DISSERTATIONES
PHILOSOPHICAE
UNIVERSITATIS
TARTUENSIS
19

RIIN KÕIV

The Content and Implications of Nativist Claims. A Philosophical Analysis

DISSERTATIONES PHILOSOPHICAE UNIVERSITATIS TARTUENSIS
19

DISSERTATIONES PHILOSOPHICAE UNIVERSITATIS TARTUENSIS
19

RIIN KÕIV

The Content and Implications of Nativist Claims. A Philosophical Analysis

UNIVERSITY OF TARTU
Press
1632

Faculty of Arts and Humanities, Institute of Philosophy and Semiotics

Dissertation has been accepted for defence of the degree of Doctor of Philosophy (PhD) in Philosophy on November 18, 2021 by the Council of the Institute of Philosophy and Semiotics, University of Tartu.

Supervisors: Professor Daniel Cohnitz, PhD, Utrecht University, Netherlands
Professor Bruno Mölder, PhD, University of Tartu, Estonia

Opponent: Professor Ingo Brigandt, PhD, University of Alberta, Canada

Defence: The thesis will be defended at the University of Tartu, Estonia, on December 27, 2021, at 16:15, via video bridge

This research has been supported by the Centre of Excellence in Estonian Studies (European Union, European Regional Development Fund), "Philosophical analysis of interdisciplinary research practices" (PRG462); European Social Fund's Doctoral Studies and International Programme DoRa, which is carried out by Archimedes Foundation; the national scholarship programme Kristjan Jaak, which is funded and managed by Archimedes Foundation in collaboration with the Ministry of Education and Research. It is related to research project IUT20-5 (Estonian Ministry of Education and Research) and ETF9083 "Keeleline reaalsus ja selle psühholoogiline alus".

ISSN 1406-9520
ISBN 978-9949-03-774-2 (print)
ISBN 978-9949-03-775-9 (pdf)

University of Tartu Press
www.tyk.ee

CONTENTS

ACKNOWLEDGMENTS

This dissertation would not exist without the support and help from my friends and colleagues in and around the University of Tartu's Department of Philosophy. I am especially indebted to Alex Davis, Taavi Laanpere and Uku Tooming who have been my main team and scaffold throughout the process of writing the thesis. They have helped me develop and test my ideas, have taken time to read my writing, have given valuable feedback on the versions of the texts that now make up the thesis. I thank my supervisors Daniel Cohnitz and Bruno Mölder for their assistance, feedback and support. I thank Endla Lõhkivi for her collegial, institutional and intellectual support that was very much needed when completing the thesis. I thank the late Karola Stotz for much appreciated conversations and feedback that helped me improve two of the papers of the thesis, and Juhani Yli-Vakkuri for a useful discussion on one of the papers. I have received feedback at various conferences. This feedback has greatly encouraged me to proceed with my work. I am grateful to Laura Viidebaum for her support, enriching and encouraging conversations, and for keeping up the spirit; to my sister Epp Kõiv who has edited and proofread bits and pieces of my writing over the years; to Juhan Hellerma, Henrik Sova, Peeter Tinits, Aaron Wendland, Lisa Zorzato for stimulating discussions; to Markus Kasemaa for being there; to Alex Davis, Jaana Eigi, Taavi Laanpere, Toomas Lott and James Stephen Pearson for proofreading parts of the thesis; to all those people who came to my aid a week before submitting the thesis; and to many others. Finally, I am much indebted to my family, parents and grandparents for their comprehensive support, patience, and example.

LIST OF ORIGINAL PUBLICATIONS

I. Kõiv, Riin 2021. Innate Mind Need Not Be Within. *Acta Analytica*. 36, 101–121. https://doi.org/10.1007/s12136-020-00441-1

II. Kõiv, Riin 2020. Elusive vehicles of Genetic Representation. *Biology and Philosophy*, 35 (24), https://doi.org/10.1007/s10539-020-9741-8

III. Kõiv, Riin 2019. Causal Social Construction. *Journal of Social Ontology*, 5 (1), 77−99. https://doi.org/10.1515/jso-2019-0006

