personality disorder.*

This paper concerns the kind *genetically caused trait.* In human genetics, certain criteria are used to determine whether, and to what extent, a trait has genetic causes, i.e., is genetically caused. We can think of these criteria as constituting the scientific category "genetically caused trait". With "*genetically caused trait*" I have in mind the kind that this category – the criteria used within human genetics to categorize traits as genetically caused – picks out. *Genetically caused trait* so defined differs from paradigmatic kinds that have been the focus of discussions around interactivity in at least one sense. Kinds such as *multiple personality disorder, homosexual* and other human kinds have *individuals* – either individual human beings, or instantiations of certain syndromes by individual human beings – as their members. *Genetically caused trait*, however, has traits as its members, where "trait" refers to traits as types (such as *eye colour, height, obesity, educational attainment*) rather than instantiations of traits by particular individuals (such as *Paul's height of 178 cm*, *Lisa's green eye colour*, *Adam's obesity, Silvia's education of 14 years*). It is traits as types that are categorized as genetically caused in most empirical contexts. Traits so understood can be more or less genetically caused (more on this in section 3). In this paper, I use "*genetically caused trait*" to include all those traits that are genetically caused to the degree that geneticists care to report them as such.

Some categories and the corresponding kinds are thought to interact. If they do, we call these kinds and categories "interactive". A kind *K* and the corresponding category "K" are interactive if classificatory practices, theories and beliefs concerning *K* (i.e., theories and beliefs that concern who, and in virtue of what, falls under "K") bring about changes in *K* and this in turn calls for further changes in theories and beliefs about *K*. Keeping with the tradition, I

---

[1] There are different philosophical accounts of the ontology of kinds (see for instance Khalidi 2013). My discussion is not committed to any particular one of them and is compatible with many of them.

call a mechanism that accounts for this effect a "feedback mechanism" or "feedback loop" and the effects of this mechanism "feedback effects" or "looping effects".[2]

There can be different types of feedback mechanism. For instance, feedback mechanisms can differ in terms of what kind of change employing "K" induces in *K*. In some cases, categorizing certain entities as members of *K* can change the constitutive properties of *K*. *Multiple personality disorder* is an often discussed example: People categorized as having multiple personality come to identify with the kind, this leads them to behave in ways and acquire properties that further distinguish them from other people, so that the kind *multiple personality disorder* comes to be associated with a new set of (defining) properties (Hacking 1999; Khalidi 2010). In other cases, employing "K" might change the extension of "K" in that the number of *K* instances increases or decreases as a result of employing "K". Hacking (2010) gives the following example. On his account, the pathological withdrawal syndrome epidemic among refugee children in Sweden between 2001 and 2006 was the outcome of the following process. At first rare instances of the syndrome were reported through the media. In response, more children began to imitate, and ultimately internalize, more and more of the symptoms of the syndrome, so that they became genuine instances of the syndrome.

Often, employing "K" can cause entities categorized as *K* to become better (more paradigmatic, more obvious) or worse (less paradigmatic, less obvious) instances of *K* by causing *K* members to acquire or lose some of the properties that define *K*. Changes like these also count as changes in the extension of "K". That an entity either becomes *K* or ceases to be *K* are two possible extreme outcomes of the process of acquiring or losing some of the K-defining properties. With *genetically caused trait*, we can view those traits that are more genetically caused as being "better", more paradigmatic, instances of the kind than those traits that are less genetically caused.

Alternatively, feedback mechanisms can differ in terms of their components. Paradigmatic instances of kind-category interaction are those where individuals categorized as *K* are self-*aware* of being so categorized and this awareness – combined with certain beliefs about what it means to be *K* – is part of the causal mechanism that brings about changes in *K* (as in the above example). But this need not always be the case. For instance, individuals categorized as having multiple personality disorder can change their behaviour in kind-changing ways because of how *other* people treat individuals who they believe to have multiple personality.

My thesis that *genetically caused trait* is interactive amounts to the following: whether and to what extent a trait in fact is genetically caused (or caused by some specific genes) in the sense studied in human genetic research can causally depend upon whether the trait is categorized as genetically caused in the context of this research. To

---

[2] There is some disagreement in the philosophical literature regarding if both these types of feedback mechanism should be counted as mechanisms of interactivity proper. Hacking (1999) and Khalidi (2010) are inclusive in this respect. Hauswald (2016) on the other hand thinks that *K* is genuinely interactive only if employing "K" causes *qualitative* changes in K. In this paper I will be assuming the more inclusive notion of interactivity as is often done (and with good reason on my view).

illustrate this thesis, I will, in section 5, outline an example of the following feedback mechanism that might cause this effect (see also Figure 3):

> Trait T is categorized by scientists as genetically caused in some relevant population P. This becomes known in P. Due to having certain conceptions about what it means for a trait to be genetically caused, members of P adopt essentialising attitudes towards T. Essentialising attitudes towards T lead some carriers of T in P to change their behaviour in a way that increases the extent to which T is in fact genetically caused in P. This increase is registered by scientific measures of the genetic causes of T – T is categorized as more genetically caused.

This is an instance of a feedback mechanism such that: (1) Applying "K" changes the extension of K. In this concrete example, categorizing a trait under "genetically caused trait" causes the trait to become more genetically caused and thus a better instance of *genetically caused trait*. (2) One part of this mechanism is the awareness of individuals categorized under "K" of being so categorized. In the case of *genetically caused trait* this awareness is, more specifically, awareness of the fact that a *trait* they carry is categorized as genetically caused.

In order to demonstrate how this feedback mechanism could plausibly occur, it is necessary to explain two things. First, it is necessary to explain in more detail the nature of what I argue is interactive, i.e., what constitutes the kind *genetically caused trait*. This is essential to my argument because the possibility that the feedback mechanism occurs has a lot to do with what it means to be a genetically caused in the first place. According to the definition introduced earlier, a trait is genetically caused insofar as it has genetic causes in the sense studied in human genetics research. Therefore, in order to explain what constitutes *genetically caused trait,* I need to unpack what it means for a trait to have genetic causes in the context of such research. I do this in the next section. Second, it is necessary to explain what *lay people* believe it means for a trait to have genetic causes. For as said, even though the kind that I argue is interactive is the kind tracked by the *scientific* concept of having genetic causes, lay beliefs about what it means for a trait to have genetic causes – essentialist beliefs in particular – are one component in the mechanism that accounts for the interactivity of this kind. I describe essentialist lay conceptions of being genetically caused in section 4.

3. **Genes as actual difference making causes**

In this paper, I defer to Water's (2007) account of what it means for genes to cause a trait in the context of an empirical claim that a (human) trait has genetic causes. Waters argues that much existing genetic research, including human genetic research, is interested in whether genes cause a trait in the sense of causing actual differences in the

trait. [3] Correspondingly, a trait is said to have genetic causes in the context of such research insofar as genes are among the "actual difference making causes" of the trait. To clarify what this means, I begin with a sketch of the methods used in human genetic research to decide if a trait has genetic causes – so as to have a better view on what these methods are meant, and in a position, to identify.

Most of the existing knowledge of the genetic causes of human traits comes from observational studies that operationalize genetic causation as a statistical association between a trait and a genome.[4] A trait T and a genome G are statistically associated if some version of T is possessed by individuals with particular versions of G significantly more frequently than individuals with some different versions of G. The relevant genetic unit (referred to with "genome" or "G") can vary from method to method – it can be a single base pair, a gene (given some meaning of "gene"), haplotype, or whole genome. Throughout the paper I use "G", "genome" and sometimes "gene" to refer to whatever genetic unit may be of interest in a given study. I use "genotype" or "g^*" to refer to an allele of a genome or a gene.