IV. Kõiv, Riin. Socially Constructed and/or Genetically Caused? Manuscript

INTRODUCTION

1. Overview of the thesis

I will call "nativist claims" all those claims that declare a trait to have significant organism internal (e.g., genetic) causes present in the organism at its birth.[1] I will call "scientific nativist claims" all those nativist claims made in scientific or other theoretical contexts. For example, scientific nativist claims include all those claims that declare a trait to be innate, heritable, genetic, hardwired, encoded in genes, naturally selected, inherited, have genetic causes etc.. Nativist claims are a frequent occurrence in different contexts of scientific research: various fields of genetic and psychological research, cognitive science, behavioral ecology, evolutionary biology etc. But the relevance of scientific nativist claims and the hypotheses that they express transcends the borders of the particular sciences in which these claims are made. As evidence, nativist claims feed into and constrain theories in yet further fields of research. Scientific nativist claims are consulted in social decision making. They are taken to (de)legitimize certain forms of social organization. They shape our understanding of the nature of human and other living beings and their place and possibilities in their environment. For this reason, it is important that the content of scientific nativist claims be interpreted correctly – as expressing the content that is in fact well-supported by evidence in a given scientific context. The aim of this dissertation is to ascertain the content of certain types of nativist claim in explanatory contexts, and clarify their implications, and non-implications, for concrete questions raised and theories advanced in certain areas of philosophical research.

This is a worthwhile aim. The content and implications of different types and instances of nativist claim made in different scientific contexts can differ significantly. The implications of a trait's being innate, given what is meant by 'innate' in immunology, can differ significantly from the implications of a trait's being innate given what is meant by 'innate' in developmental psychology. The latter, in turn, can differ from the implications of a trait's having genetic causes given what is understood by 'has genetic causes' in the context of population genetics. These are different yet again from the implications of a trait's having genetic causes given what is identified as having genetic causes by molecular genetics, and the implications of both of these can differ from the implications of a trait's being naturally selected. And so on. Yet, the different types of nativist claim – and this is partly what makes the otherwise loose category "nativist claim" an important category – tend to be interpreted as though they were expressing the same or very similar proposition. Moreover, as empirical research shows, they tend to be interpreted in light of mistaken essentialist folk biological conceptions (in a way that I will clarify in section 2). In sum, the content and implications of different scientific nativist claims are prone to be obscured and

[1] Throughout this introduction I will use 'trait' broadly to include all sorts of properties of biological organisms.

veiled by folk conceptions, which can give rise to various false inferences. My dissertation serves to mitigate this danger.

The dissertation consists of three published articles, one article manuscript and an extended introduction to the articles. Each of the articles focuses on one type of nativist claim. "Elusive Vehicles of Genetic Representation" (paper I) is about the concept of genetic information and, more specifically, about the claim frequently made in biological science that organisms develop certain phenotypic traits in virtue of their genes bearing information about these traits. The paper argues that a prominent philosophical theory according to which this is indeed the case – the teleosemantic theory of genetic information – is not supported by empirical facts. "Innate Mind Need Not Be Within" (paper II) scrutinizes claims to the effect that some mental concepts are innate when such claims are made within the cognitive and other psychological sciences. The paper clarifies the implications of such claims for causal externalist theories of mental content. It explains why there is no shortcut inference from the thesis that a concept is innate to the conclusion that causal externalism (given its different versions) is false about the content of the concept. "Socially Constructed and/or Genetically Caused?" (paper III) in combination with "Causal Social Construction" (paper IV) propose a conceptual framework for assessing the implications of findings to the effect that a human trait has genetic causes for social constructionism about the trait. I draw upon a general theory of causation – the contrastive counterfactual dependence account of causation – to outline the conditions under which an empirically supported thesis to the effect that a trait has genetic causes is in conflict with a corresponding social constructionist claim, and the conditions under which it isn't. "Causal Social Construction" proposes an analysis of the concept of causal social construction in terms of the contrastive counterfactual dependence account of causation. This analysis can be viewed as an operationalization of the concept of causal social construction for the purposes of the analysis undertaken in "Socially Constructed and/or Genetically Caused?". However, the analysis also makes a self-standing contribution to philosophical social constructionist debates.

Two general assumptions underlie the project undertaken in the papers of the dissertation, assumptions that they also illustrate and corroborate. First, in the context of scientific nativist claims that are in good epistemic standing, 'innate', 'genetic', 'genetically encoded' etc. mean something different from what laypeople tend to take them to mean. Second, these words should not be assumed to mean the same thing and instances of the same word (say, 'innate', 'genetically caused') shouldn't be assumed to mean the same thing when used in *different contexts*. They often do not. Consequently, nativist claims in varying research contexts can report significantly different truths – if truths they report. These two assumptions fix some general methodological guidelines for evaluating the content and implications of a scientific nativist claim. First, one should abstain from whatever prior lay intuitions one may have about the concepts deployed in association with 'innate' and the like. Second, it is reasonable to engage in relatively local case studies of the meaning of 'innate', 'genetically caused' etc. in specific explanatory contexts at issue. Even if some of the nativist claims made

in different research contexts express the same proposition, this should be a discovery based on such local analyses, and not an assumption or desideratum to begin with. The papers of the dissertation each provide a case study like this.

There is a venerable philosophical tradition that discusses the concepts that the papers of the dissertation address (of innateness, of genetic causation, genetic information). The dissertation continues and builds upon this tradition. However, it differs from typical existent discussions in at least three respects. First, rather than focusing on but one of these concepts, I view these concepts comparatively in a shared framework. Second, a typical philosophical analysis of, say, a concept of innateness or a concept of genetic causation aims at being adequate in relation to as many scientific usages of the relevant word as possible. In contrast, my dissertation emphasizes the possible plurality of such adequate analyses. Such an emphasis and focus on possible plurality is especially relevant if the purpose of such an analysis is to extract the implications of a nativist claim for some further question (as explained in section 3.1). Third, the papers open up existing discussions in the philosophy of science on such nativist concepts to audiences beyond philosophers of science. Each paper is a project of bridging and translating between different research fields – empirical disciplines and different subfields of philosophy – and correcting possible and plausible mistranslations between them. This means that I often rely and build upon existing theories. For example, in "Innate Mind Need Not Be Within" and "Socially Constructed and/or Genetically Caused?" I do not provide an original analysis of the concept of innateness or of genetic causation. Instead, I use existing philosophical accounts – organizing and modifying them in accordance with my purposes – in novel ways to clarify certain philosophical questions that have typically been uninformed by these accounts. Likewise in "Elusive Vehicles of Genetic Representation" where I first dissect the commitments of the teleosemantic theory of representation (itself drawn from work in the philosophy of mind), then consult existing philosophical and empirical theories about the nature and role of genes in development and evolution, and then demonstrate that in light of these theories, the teleosemantic account of genes as representing phenotypes reveals itself to be a misapplication of the teleosemantic notion of representation.

In this extended introduction to the articles, I do three things. In section 2, I elucidate the motivation for the general project that the individual papers contribute to. This motivation stems from two empirical hypotheses: first, that we are psychological essentialists by nature and, second, that we tend to (mis)interpret scientific nativist claims within an essentialist mindset. In section 3, I discuss the background of each individual paper and locate each paper in relation to its topic-relevant literature. As was said above, all the papers are to a large extent projects of theory-bridging. As such, the papers – individually and jointly – bring together topics and research traditions that have run down more or less distinct paths in academic philosophy. When giving background to each paper, I focus on that part of its background that concerns the particular type of nativist claim at issue. In section 4, I address a possible "eliminativist" objection to one motivational premise of my dissertation. My response to this objection contains seeds for intended future research.

2. General motivation. Psychological essentialism and nativist claims

2.1. Psychological essentialism

Ample evidence shows that we, humans, are predisposed to think of biological organisms, including organisms of our own species, as possessing an underlying invisible causally potent inner nature or inner essence (as I shall be calling it) (Berent, 2020; Gelman, 2003, 2009; Gelman & Wellman, 1991; Keil, 1989; Medin & Ortony, 1989). We view this inner essence as something that an organism inherits from its parents, that the organism shares with other organisms of the same kind (especially of the same generic species), that is developmentally fixed, that defines the organism as the kind of organism that it is, that survives changes in the organism's superficial properties. As a manifestation of this predisposition, we are prone to view some traits of organisms as being caused by this inner essence. We view such traits also as developmentally fixed, inherited from biological parents, and typical to organisms of the same biological species. Our predisposition to essentialize appears to have remarkably wide scope. In addition to non-human biological phenomena, it has been argued to explain a large variety of aspects of how we reason about human categories (Diesendruck & Menahem, 2015; Haslam & Whelan, 2008; Mandalaywala et al., 2018; Prentice & Miller, 2007; Rhodes & Moty, 2020, 2020).[2]