Consider a simple toy example. Let our trait be obesity (represented by variable O). And let the trait come in two versions: obese (represented by value o^+ of O) and not-obese (represented by value o^- of O).[5] Population 1 (in Figure 1) depicts an imaginary population where there is association between G and O (which, let's stipulate, is statistically significant).

Figure 1

Population 1. G-O association.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
G	g^1	g^1	g^1	g^1	g^2	g^2	g^2	g^2

---

[3] Various accounts have been proposed to articulate what it means for genes to cause a trait in empirical contexts (Bourrat, 2019, 2020; Gannet, 1999; Lynch & Bourrat, 2017; Waters, 2007; Weber, 2017; Woodward, 2010). I take Waters' account to be compatible with most of them.

[4] Methods used in such studies include twin, family and adoption studies, linkage studies, candidate locus studies, genome wide association studies (GWAS). The details of what these different methods show of a trait when showing the trait to have genetic causes can vary along many dimensions. These details don't concern us. See Lynch (2021) for a more fine grained discussion of the content of "has genetic causes" across different research contexts.

[5] The example is not far-fetched. Obesity is often reported to have significant genetic causes (Chami et al., 2020; Loos & Yeo, 2022; Namjou et al., 2021; Wang et al., 2011). Also, studies into the genetic causes of obesity often treat the trait as a binary trait where an individual counts as obese if her body mass index is higher than 40, and not obese otherwise (e.g., Wang et al. 2011). Be it stressed, however, that nothing in my argument depends upon whether a trait is binary or continuous. I have chosen to use a binary trait as my example for simplicity of presentation.

| O | o$^+$ | o$^+$ | o$^-$ | o$^-$ | o$^-$ | o$^-$ | o$^-$ | o$^-$ |

Population 2. No G-O association.

	1	2	3	4	5	6	7	8
G	g^1	g^1	g^1	g^1	g^2	g^2	g^2	g^2
O	o$^-$	o$^-$	o$^-$	o$^-$	o$^-$	o$^-$	o$^-$	o$^-$

Figure 1: Two populations of 8 individuals are compared. Each individual either has genotype g^1 or genotype g^2, and is either obese (o$^+$) or not obese (o$^-$). In Population 1, there is an association between O and G. In Population 2 there is no association between O and G. To add a touch of realism to the example, we can think of the 8 individuals as datasets of individuals. We can also take "g^1" to stand for a genotype that comprises some sufficiently large set of all those alleles at different loci on the human genome that are known to increase the risk of obesity. We can take "g^2" to stand for a genotype that does not comprise such a set.

Association between a trait and a genome (O and G) is of course not yet causation but a *test* for causation. What interests us is what such association is a test *for* – what is this thing called "causation" that significant genome-trait associations are meant to detect and, if successful, in fact do detect? This much is clear without theory that detecting gene-trait associations is meant to detect a causal relation between genes and a trait insofar as it is meant to detect whether the instantiation of a given value of a trait *depends* upon which genotype one carries.[6] Waters' (2007) actual-difference-making account of genetic causation specifies the nature of the relevant kind of dependence. Waters' account explicitly builds upon James Woodward`s influential version of counterfactual account of causation (known as "interventionism"). Thus, an outline of Woodward's core idea is needed before I can move on to Waters' application of this account to genetic causation in particular.

Woodward casts causation as a relationship between two variables (anything that can take on at least two different values). According to Woodward, one variable X causes another variable Y if the following – call it "Woodward's criterion" – is true:

---

[6] As is well known, for example in the case of GWAS, a genetic marker G found to be associated with T need not be itself causally related to T. Instead, G-T association might be explained by the fact that G is linked to some other "gene" G' that is causally related to T. Therefore, strictly speaking, G-T association is not a test for whether *G* causes T but whether G *or* some other gene G' in the vicinity of G causes T. This nuance does not bear upon my argument.

There are background circumstances B such that if some (single) intervention that changes the value of X (and of no other variable) were to occur in B, then the value of Y or the probability distribution of Y would change.

In the context of this criterion, "background circumstances" refers to all those parts of the context of a (possible) intervention on X with respect to Y that are not part of the X-Y relation. "Intervention" is a technical term for a specific kind of manipulation (changing) of the causal variable.[7] As the technical meaning of "intervention" plays no role in my argument, I will not explicitly use the term in the following analysis. Instead, I will be simply talking about "changing the value of X", tacitly assuming that a given instance of such changing qualifies as an intervention in Woodward's sense.

But notice that Woodward's criterion is too permissive to provide an adequate explication of what is meant by "genes cause a trait" in the context of an empirical finding that genes cause a trait. As per Woodward's criterion, G counts as a cause of T whenever there exists but *one* possible background circumstance b*, *one* pair of possible G values, g* and g**, and *one* pair of possible T values, t* and t**, such that if an individual with t* and g* would have g** then the individual would have t**. For any trait, we can find a genetic variable, and a background circumstance, of which this is true. For example, consider the trait *speaks Estonian* with two values "speaks Estonian" and "does not speak Estonian". It is true of most actual adult Estonian speakers that if instead of their actual genotype they had had a certain mutation in the genetic region associated with Hutchinson-Gilford syndrome, then they would not speak Estonian because they would have died in their teens and would not exist. Yet, "speaking Estonian" is not a trait that would be called genetically caused in any sense of interest in human genetic research – at least not for the reason cited. Similar examples can be constructed for all traits. Moreover, Woodward's criterion trivially renders all traits genetically caused because it is true of all traits that if G, understood as a whole genome, was made to have the value "absent" then each and every trait, whatever its prior value, would also have the value "absent" – without a genome there is no organism, therefore no trait instantiations. Yet, gene-trait association studies only identify *some* traits as genetically caused (to some significant degree). This suggests that when these studies show a trait to have genetic causes, they show something more specific about the trait than that the trait relates to some genes as described by Woodward's criterion. This suggests that in the context of human genetic research, not all possible T and G values and not all possible background circumstances can be relevant for determining whether genes cause a given trait. The question is: which ones are?

Waters (2007) proposes an answer. He argues that the various association methods used in genetic research are designed to identify a subset of those G-T relations that meet Woodward's criterion, a subset that meets Woodward's criterion *for the values* of G, T and B that are actually instantiated in some actual population. Here's what that means. In principle, any G and T can have many different values and whether a specific change in the value of G would result in a change in the value of T can be assessed against various possible background

---

[7] An intervention on *X* with respect to *Y* is a manipulation of *X* such that the manipulation changes the value of *X* without changing, independently of the change in the value of *X,* the value of any other causes of *Y* (here paraphrased from Waters 2007, 12; see also Woodward 2003, 98).

circumstances. However, an observational study into the genetic causes of a trait always targets some concrete actual population. And in an actual population, we typically only some of the possible background circumstances obtaining, and only some of the possible values of T and G are actually instantiated by the members of the population and distributed in a certain way. Which population is in question varies from research context to research context: it can be some "natural population" (like the Finnish population or the Caucasian population), some relevant subset of individuals from a natural population (e.g., Finnish men or Finnish men with higher education born between 1940 and 1950), individuals dwelling in a given geographic location, some "time-slices" of a relevant group of individuals (e.g., Finns aged between 35 and 40), and so on. Depending on what the target population is, the actually instantiated values of B, G and T, and their distribution, can differ. In the two populations in Figure 1, two G values are instantiated: g^1 and g^2. But in some different population yet a different value of G, g^3, might be instantiated. As for O, in our example we construed O as a binary trait with only two values: obese/not-obese. Given this, the O values instantiated in Population 1 exhaust the possible O values, but not in Population 2 where only o⁻ is instantiated.[8] Now, Waters argues that a typical genetics study seeks and provides knowledge about whether G causes T according to Woodward's criterion, given the values of G and T that are actually instantiated in the target population (and represented in the data), and given the background circumstances that actually obtain in this population. If this is so, G is what Waters calls "an actual difference making cause" of T – a cause that causes actual differences in T in the relevant population.

Apply all this to our example: if an empirical study shows that in Population 1 G is associated with O, we don't merely learn from this that for *some* background circumstances b*, *some* values g* and g** of G, and *some* values o* and o** of O, and in *some* background circumstance b* it is true that if an individual had g* instead of g** then the individual would have o** instead of o*; nor do we learn that this is the case for *all* values of G, O, and B. Instead, we learn something both more specific and less general: that the above counterfactual is true of those values of G, O and B that are actually instantiated in Population 1 (g^1, g^2, o⁺ and o⁻, b^{actual}). This of course does not rule out the possibility that setting G to have a value that is *not* instantiated in Population 1 (e.g., g^3) would change the value of O, nor that changing the value of G would change the value of O in some background circumstances that do *not* obtain in Population 1. But this need not be, and often is not, the case (as I show below).

So, when the claim is made in a given scientific context that some trait has genetic causes, I will understand it to mean that genes are among the actual difference making causes of the trait, in the sense outlined. Note that on this account, being a trait that has genetic causes is much less inclusive than if one defined being a genetically caused trait in terms of Woodward's criterion. While it is trivially true that with all traits Woodward's criterion holds for *some* values of G, T and B, it is not trivially true that it holds for those values that are actually instantiated in some actual population. Whether that is the case is an empirical question. Moreover, it is also a non-trivial empirical

---

[8] Or suppose that we had construed O in a more fine-grained way as having four values (as is sometimes done): not obese (BMI < 30), class 1 obese (BMI of 30 to < 35), class 2 obese (BMI of 35 to < 40), class 3 obese (BMI of 40 or higher). Had we done so, we might as well had found that in Population 1 only "not-obese" and "class 1 obese" are instantiated.

question *what proportion* of the actual trait differences are caused by genes in a given population (which is something that I turn to in a moment).[9] It is these questions that empirical research into the genetic causes of human traits provides answers to.

With this I have articulated what constitutes the kind *genetically caused trait*: *genetically caused trait* is a trait such that genes are actual difference making causes of this trait in the sense articulated by Waters. It is *genetically caused trait* so understood that I argue is interactive.[10] My argument will make use of two implications of this account of being a genetically caused trait. First, on this account, being a genetically caused trait is a matter of degree: genes can cause more or less of the actual trait differences in a population. This "more or less" can be fleshed out along many dimensions. The dimension I will make use of is the following. For a course-grained division, let's say that G can cause all or only some actual T differences in some population P. G causes *all* T differences in P if T and G are instantiated with different values in P and it is true that if every individual in P had the same G value g* (any value actually instantiated by one of these individuals), and keeping everything else fixed as background circumstances, then all individuals would instantiate the same T value t* (a T value actually instantiated by this individual). G causes *some* T differences in P if T and G are instantiated with different values in P and it is true that if every individual in P had the same G value g* (a G value actually instantiated by one of these individuals), and keeping everything else fixed as background circumstances, then the actual differences in T would change in P but would not be eliminated.[11] The fact that in Population 1 two carriers of g^1 (3 and 4) do not have o^+ like most g^1 individuals suggests that if everyone in this population had g^1 (keeping everything else unchanged) then O differences would change – plausibly, more individuals would instantiate o^+ – but would not be eliminated (*mutatis mutandis* for g^2 and o^-); it therefore suggests that in Population 1 G causes some (and not all) O differences.

The second important implication of Waters' account is that the degree to which a trait is genetically caused can vary across populations only *in virtue of different background circumstances*. Here's an illustration. Take for

granted that in Population 1, given the background circumstances b^{actual} that actually obtain in this population, G causes some actual O differences. We have said nothing about what b^{actual} consists in in Population 1. However, whatever b^{actual} is, let's suppose that instead of b^{actual}, b* would have obtained in Population 1: b* = all individuals in Population 1 have consumed no more calories than is necessary for normal biological functioning. Consuming a certain excess amount of calories is biologically necessary for anyone to have o^+. Therefore, if b* obtained in Population 1, both g^1 and g^2 individuals would all have o^- as in Population 2 (Figure 1). But notice that this change in background circumstances has not merely changed facts about the frequency of o^+ among g^1 and thus facts about how well G and O are associated. The extent to which O is associated with G has changed because facts of *causation* have changed. In Population 1 it is true (as we assumed) that if everyone had the same genome, say, g^1, then the distribution of O in Population 1 would change: some of the g^2 individuals who currently have o^- would have o^+. In Population 2 this is not the case. If everyone in Population 2 had had g^1 then everyone would have consumed very few calories just like they actually did, and everyone would have o^- just like they actually do. This means that in Population 1 G causes *some* actual O differences, whereas in Population 2 G causes *no* actual O differences, despite the fact that genetically these two populations are identical.

That traits can be genetically caused to a different degree and this degree can vary with background circumstances is well-known in the empirical literature. It is reflected, for example, by different heritability estimates for different traits, and different heritability estimates for the same trait in different populations. For instance, the heritability of height is known to be higher in richer populations compared to poorer populations even where genetically these populations do not differ (Silventoinen et al., 2003). The heritability of many social outcomes is higher in politically liberal societies compared to authoritarian societies, again despite the genetic similarity of these societies (Rimfeld et al., 2018; Uchiyama et al., 2021). It is also known that that which particular genetic loci causally contribute to trait differences, and what proportion of all trait differences a given locus explains, varies from population to population with background circumstances (Mathieson, 2021; Matthews, 2022; Mostafavi et al., 2020). The actual difference making account of genetic causation makes clear that such variable estimates are indeed estimates of genetic *causation* and that the possibility and plausibility of such variation is written into the very concept of being a cause that is operative in empirical studies.

My thesis that *genetically caused trait* is interactive amounts to the claim that a shift in whether T is categorized as caused by G, and broadly believed to be so categorized, can constitute the relevant shift in background circumstances that changes facts about whether, and what proportion of, the actual T differences G actually causes in a given population. As just explained, it is built into the concept of being an actual difference making cause that the extent to which G causes T in a population depends upon which background circumstances obtain in this population. That the relevant background circumstances can consist in the beliefs of the members of the target population has been empirically demonstrated (Burt, 2022; Mezquita et al., 2018; Rimfeld et al., 2018). I will now expand upon the possibility that the relevant background circumstances consist more specifically in beliefs about whether or not the relevant trait is categorized as genetically caused. As a first step in doing this, I need to address empirical research on lay attitudes towards genetic causation.