This predisposition to view organisms as possessing such inner essences is called "psychological essentialism" by psychologists. The usual way to construe and speak of psychological essentialism is to construe and speak of it as a set of implicit folk biological beliefs that forms a part of our folk biological theory. Central to this set of beliefs is the belief that animate things possess inner essences where a constituent of this belief is the concept of inner essence (henceforth INNER ESSENCE). The majority view is that, first, INNER ESSENCE is a placeholder concept: people share the belief *that* organisms have inner essences, however, need not have any beliefs about *what* this inner essence is (Gelman, 2003; Medin & Ortony, 1989). At different times in different contexts, different things can be assumed to play the role. Second, it is commonly agreed that INNER ESSENCE is an attractor concept, i.e., a concept that we disposed to apply in our thinking across a variety of domains of the animate world, or, in other words, that attracts our thinking about the animate world (e.g., Machery et al.. 2019; Machery 2021). I will adopt this view and way of speaking.

I will take psychological essentialism (as just described) for granted. That is, I will assume that humans have the implicit belief that animate beings possess inner essences and, as a component of this belief, the concept INNER ESSENCE; that this belief together with INNER ESSENCE is a human universal; and that

[2] Some authors argue that essentialist biases are also at work in reasoning about inanimate objects (Newman et al., 2008).

we frequently employ this belief when reasoning about the animate, including the social, world.[3] The papers of this thesis do *not* engage with issues directly concerning psychological essentialism. However, the case studies that I undertake in the individual papers each contribute to a general project that is partly motivated by the assumption that something like psychological essentialism is true. In this section I explain this motivation.

The view that both INNER ESSENCE, and the belief that animate things possess inner essences, are human universals says nothing in itself about the causes of possessing the concept and the belief. The received view is that psychological essentialism is, in some qualified sense, innate.[4] This view is supported by the following considerations. First, the essentialist stance has been observed to be operative in an array of natural and cultural environments (Astuti et al., 2004; Haslam et al., 2000; Sousa et al., 2002; Waxman et al., 2007). Second, essentialist reasoning biases appear early in human development (at four years of age at the latest (Gelman, 2003, 2009; Rakoczy & Cacchione, 2019)). These observations suggest that the essentialist stance does not require any specific environmental, cultural or linguistic experiential input to develop. Third, there is some evidence that psychological essentialism has homologies in great apes (Cacchione et al., 2016). This suggests that it is a trait that humans possess because of genes that got fixed in the human population prior to the time when human evolution parted ways with the evolution of other great apes. Fourth, there is negative developmental evidence that, on account of poverty of stimulus, the essentialist stance cannot be (entirely) explained by socialization (e.g., Gelman et al., 2004). However, many of these observations are also compatible with other hypotheses about the origin of essentialism, e.g., that psychological essentialism develops due to some universal features of human developmental environment or human language (Khalidi & Mugg, 2014); or that it is a by-product of some other universal innate dispositions or concepts (Cimpian & Salomon, 2014; see also Gelman, 2003, Ch. 11).[5] Regardless of which of these etiologies is correct, they all imply that the development of the essentialist stance and INNER ESSENCE is environmentally invariant. They are also all consistent with the view that INNER ESSENCE is a cognitive attractor. These are the two features of psychological essentialism that here matter.

Let it be stressed that the proposition that INNER ESSENCE is a humanly universal attractor concept, possibly hardwired into our cognitive architecture, does *not* imply that our reasoning about the animate world is necessarily mediated by this concept nor that the effect of this mediation couldn't be smoothed, canalized or screened out. Essentialist thinking is a *bias* just like any other and as

[3] I do not mean to commit myself to any specific view on the nature of the belief *qua* belief and the nature of the concept *qua* concept. What follows is compatible with different views on this matter.

[4] As will become clear in the next section, 'innate' can mean different things. I here use it in the absence of a better word, assuming whichever meaning it is assumed to have when psychological essentialism is hypothesized to be innate.

[5] This might not rule out that it is still "innate" given some relevant definition of the word.

such does not manifest itself in a deterministic manner. Depending on cultural, social, experiential and informational circumstances, the bias can be more or less prone to manifest itself, can manifest itself in different forms, can be overridden, screened out etc. (Mandalaywala, 2020; O'Connor & Joffe, 2014; Rhodes et al., 2012; Uhlmann et al., 2014). For instance, Norenzayan & Heine (2005) demonstrate that if primed with information that essentialism is mistaken about a category, subjects are less likely to use essentialist reasoning strategies. In sum, it is possible to bring about conditions that either encourage or inhibit our essentialist biases.