4. Genetic essentialism

Having explained what it means for a trait to have genetic causes in the context of human genetic research, I now turn to what *lay* people think it means for genes to cause a trait and, correspondingly, how they interpret scientific reports that a trait has genetic causes. Let it be noted that the two need not align. Multiple factors have been shown to impact lay people's assessment of the implications of the claim that a trait is genetically caused (see Lynch et al. 2021 for an overview). However, one factor that appears to be salient and have a relatively stable impact on lay interpretations of genetic findings is genetic essentialism. Namely, empirical research on the lay concept of genetic causation suggests that the way lay people conceive of genes and genes' relation to traits expresses a more general well-evidenced psychological bias called "psychological essentialism" (Cheung et al., 2014; Dar-Nimrod et al., 2021; Dar-Nimrod & Heine, 2011; Gould & Heine, 2012; Heine, 2016; Heine et al., 2017). Psychological essentialism refers to the assumedly universal implicit human tendency to think of biological organisms, including humans, as possessing an invisible causally potent inner "essence" (or "nature", as it is sometimes called) (Berent, 2020; Gelman, 2003, 2009; Gelman & Wellman, 1991; Keil, 1989; Medin & Ortony, 1989). This inner essence is viewed as something that an organism inherits from its parents, that the organism shares with other organisms of the same kind, that defines the organism as the kind of organism that it is, that is developmentally fixed, and that survives changes in the organism's superficial properties. As a manifestation of this tendency, we, humans, are prone to view some traits of organisms as emerging from this inner essence. As the essence itself, we view such "essence-caused" traits as developmentally fixed, biologically inherited, difficult to manipulate by experiential intervention etc.

Importantly, this lay concept of inner essence appears to be a placeholder concept. People universally share the belief that there is *something* within the organism that plays the role of inner essence but need not have beliefs about *what* this something is. At different times in different contexts, different things are believed to play the role (e.g., blood, heart) (Gelman, 2003; Medin & Ortony, 1989). Dar-Nimrod et al. argue that in modern societies, laypeople tend to view genes as the material carriers of an organism's inner essence. Correspondingly, they tend to view the traits they believe to be genetically caused as caused by this essence. This is evidenced by the observation that people attribute to genes and to traits they believe to be genetically caused the very same characteristics they associate with "essences" and "essence-caused" traits (see Dar-Nimrod and Heine 2011 for a review). Upon hearing that a trait has genetic causes lay people are likely to infer that the development of the trait is to a significant degree predetermined, unavoidable and that, once developed, the trait is difficult to change. Notably, whether this inference is drawn seems to be insensitive to information about the strength of genetic influence on the trait (Heine 2016).[12]

---

[12] This intuitive distinction between features that originate from the internal essence of the organism and those that are of external, experiential, origin is argued to underlie the lay categories of "natural" versus "non-natural" (Haslam

Essentialising interpretations of findings in genetics have also been shown to induce certain systematic behavioural responses, for instance, fatalistic behaviour. Here is an example. Dar-Nimrod et al. (2011, 2014) conducted an experiment to investigate people's behavioural response to exposure to scientific claims regarding the existence of genes associated with obesity.

> Participants read one of three different articles: an article describing evidence for an "obesity gene," an article describing evidence for how environmental factors (specifically social networks) relate to obesity, or a neutral article. Following the manipulation, participants took part in an experiment that purported to investigate their food preferences; they were provided with some cookies to evaluate. Those participants who learned of the existence of obesity genes subsequently consumed more cookies than participants in either of the two other conditions (which did not differ from each other). In this instance, it seems that people's default explanation for obesity is that it is under an individual's control, however, when exposed to a genetic argument people appear to discount relevant variables such as their own eating behaviors, suggesting an increase in their deterministic perceptions of one's weight. (Dar-Nimrod et al. 2011; these results were later published in (Dar-Nimrod et al., 2014))

The authors take the following mechanism to be at work here. Subjects interpreted the information that obesity has genetic aetiology as implying that whether or not one becomes obese is determined by one's "essence" and therefore is difficult to prevent; this made them adopt fatalistic attitudes towards their weight; this led them to control their calorie intake less than prior to being primed with information about the genetic causes of obesity. Dar-Nimrod et al. argue that this response is representative of a more general tendency in how lay people respond to the (scientific) information that a trait is genetically caused.

Such essentialising fatalist representation of scientific reports of genetic causation are, typically, misguided. That genes cause some or all actual differences in trait in some actual population has none of the above-described essentialist implications (see Dar-Nimrod&Heine 2011 for discussion). However, as I will now show, these often misguided lay representations can play a role in a process that ends up actually changing facts about the extent to which genes cause actual differences in a trait. The next section describes one possible, and empirically plausible, course of events whereby – in the context of essentialist attitudes towards genetically caused traits – the fact that obesity is categorized as genetically caused in a scientific context increases the extent to which genes cause actual obesity differences in a population.

5. **The interactivity of *genetically caused trait*: an example of a feedback-loop.**

---

et al., 2000). Given that more and more "natural" tends to be equated with "genetic", it is not surprising that both categories are associated with the same essentialist beliefs.

Suppose that Population 1 is our target population (see Figure 1 or Figure 2). Also suppose that in Population 1 G indeed causes some actual differences in O. And suppose that a genetics study shows this to be the case. The finding that O is genetically caused is broadly advertised in Population 1 and knowledge of it spreads. Soon, most members of Population 1 have formed the belief "O is genetically caused". Findings described in the previous section allow us to make predictions about which further course of events is likely to unfold if this happens. The first prediction is that many of these individuals interpret the empirical claim that O is genetically caused through the essentialist lens. In order to predict which further consequences this might have, we first need to speculate about the reasons why G is an actual difference making cause of O in Population 1 in the first place – why is it that O values depend upon G values in Population 1?

We can safely assume that the reason why G is an actual difference making cause of O in Population 1 is that G somehow participates in a biological pathway that contributes to the morphological characteristics (height, mass) that the different O values supervene upon. Not much is known about the biological function of the numerous genes associated with obesity or the pathways via which they contribute to this trait. But given existent evidence, many of those genes participate in regulating appetite and hunger. Differences in such genes cause actual obesity differences because individuals with certain alleles of these genes ("large appetite alleles") tend to crave for more food than individuals with different alleles ("small appetite alleles") (Abdella et al., 2019; Larder et al., 2017; Namjou et al., 2021; Silventoinen & Konttinen, 2020). In background conditions where food is easily accessed, carriers of the large appetite allele eat more, put on more excess weight and, consequently, have o^+ more frequently than carriers of the small appetite allele.

Let's suppose that this is indeed the reason why G causes O in Population 1: g^1 is the "large appetite allele", g^2 is the "small appetite allele", g^1 individuals tend to eat more than g^2 individuals and thus become obese more frequently than g^2 individuals. Supposing this, the following course of events may be triggered when members of Population 1 learn that O is found to be genetically caused. Being genetic essentialists, many members of Population 1, both g^1 and g^2 carriers, adopt a fatalistic laissez-fair attitude towards their bodyweight. They now exercise less control over their calorie intake than they did prior to believing that O is genetically caused. But this shared response of reduced control over how much one eats has different consequences for g^1 and g^2 individuals. Carriers of g^1 (those with a large appetite) now systematically eat more than they ate prior to believing that O is genetically caused. Carriers of g^2 (those with small appetite) either eat less than they did prior to believing that O is genetically caused (if they have a really small appetite), don't change how much they eat, or eat more but to a lesser degree than carriers of g^1. If this pattern persists in the population for long enough, g^1 individuals on average end up putting on more extra weight compared to g^2 individuals (who either do not put on extra weight or do so less than g^1 individuals). More and more of the g^1 individuals therefore surpass the threshold of being obese and the proportion g^1 individuals with o^+ increases in the population.[13] Let's stipulate that by some time, *all* g^1 individuals surpass the threshold of being

---

[13] With a continuous trait like body mass index (BMI) the change would result in increased difference between the average BMI of g^1 and g^2 individuals.

obese, so that the distribution of O in our population is now as in Population 3 (Figure 2). Be it stressed that the g^1 and g^2 carriers whose actual O values account for the new distribution of O in Population 3 need not be the same g^1 and g^2 carriers whose actual O values accounted for the distribution of O in Population 1. What matters is that the proportion of g^1 individuals with o^+ in Population 3 has grown compared to Population 1, regardless of whether the g^1 individuals in Population 1 are numerically identical to the g^1 individuals in Population 1.

Figure 2

Population 1. Background circumstance: O is not categorized as a genetically caused trait.

	1	2	3	4	5	6	7	8
G	g^1	g^1	g^1	g^1	g^2	g^2	g^2	g^2
O	o^+	o^+	o^-	o^-	o^-	o^-	o^-	o^-

Population 3. Background circumstance: O is categorized as a genetically caused trait.

	1	2	3	4	5	6	7	8
G	g^1	g^1	g^1	g^1	g^2	g^2	g^2	g^2
O	o^+	o^+	o^+	o^+	o^-	o^-	o^-	o^-

Added O-differences

Figure 2. O distribution before (Population 1) and after (Population 3) O is categorized by scientists as genetically caused and broadly believed to be so categorized. We can think of Population 1 and Population 3 as two different time phases of the same superpopulation. "Added O-differences" signifies the segment of O differences in Population 3 that is not present in Population 1.

But importantly, it is not merely the frequency of o^+ among g^1 individuals and thereby the extent to which O is associated with G that has increased in Population 3 compared to Population 1. Assuming Waters` account of what the relevant notion of genetic causation is, and our story about the biological function of G, *causal* facts – what proportion of O differences G causes – have changed too. Recall the distinction made in section 3 between G causing *some* and G causing *all* actual O differences in a population. G causes *all* T differences in P if T and G are instantiated with different values in P and it is true that if every individual in P had the same G value g* (any value actually instantiated by one of these individuals), and keeping everything else fixed as background circumstances,

then all individuals would instantiate the same T value t*. G causes *some* T differences in P if T and G are instantiated with different values in P and it is true that if every individual in P had the same G value g* (a G value actually instantiated by one of these individuals), and keeping everything else fixed as background circumstances, then the actual differences in T would change in P but would not disappear. The fact that not all g^1 carriers have o^+ in Population 1 indicates that even if every individual in this population had, say, g^1 then even though it is likely that *more* individuals would have o^+, some would still have o^- (for instance, because they would have restricted their calorie intake despite large appetite). Thus, in Population 1 where O is *not* known to be categorized as genetically caused, G counts as causing only *some* of the existing O differences. However, in Population 3, G counts as causing all existing O differences. In Population 3 – where the background circumstances have changed to include the scientific finding that O has genetic causes and everyone is less motivated to control how much they eat – it is true that if all the members of the population had g^1 then everyone would have a large appetite, would eat enough to become obese, and, consequently, would have o^+. Thus, categorizing O as genetically caused has increased the extent to which O *is* genetically caused in the empirically relevant sense – it has caused O to become a better, more paradigmatic, instance of *genetically caused trait*. If this change gets registered by empirical studies, the loop is reinforced.

Of course, it is extremely unlikely that in a natural population (such as, say, Finnish population) a shift in whether O is categorized as genetically caused would result in G causing all actual O differences. If only for the reason that it is extremely unlikely with any complex trait that genes cause all actual differences in the trait in natural populations. However, this is beside the point. First, the purpose of the example is to demonstrate *how* the scientific practice of categorizing a trait as genetically caused might change the degree to which the trait is in fact genetically caused, and not how big the change is likely to be. Second, we can easily think of Population 1 and Population 3 as those subpopulations of some natural population that Figure 2 *does* accurately describe. That such (even if tiny) subpopulations exist is reasonably plausible given the empirical premises that the above example built upon. We can even add to this plausibility by assuming that G is pleiotropic for self-control: g^1 not only increases appetite but also reduces self-control (see e.g., Meyre et al., 2019). If so, then not only are g^1 individuals prone to crave after more food than g^2 individuals, but they are also less likely to resist their cravings. This will magnify the effect of learning O to be genetically caused in terms of g^1 individuals eating more than g^2 individuals. That G goes from causing some to causing all O differences in a subpopulation manifests in the superpopulation as G going from causing some to causing more (but not necessarily all) actual O differences (see also Waters 2007, 21).

This toy example exemplifies one type of mechanism by which a trait's being categorized as genetically caused can change the degree to which the trait is in fact genetically caused in the sense of genes causing actual differences in the trait in some relevant population. Although I used the example of obesity, the same kind of mechanism could also be operating on other (quantitative and qualitative) traits. Plausible candidates include psychological, behavioural and disease traits such that: (a) these traits are in fact partly genetically caused in some population and (b) the influence of genes on these traits is mediated by motivational and self-control traits. Consider "educational attainment" – operationalized as the number of years spent in education. There is evidence that (a) certain genes

contribute to differences in years spent in education because (b) individuals with certain alleles of such genes tend to be more disciplined and committed to long-term goals than individuals with alternative alleles. If knowledge of the genetic causes of educational attainment induces fatalism, as predicted by genetic essentialism, the causal impact of such genes on educational attainment is likely to grow in a manner similar to that found in the obesity example, if this knowledge becomes prevalent. However, let me stress that the sketched mechanism depicts but *one* possible way how a feedback loop between "genetically caused trait" and *genetically caused trait* might operate. In different contexts, with regards to different traits, different types of feedback-mechanisms might be at work. For instance, in some circumstances categorizing a trait as genetically caused might reduce, rather than increase, the extent to which genes cause the trait (a brief example will be given in the next section).
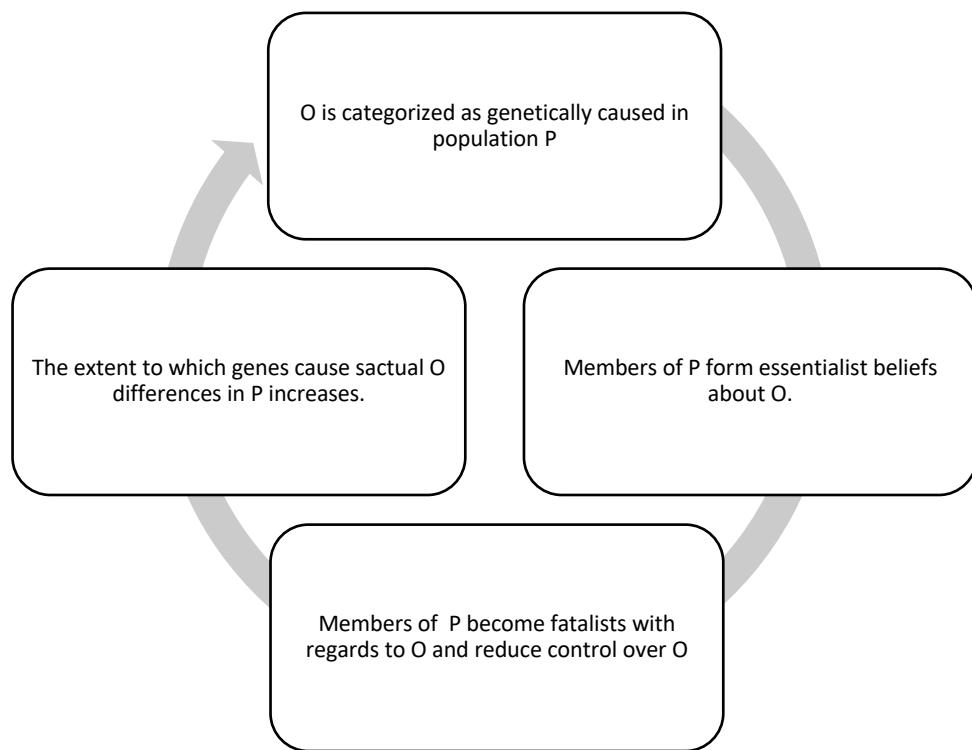
Figure 3

Figure 3. A mechanism via which the fact that obesity (O) is categorized as genetically caused increases the extent to which O is in fact genetically caused in a population P.