And sometimes it is desirable to bring such conditions about, especially conditions that inhibit our essentialist biases. This is because, often, essentialist biases make us form epistemically (and practically) problematic beliefs. Indeed, with regards to some categories and in some contexts, the essentialist "model" of the biological world might be a good enough heuristic in that it grounds accurate enough inferences and judgements, inferences and judgements that align with what the best theory about the topic would likewise predict. Psychological essentialism has been thought to explain why people assume category membership to travel with something inner and non-obvious and survive changes in observable properties. This assumption is correct in many cases: given our best raccoon-theory, a raccoon will surely continue to be a raccoon even if made to stink and look like a skunk. The folk-essentialist assumption that a seed taken from an apple hides in it something non-observable due to which the seed develops into an apple tree rather than a dandelion, and does so even if grown in a flowerbed together with dandelions, is also correct. So is the prediction that a child born to brown-skinned parents but raised in a light-skinned family will typically develop dark rather than light skin and does so because of something inside the child that it inherits from its biological parents. However, often essentialist inferences break down. First, they break down with some categories and traits, especially with social categories and human traits. People often assume that the shared features of a social group (e.g., races, genders, ethnic groups) are pre-eminently caused by something that resides deep inside the members of the group and is biologically inherited (Haslam et al., 2006; Haslam & Whelan, 2008). This assumption, commonly attributed to psychological essentialism, is straightforwardly false of many social groups and features characteristic to these groups. Also false is the common assumption that if a trait is shared in a group then its development is insensitive to variation in developmental environment and difficult to change by environmental intervention. Secondly, essentialist biases are likely to result in misrepresentation where the belief that members of a kind share an essence couples up with a specific belief about what the essence is (e.g., with the belief that the essence is the organism's DNA, heart, blood or whatever). Thirdly, mistakes are likely to occur where essentialist beliefs mediate

interpretations of scientific nativist hypotheses. This is the type of mistake that this dissertation is concerned with. I will say more on this in section 2.4.[6]

[6] It is sometimes said that the folk belief that biological categories possess essences is *tout court* and simply false; in reality, so it is said, many or most of the categories that we essentialize, like biological species, do not have essences (e.g., Gelman, 2004). I am hesitant about this strong claim of falsity (partly because I am not sure under which conditions the folk psychological belief, vague as it is, *would* be true). For one reason to believe that the lay belief that kinds possess essences is false, Gelman points out that essentialism as a philosophical metaphysical theory about the nature of kinds is generally agreed to be false; at least when the kind at issue is a biological taxon like species. This reasoning makes the assumption that *if* metaphysical essentialism is false, then the folk-psychological belief that biological kinds have essences is also false; which in turn assumes that the "essence" posited by lay mind and the "essence" of philosophical metaphysical essentialism about biological kinds are the same thing. But as far as I can see, this is not the case. Indeed, different metaphysical essentialist accounts of natural kinds in general and biological kinds in particular have understood what defines an essence of a kind in different ways. It is possible that according to some of these views the "essence" of psychological essentialism and the "essence" of metaphysical essentialism come pretty close. But with other prominent versions of metaphysical essentialism this does not seem to be the case. E.g., according to one once popular version of metaphysical essentialism, the essence of a kind is what is both necessary and sufficient for belonging to a category. It is this version of essentialism that Gelman appears to have in mind when she claims (correctly) that metaphysical essentialism about biological kinds is generally deemed false – arguably there is no such thing that all members of the same species necessarily share. But even if she is right, the falsity of this kind of essentialism does not seem to undermine the folk essentialist belief that species (or other biological kinds) have essences. Folk essentialism is not committed to the view that a kind possesses an essence only if there is something that all members of the kind necessarily share.

Here's why. Recall that INNER ESSENCE is likely a cluster concept, a concept that specifies a cluster of features such that possessing none of these features is by itself necessary for something, X, to count as the essence of a kind, however, possessing some relevant subset of these features is. Assuming this, the following would be a way to represent the content of INNER ESSENCE (where 1–8 are the candidate features laypeople associate with inner essence according to Gelman 2003):

> X is the inner essence of a kind K iff most (but not necessarily all) of (1)–(8) are true of X: (1) organisms that belong to K share X, (2) X persists through changes in an organism's superficial properties, (3) X accounts for the identity of an organism as a member of K, (4) X is not obvious, (5) X is inside the organism, (6) X is inherited from biological parents, (7) X is present in the organism at its birth, (8) X causes an organism to possess other K-typical traits.