6. Responses to two objections

I will now consider, and respond to, two potential objections to what I have said.

First, one might reject my thesis that what has increased in the above toy example in the result of O being categorized as genetically caused is the degree of *genetic* causation, i.e., the extent to which *G* causes actual T differences. One might reject this thesis by rejecting one of the assumptions of the example. The example assumes two things. First, it assumes that there is a segment of O differences in Population 3 that is not present in Population 1 – the difference between the O values of individuals 3-8 (in Population 1, individuals 3-8 all have o^- and there are no differences in this segment of Population 1, whereas in Population 3 there *is* a difference: a difference between the O values of individuals 3 and 4 (both with o^+) and individuals 5-8 (each with o^-)). Call this segment "added O-differences". Second, it assumes that within added O-differences, O differences are entirely caused by G. And one might reject this second assumption. One might insist that O differences within added O difference are *not* caused by G but, instead, by calorie intake. Here's how one might argue in support of this claim. Given the set-up of our scenario, it is true that, keeping the values of all other variables (including G) fixed, if all of the individuals 3-8 in Population 3 (individuals whose O values constitute added O-differences) had consumed as few calories as individuals 5-8 actually did, then all of these individuals would have o^- just like individuals 5-8 actually do – there would be no O differences among individuals 3-8. And if all of the individuals 3-8 in Population 3 had consumed as many calories as individuals 3 and 4 actually did, then all of the individuals 3-8 would have o^+ just like individuals 3 and 4 actually do – again, there would be no O differences among individuals 3-8. This means that calorie intake satisfies the criterion (outlined in section 3) for causing all actual O differences within added O-differences. Therefore, all O differences within added O-differences are caused by calorie intake, not by G, and the G-caused portion of O variation in Population 3 has not changed compared to Population 1. If some empirical estimation of what proportion of the actual O differences G causes in Population 3 happens to count this portion as caused by G, then one has mistakenly inflated of the estimation.

As a first comment, even if the mechanism described with the above scenario is not a mechanism by which genetic facts about genetic causation get changed but rather a mechanism by which empirical estimates of genetic causation get inflated, as the objection has it, it is still a relevant mechanism that we need to be aware of if we are to avoid such inflation. However, there is good reason to maintain that added O-difference *is* caused by G and thus that it *is* facts about genetic causation *qua* actual difference making that are changed in the envisaged scenario. What we here witness is the phenomenon of gene-environment correlation (G-E). We have G–E if individuals with a certain genotype g* tend to experience certain environments more frequently than individuals with a different genotype g**, and these differences in experience lead g* individuals to instantiate a particular trait value more frequently than g** individuals. In our example, g^1 individuals within added O-differences ended up instantiating o^+ more frequently than g^2 individuals because, due to their large appetite, they systematically consumed more calories than g^2 individuals with small appetite. The influence of G on O is mediated by calorie intake (E), G and calorie intake correlate, and both E and G pass the criterion of causing all of the added O-differences. There are discussions within biology and the philosophy of biology about whether trait differences so produced should be ascribed to genetic or

environmental causes (Burt, 2022; Kaplan & Turkheimer, 2021; Lynch, 2017). Quite possibly there is no fact of the matter about this. However, there are at least two reasons to classify such cases as cases of *genetic* causation.

First, within empirical research, cases like these are routinely assigned to genetic causes: having a large appetite and the resulting eating behaviour are investigated as pathways via which G causes O rather than independent environmental causes of O (see references on page 14 of this paper). So, plausibly, in actual research contexts O, differences within added O-differences in Population 3 (and therefore *all* O differences in Population 3) would be ascribed to genetic causes. This is relevant because, recall, according to the definition assumed in this paper, a trait is genetically caused insofar as it has genetic causes *given the notion of having genetic causes* operative within empirical research. Secondly, assigning O differences within added O-differences to genetic causes is also supported by systematic philosophical considerations, where those have been laid out (e.g., Lynch, 2017). It is common to distinguish between different types of G-E: reactive (or evocative) and active. Reactive G-E occurs when the cause of the fact that individuals with g* experience some environment e* more frequently than individuals with g** is *exogenous* to organisms with g*. In such cases, experiencing e* is something that is "done to" g* individuals. Active G-E occurs when the cause of the fact that individuals with g* experience some environment e* more frequently than individuals with g** is *endogenous* to organisms with g*. In such cases, experiencing e* is something that g* individuals "do to" themselves by actively seeking out, and exposing themselves to, e*. According to Lynch (2017), if trait differences emerge because of reactive G-E then these differences should be counted as environmentally, rather than genetically, caused. If they emerge because of active G-E, these differences should be counted as genetically rather than environmentally caused. Our scenario with its assumptions about the biological function of G fits best to the active G-E case: g^1 individuals eat more than g^2 individuals because they actively seek out more food, and they do that because of their endogenous disposition to crave after more food (and, if we assume that G is pleiotropic for impulsivity, to not resist this craving). As such it would be classified as a case of active G-E. Consequently, O differences within added O-differences would qualify as caused entirely by G whereby caloric intake would not be classified as part of the "environment" but as an endophenotype for O.

The second objection targets my claim that the interactivity of *genetically caused trait* is empirically plausible. Specifically, one might argue that it is empirically implausible that the kind of feedback loop exemplified in my toy example ever gets instantiated. To show that this is so, one would have to show that one or other of the empirical premises that the example relied upon is false or weakly supported. For instance, my example assumed that essentialist attitudes about genes and genetically caused traits are relatively pervasive and their effect on people's behavioural response to genetic information is relatively big. And one could reasonably argue that existing evidence is inconclusive on this matter. Although there is solid evidence that essentialist biases indeed influence people`s responses to genetic information, they are far from being the only factor influencing this. Furthermore, there is evidence that the impact of essentialist beliefs can be outweighed, screened off by other factors. (see e.g. Condit 2010, 2019; Dar-Nimrod et al., 2021; Dar-Nimrod & Heine, 2011; Marteau et al. 2010; McBride et al. 2010).