Given this articulation of INNER ESSENCE, the lay belief that K has an essence is true if there is an X of which most of (1)-(8) are true and false if there is no such X. There are two reasons why the lay belief so understood can be true even if metaphysical species essentialism is not. First, notice that condition (1), "organisms that belong to K share X", is a generic and as such ambiguous. For current purposes, it can be read as either (a) or (b): (a) all organisms that belong to K share X, (b) some (e.g., most or paradigmatic) organisms that belong to K share X. Let's suppose that (b) is the correct reading. Assuming this reading, for the lay belief that K has an essence to be true, it suffices if *some* K members share X, where X meets many of the rest of the conditions (2) – (8). This surely can be the case even if metaphysical essentialism is false and there is nothing such that *all* members of K necessarily share. Now

2.2. INNER ESSENCE and language

So, we are predisposed to think thoughts about the biological sphere that contain INNER ESSENCE. Call such thoughts essentialist thoughts. Taking this much for granted, it is to be expected that we sometimes manifest this disposition by entertaining essentialist thoughts. Given this, it is also to be expected that we sometimes express these essentialist thoughts in language and sometimes interpret certain linguistic utterances as expressing essentialist thoughts. This in turn makes it plausible that there will be specific words and expressions associated with INNER ESSENCE where by 'associated' I mean something as loose as *Association*.

Association. An expression E is associated with INNER ESSENCE within a group of language users if

i) E is to some significant extent more likely than some relevant set of other expressions to be used to express essentialist thoughts or

ii) that a claim contains E increases to some significant extent the likelihood that the claim is interpreted as expressing essentialist thoughts

iii) that a claim contains E increases to some significant extent the likelihood that the claim triggers essentialist thoughts (e.g., because what one infers from a claim like this tends to contain INNER ESSENCE).

Do such expressions exist? And which expressions are these? Answers to these questions might be interesting if only for the following reason. *If* we want to prevent people (us) from interpreting certain linguistic information in an essentialist fashion, or from forming essentialist beliefs in response to such linguistic information, then a way to do this would be to avoid using these expressions or to modify the interpretation context in such a way that essentialist interpretations would be blocked or mitigated. And indeed, there is empirical research that its authors take to have identified such expressions. For instance, consider the following studies.

Griffiths et al. (2009) hypothesized that the vernacular concept of innateness, i.e., the concept associated with 'innate' in vernacular usage, is a concept that has its origin in the folk biological belief that organisms possess inherited inner

---

let's suppose that (a) is the correct reading of (1). In this case, the falsity of metaphysical essentialism would imply the falsity of folk essentialism if, according to folk essentialism, X would count as the essence of K only if X meets *all* of (1)-(8). But this is not the case. As per the above definition, X falls under INNER ESSENCE even if X does not meet some of the conditions (1)–(8), for instance (1).

So, the falsity of metaphysical essentialism does not imply the falsity of the folk belief that kinds have essences. Moreover, it is plausible that many biological species and some other paradigmatic essentialized kinds (but of course not all) do have an essence, given the above rather permissive articulation of the folk concept INNER ESSENCE.

essences. This hypothesis, according to the authors, predicts that if lay people believe a trait to be innate, then they believe that the trait will possess all or some combination of the three features associated with origin from inner essence: (1) it is species-typical, (2) its development is resistant to environmental influences, (3) it is functional (has a purpose) (Griffiths et al., 2009, p. 607). To test this hypothesis, Griffiths and colleagues set out to test if lay people are more likely to judge a trait innate if they believe the trait to have the three features. They had college students with no background in biological or behavioral science to read eight birdsong descriptions that corresponded to each combination of the presence/absence of the three features. They then measured how likely a birdsong corresponding to each combination is to be assessed innate. They found that, indeed, the presence of all three features contributes to people's judgements about whether birdsong is innate, and that each feature contributes to such judgements independently.

The study of Linquist et al. (2011) is a follow-up to the one of Griffiths et al. (2009). They set off to see whether the three features also predict subjects' endorsement of the thesis that a trait of an organism is 'in the DNA'. They found this to be the case (more so than with 'innate'). The authors treat 'in the DNA' as representative of a broader family of gene/genetics-related expressions and thus their findings to generalize to a relevant subset of these expressions.