Whether, and how, a feedback loop between application of the category "genetically caused trait" and the kind *genetically caused trait* occurs for obesity or any other trait is an empirical, and empirically testable, question. However, *prima facie* I think there is no reason to dismiss the empirical plausibility of either the concrete example or any other example of the same kind. As to the concrete example, although it is a toy example that significantly simplifies things, its core empirical foundations are strongly realistic depictions of our current best empirical knowledge. Thinking in particular of premise that people are genetic essentialists, I have two further points. First, acknowledging that my representation of the genetic essentialist framework was a simplification, existing evidence is certainly enough to warrant acceptance of at least the following: in *some* conditions, people are genetic essentialists and behave fatalistically with regards to traits they believe to be genetically caused. So, at least in *some* conditions the kind of mechanism that I described is not unlikely to be in effect. It is a further question then, what these conditions are. This paper serves precisely as a launch pad for beginning to address these questions more closely. But secondly, the emergence of a feedback loop between the category "genetically caused trait" and the kind *genetically caused trait* does not ultimately depend upon the presence of genetic essentialism. To see this, consider the following example. Let the trait of interest is a disease trait (D) with values "present" (d^+) and "absent" (d^-), and d^+ is generally thought to be very undesirable. And suppose that in some population, G is among the actual difference makers with regards to D with carriers of a certain allele of G, g^+ being to some significant degree more likely to develop d^+ than carriers of the alternative allele g^-. Suppose also that g^+ carriers can prevent developing d^+ if they follow a strict healthy lifestyle. Now, for the sake of the example, let's suppose that instead of being genetic essentialists, members of this population are genetic "neutralists": they take genetic causes to be no different from non-genetic causes of traits in terms of whether their influence on a trait can be counteracted or neutralized by intentional action. Assuming this, the discovery that D is genetically caused might trigger the following scenario. Members of the relevant population all know that G is causally related to D in the way described. They also correctly believe that g^+ carriers unlike g^- carriers are highly likely to develop d^+ *unless* they commit to some specific demanding lifestyle. Most individuals in this population do not know if they carry g^+ of g^-. However, motivated by fear of developing d^+, all individuals, including g^+ carriers, take the necessary measures to prevent developing d^+. In consequence, no one in this population develops d^+. As there are now no D differences in this population, nothing, including G, causes such differences. The degree to which G causes (actual differences in) D has decreased compared to when G was not yet known to cause D.

As this example illustrates, for a feedback loop between application of the category "genetically caused trait" and the kind *genetically caused trait* to emerge, it is not necessary that people hold *essentialist* beliefs towards genetically caused traits. A feedback loop like this requires only that in a given population people have *some* beliefs about genetically caused traits which, in conjunction with certain background beliefs and other background circumstances, impact – in some right manner – people's behaviour with regards to a trait they believe to be genetically caused. That this is sometimes the case is a much weaker assumption than the assumption of genetic essentialism.

7. Conclusion

I defended the idea that whether, and to what extent, a human trait is genetically caused given the empirically relevant concept of genetic causation can be influenced by whether or not the trait is categorized as genetically caused in the context of human genetic research. That this is so becomes clear once we unpack what it means for genes to cause a trait in such context. I unpacked this meaning using Kenneth Waters' account according to which genes cause a trait in the empirically relevant sense insofar as genes are among the actual difference making causes of the trait. I then fleshed out my thesis by sketching a hypothetical but empirically plausible model mechanism that might account for this effect. This mechanism drew upon the empirical hypothesis that laypeople have essentialist and therefore fatalistic attitudes towards traits they believe to have genetic causes. I also stressed that the concrete example depicts but one kind of possible mechanism whereby categorizing a trait as genetically caused changes the extent to which the trait is in fact genetically caused. In different contexts, depending upon different factors, different kinds of feedback mechanisms might be at work. Given the ever-growing prominence of genetic knowledge, it is important to further explore the possibility of such feedback loops – if and under which conditions they emerge. This paper serves to raise the alarm about this possibility and gestures towards a more detailed philosophical and empirical investigation into the matter. It also serves to highlight yet another reason why human behavioural genetics is akin (probably more than typically recognized) to the *human* and *social* sciences.

References

Abdella, H. M., El Farssi, H. O., Broom, D. R., Hadden, D. A., & Dalton, C. F. (2019). Eating Behaviours and Food Cravings; Influence of Age, Sex, BMI and FTO Genotype. *Nutrients*, *11*(2), 377. https://doi.org/10.3390/nu11020377

Allen, S. R. (2021). Kinds behaving badly: Intentional action and interactive kinds. *Synthese*, *198*(12), 2927–2956. https://doi.org/10.1007/s11229-018-1870-0

Berent, I. (2020). *The Blind Storyteller*. Oxford University Press. https://doi.org/10.1093/oso/9780190061920.001.0001

Bourrat, P. (2019). Heritability, Causal Influence and Locality. *Synthese*, *198*(7), 6689–6715. https://doi.org/10.1007/s11229-019-02484-3

Bourrat, P. (2020). Causation and Single Nucleotide Polymorphism Heritability. *Philosophy of Science*, *87*(5), 1073–1083. https://doi.org/10.1086/710517

Burt, C. H. (2022). Challenging the Utility of Polygenic Scores for Social Science: Environmental Confounding, Downward Causation, and Unknown Biology. *The Behavioral and Brain Sciences*, 1–36. https://doi.org/10.1017/S0140525X22001145

Chami, N., Preuss, M., Walker, R. W., Moscati, A., & Loos, R. J. F. (2020). The role of polygenic susceptibility to obesity among carriers of pathogenic mutations in MC4R in the UK Biobank population. *PLOS Medicine*, *17*(7), e1003196. https://doi.org/10.1371/journal.pmed.1003196

Cheung, B. Y., Dar-Nimrod, I., & Gonsalkorale, K. (2014). Am I My Genes? Perceived Genetic Etiology, Intrapersonal Processes, and Health. *Social and Personality Psychology Compass*, *8*(11), 626–637. https://doi.org/10.1111/spc3.12138

Condit, C. M. (2019). Laypeople Are Strategic Essentialists, Not Genetic Essentialists. *Hastings Center Report*, *49*(S1), S27–S37. https://doi.org/10.1002/hast.1014

Cooper, R. (2004). Why Hacking is Wrong About Human Kinds. *British Journal for the Philosophy of Science*, *55*(1), 73–85. https://doi.org/10.1093/bjps/55.1.73

Czene, K., Lichtenstein, P., & Hemminki, K. (2002). Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *International Journal of Cancer*, *99*(2), 260–266. https://doi.org/10.1002/ijc.10332

Dar-Nimrod, I., Cheung, B. Y., Ruby, M. B., & Heine, S. J. (2014). Can merely learning about obesity genes affect eating behavior? *Appetite*, *81*, 269–276. https://doi.org/10.1016/j.appet.2014.06.109

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, *137*(5), 800–818. https://doi.org/10.1037/a0021860

Dar-Nimrod, I., Kuntzman, R., MacNevin, G., Lynch, K., Woods, M., & Morandini, J. (2021). Genetic essentialism: The mediating role of essentialist biases on the relationship between genetic knowledge and the interpretations of genetic information. *European Journal of Medical Genetics*, *64*(1), 104119. https://doi.org/10.1016/j.ejmg.2020.104119

Gannet, L. (1999). What's in a Cause?: The Pragmatic Dimensions of Genetic Explanations. *Biology and Philosophy*, *14*(3), 349–373.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought* (pp. x, 382). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195154061.001.0001

Gelman, S. A. (2009). Essentialist Reasoning about the Biological World. In A. Berthoz & Y. Christen (Eds.), *Neurobiology of "Umwelt": How Living Beings Perceive the World* (pp. 7–16). Springer. https://doi.org/10.1007/978-3-540-85897-3_2

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38*(3), 213–244. https://doi.org/10.1016/0010-0277(91)90007-Q

Gould, W. A., & Heine, S. J. (2012). Implicit Essentialism: Genetic Concepts Are Implicitly Associated with Fate Concepts. *PLOS ONE*, *7*(6), e38176. https://doi.org/10.1371/journal.pone.0038176

Hacking, I. (1999). *The Social Construction of What?* Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674004122

Hacking, I. (2007). Kinds of People: Moving Targets: British Academy Lecture. In *Proceedings of the British Academy, Volume 151, 2006 Lectures*. British Academy. https://doi.org/10.5871/bacad/9780197264249.003.0010

Hacking, I. (2010). Pathological withdrawl of refugee children seeking asylum in Sweden. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *41*(4), 309–317. https://doi.org/10.1016/j.shpsc.2010.10.001

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, *39*(1), 113–127. https://doi.org/10.1348/014466600164363

Hauswald, R. (2016). The Ontology of Interactive Kinds. *Journal of Social Ontology*, *2*(2), 203–221. https://doi.org/10.1515/jso-2015-0049

Heine, S. (2016). *DNA Is Not Destiny: The Remarkable, Completely Misunderstood Relationship Between You and Your Genes*. W.W.Norton & Co.