This coheres with what other authors have found about people's interpretation of information about genes and genetically caused traits (Dar-Nimrod & Heine, 2011; Gould & Heine, 2012; Heine, 2016). Recall that INNER ESSENCE is thought to be a placeholder concept: in different circumstances, different things can be perceived to play the causal role that INNER ESSENCE specifies. Dar-Nimrod and Heine (2011) propose that in contemporary societies where information about, and talk of, genes and genetically caused traits is widespread, people believe genes, the DNA, to occupy this role. Correspondingly, people tend to view traits that they believe to be genetically caused *as if* originating from the inner essence of the organism and thus as having the corresponding features of fixity, category-typicality etc.[7] Dar-Nimrod and Heine call this account 'genetic essentialism framework'. They base this account on an array of studies into people's attitudes towards genes and genetically caused traits and behavioral responses to exposure to scientific claims about genetic causation. For instance, one experiment that the authors take to support genetic essentialism is the following:

> 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

[7] The set of properties that Dar-Nimrod and Heine list in association with the folk biological theory of inner essences differs to some extent from the one that Griffiths et al. and Linquist et al. (2011) operate with.

> 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))

According to the authors, these results show that people adopt fatalist attitudes towards perceived genetic etiology. The authors take these fatalist attitudes to be explained by the fact that people believe genes to constitute the inner essence of an organism and see "obesity" (or high BMI), insofar as having genetic causes, as an expression of this essence.[8]

This cookie-experiment does not directly test hypotheses about whether INNER ESSENCE is associated with specific expressions but a hypothesis about people's beliefs about genes and their role in causing phenotypes. However, it also provides information about the former. The cookie-experiment measures people's gene-related beliefs by measuring how people respond to information that a trait has genetic etiology. This information is presented to them in language containing gene-related expressions. Thus, insofar as the cookie-experiment shows that people view traits that they believe to be genetically caused *qua* traits originating from the organism's essence, it also shows that certain gene-related expressions like 'has genetic causes' trigger essentialist beliefs about the trait and thus that these expressions are associated with INNER ESSENCE as per Association.

The reader might sense some tension in the thesis that laypeople take 'genetically caused', 'in the DNA' and 'innate' to indicate that a trait is caused by an organism's inner essence. According to the description given earlier, essences are taken to be shared by organisms of the same biological species. At least this much is explicitly assumed in the studies of Griffiths et al (2009) and Linquist et al. (2011). Yet, we can observe that oftentimes it is idiosyncratic traits and individual differences *within* a species – as for example with the cookie-experiment – that both scientific findings and lay persons judge to be genetically caused or innate. Does this observation not conflict with the hypothesis that 'innate' and 'genetically caused' are associated with INNER ESSENCE, that is, that INNER ESSENCE is the concept that guides our application and interpretation of these expressions? Not really and for various reasons.

First, according to Association, an expression, say, 'in the DNA', is associated with INNER ESSENCE if statements that contain this expression *tend to* be interpreted in terms of, or trigger, beliefs that contain INNER ESSENCE. However, this much can be true without it being the case that *all* claims to the effect that a trait is in the DNA are interpreted in terms of such beliefs. The

[8] Similar effect has been observed with other traits (e.g., McBride et al. 2010, Dar-Nimrod 2017).

interpretation of some such claims, for example when an idiosyncratic trait is being claimed to be "in the DNA", might as well *not* involve INNER ESSENCE.

Second, in the statement that inner essences are taken to be shared by members of the same species, 'species' should not be understood as a taxonomic category of biological science. Rather, it should be understood to mean *biologically salient category*. Indeed, the generic species level is favored when ascribing shared category-typical essences. However, often, other kinds of categories are also essentialized. In the human sphere, these categories have been shown to include ethnic groups, sexes, different socio-economic classes etc. (Haslam et al., 2000). In general, there is flexibility to which categories, and when, we essentialize.

Third, recall that the studies of Griffiths et al. (2009) and Linquist et al. (2011) both showed that the features associated with the folk notion of inner kind essence – species typicality, fixity and teleology – contribute to innateness/ in-the-DNA judgements additively. That is, each of these features independently increases the likelihood that a trait is judged innate/in the DNA, rather than having this effect only in conjunction with the other(s). As the authors' summarize the results of these studies in a later paper:

> we do not view fixity, typicality, and teleology as separately necessary and jointly sufficient conditions for a trait to count as innate. Rather, we view them as prototypical features of the concept of innateness; they characterize what a prototypical innate trait would look like, and a trait is more likely to be judged innate to the extent that it resembles this prototypical innate trait. (Machery et al., 2019, p. 177)