Heine, S. J., Dar-Nimrod, I., Cheung, B. Y., & Proulx, T. (2017). Chapter Three - Essentially Biased: Why People Are Fatalistic About Genes. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology* (Vol. 55, pp. 137–192). Academic Press. https://doi.org/10.1016/bs.aesp.2016.10.003

Kaplan, J. M., & Turkheimer, E. (2021). Galton's Quincunx: Probabilistic causation in developmental behavior genetics. *Studies in History and Philosophy of Science*, *88*, 60–69. https://doi.org/10.1016/j.shpsa.2021.04.001

Keil, F. C. (1989). *Concepts, Kinds, and Cognitive Development*. A Bradford Book.

Khalidi, M. A. (2010). Interactive Kinds. *The British Journal for the Philosophy of Science*, *61*(2), 335–360. https://doi.org/10.1093/bjps/axp042

Kuorikoski, J., & Pöyhönen, S. (2012). Looping Kinds and Social Mechanisms. *Sociological Theory*, *30*(3), 187–205. https://doi.org/10.1177/0735275112457911

Larder, R., Sim, M. F. M., Gulati, P., Antrobus, R., Tung, Y. C. L., Rimmington, D., Ayuso, E., Polex-Wolf, J., Lam, B. Y. H., Dias, C., Logan, D. W., Virtue, S., Bosch, F., Yeo, G. S. H., Saudek, V., O'Rahilly, S., & Coll, A. P. (2017). Obesity-associated gene TMEM18 has a role in the central control of appetite and body weight regulation. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(35), 9421–9426. https://doi.org/10.1073/pnas.1707310114

Loos, R. J. F., & Yeo, G. S. H. (2022). The genetics of obesity: From discovery to biology. *Nature Reviews Genetics*, *23*(2), Article 2. https://doi.org/10.1038/s41576-021-00414-z

Lynch, K. (2017). Heritability and Causal Reasoning. *Biology and Philosophy*, *32*(1), 25–49. https://doi.org/10.1007/s10539-016-9535-1

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

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

Mathieson, I. (2021). The omnigenic model and polygenic prediction of complex traits. *The American Journal of Human Genetics*, *108*(9), 1558–1563. https://doi.org/10.1016/j.ajhg.2021.07.003

Matthews, L. J. (2022). Half a century later and we're back where we started: How the problem of locality turned in to the problem of portability. *Studies in History and Philosophy of Science*, *91*, 1–9. https://doi.org/10.1016/j.shpsa.2021.10.021

Medin, D., & Ortony, A. (1989). Psychological essentialism. *Vosniadou and Ortony*, *1989*, 179–195.

Meyre, D., Mohamed, S., Gray, J. C., Weafer, J., MacKillop, J., & de Wit, H. (2019). Association between impulsivity traits and body mass index at the observational and genetic epidemiology level. *Scientific Reports*, *9*(1), 17583. https://doi.org/10.1038/s41598-019-53922-8

Mezquita, L., Sánchez-Romera, J. F., Ibáñez, M. I., Morosoli, J. J., Colodro-Conde, L., Ortet, G., & Ordoñana, J. R. (2018). Effects of Social Attitude Change on Smoking Heritability. *Behavior Genetics*, *48*(1), 12–21. https://doi.org/10.1007/s10519-017-9871-1

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *ELife*, *9*, e48376. https://doi.org/10.7554/eLife.48376

Namjou, B., Stanaway, I. B., Lingren, T., Mentch, F. D., Benoit, B., Dikilitas, O., Niu, X., Shang, N., Shoemaker, A. H., Carey, D. J., Mirshahi, T., Singh, R., Nestor, J. G., Hakonarson, H., Denny, J. C., Crosslin, D. R., Jarvik, G. P., Kullo, I. J., Williams, M. S., & Harley, J. B. (2021). Evaluation of the MC4R gene across eMERGE network identifies many unreported obesity-associated variants. *International Journal of Obesity*, *45*(1), Article 1. https://doi.org/10.1038/s41366-020-00675-4

Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R. I., Selzam, S., Dale, P. S., Esko, T., Metspalu, A., & Plomin, R. (2018). Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nature Human Behaviour*, *2*(4), 269–275. https://doi.org/10.1038/s41562-018-0332-5

Silventoinen, K., Jelenkovic, A., Sund, R., Latvala, A., Honda, C., Inui, F., Tomizawa, R., Watanabe, M., Sakai, N., Rebato, E., Busjahn, A., Tyler, J., Hopper, J. L., Ordoñana, J. R., Sánchez-Romera, J. F., Colodro-Conde, L., Calais-Ferreira, L., Oliveira, V. C., Ferreira, P. H., … Kaprio, J. (2020). Genetic and environmental variation in educational attainment: An individual-based analysis of 28 twin cohorts. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-69526-6

Silventoinen, K., & Konttinen, H. (2020). Obesity and eating behavior from the perspective of twin and genetic research. *Neuroscience and Biobehavioral Reviews*, *109*, 150–165. https://doi.org/10.1016/j.neubiorev.2019.12.012

Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D. I., Cornes, B. K., Davis, C., Dunkel, L., Lange, M. de, Harris, J. R., Hjelmborg, J. V. B., Luciano, M., Martin, N. G., Mortensen, J., Nisticò, L., Pedersen, N. L., Skytthe, A., Spector, T. D., Stazi, M. A., Willemsen, G., & Kaprio, J. (2003). Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Research and Human Genetics*, *6*(5), 399–408. https://doi.org/10.1375/twin.6.5.399

Turkheimer, E. (2000). Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science*, *9*(5), 160–164. https://doi.org/10.1111/1467-8721.00084

Uchiyama, R., Spicer, R., & Muthukrishna, M. (2021). Cultural Evolution of Genetic Heritability. *Behavioral and Brain Sciences*, 1–147. https://doi.org/10.1017/S0140525X21000893

Wang, K., Li, W.-D., Zhang, C. K., Wang, Z., Glessner, J. T., Grant, S. F. A., Zhao, H., Hakonarson, H., & Price, R. A. (2011). A Genome-Wide Association Study on Obesity and Obesity-Related Traits. *PLOS ONE*, *6*(4), e18939. https://doi.org/10.1371/journal.pone.0018939

Waters, C. K. (2007). Causes That Make a Difference. *The Journal of Philosophy*, *104*(11), 551–579. https://doi.org/10.5840/jphil2007104111

Weber, M. (2017). *Causal Selection vs Causal Parity in Biology: Relevant Counterfactuals and Biologically Normal Interventions* [Preprint]. http://philsci-archive.pitt.edu/13382/

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318. https://doi.org/10.1007/s10539-010-9200-z