This description of the way in which the concept associated with 'innate' – which according to Machery et al. is nothing but the concept of being caused by inner essence – is perfectly consistent with the observation that people sometimes judge a trait to originate from the inner essence of an organism even if the trait lacks some of the three features, e.g., category typicality. Moreover, the latter is to be expected given plausible general accounts of the nature/structure of concepts. The consensus appears to be that most concepts, and definitely primitive ones like INNER ESSENCE, do *not* have the structure of a definition that specifies necessary and sufficient conditions for falling under the concept. More plausibly, concepts specify a cluster of properties that are typical to the category that the concept is about, so that an entity falls under the concept if it has some relevant set (and not necessarily all) of the properties from this cluster. Among the views according to which this is the case is the dominant view in psychology that concepts are prototypes (Carey, 2009; Hampton, 2012). If INNER ESSENCE and CAUSED BY INNER ESSENCE likewise function as prototypes, then species typicality is not to be seen as a necessary condition for falling under these concepts. Considering all this, the hypotheses that 'in the DNA', 'innate', 'genetically caused' are associated with INNER ESSENCE is consistent with the fact that we often predicate 'in the DNA', 'innate' of traits that are idiosyncratic or vary in a species in some other obvious way.

2.3. I-expressions, science and two threats

We saw evidence that certain expressions, among them 'innate', 'in the DNA', 'genetic', are associated with INNER ESSENCE or at least a concept very much similar to INNER ESSENCE. And there is reason to believe that there are many other expressions that are, have been, or will be associated with INNER ESSENCE: expressions that tend to elicit chains of inference similar to those elicited by 'innate', 'in the DNA', 'genetically caused'. These reasons include the following. First, as a developmentally robust human universal, INNER ESSENCE is also possessed by non-English speakers who probably associate certain *non*-English expressions with the concept (e.g., 'kaasasündinud' in Estonian, 'angeboren' in German). Second, even in contemporary English there are expressions that either are used synonymously, or have substantial semantic overlap, with 'innate' and therefore are likely associated with INNER ESSENCE (Mameli and Bateson 2006). After all, association is a matter of degree. Some expressions may be associated with INNER ESSENCE more strongly than others. Some expressions might trigger thoughts containing INNER ESSENCE more robustly, others only in specific circumstances. Yet they all would count as associated with INNER ESSENCE as per Association. Third, which expressions are associated with INNER ESSENCE and how strongly is likely to vary within a language in time. For example, the association of 'in the DNA' and other gene-related expressions with INNER ESSENCE must obviously be a recent, no earlier than a 20th century phenomenon.

Those expressions that orbit around INNER ESSENCE in vernacular usage are the focus of my thesis. I will clarify how in a moment. But first, for convenience of expression, let me introduce a name for such expressions: 'i-expressions' [9] ('i' for 'inner').

> **i-expression.** An expression that in some relevant lay population is associated with INNER ESSENCE (as per Association)

I will also be using the word 'i-claim':

> **i-claim.** A claim that contains an i-expression (notice that not all nativist claims as defined in section 1 need to be i-claims and the other way around).

Which expressions count as i-expressions can change and can vary from population to population. For example, in the English language, in addition to 'innate' and 'in the DNA' (and its cognates) the set of i-expressions is likely to include

[9] Although the term 'i-expression' takes inspiration from Mameli (2008) and Shea (2012) who use the term 'i-properties' ('i' for 'innate'), the former should not be confused with the latter term. According to Mameli's definition, i-properties are properties that regularly occur in inferences of the following form: *trait T is innate, therefore, T has property P* and *trait T has property P, therefore T is innate* (Mameli, 2008, p. 735).

'inborn', 'part of core knowledge', 'inner essence', 'instinct', 'evolved', 'adaptation', 'present in newborns', 'hardwired', 'natural', 'inherited', 'heritable' etc.[10] Identifying which words are i-expressions is an empirical issue that need not concern us right now. For now, let's just take for granted that i-expressions exist.

Some i-expressions are also used within scientific nativist claims. For instance, in the cognitive sciences it is (or at least used to be) a common claim that some cognitive capacity, concept or bias (e.g., psychological essentialism and the concept of inner essence) is innate. In behavioral genetics most traits have been found to be heritable. In medical genetics some diseases are explained to be genetically caused. In evolutionary disciplines traits are claimed to be products of natural selection. Etc. Now, there is reason to believe that if these and other scientific nativist claims are to be interpreted correctly, then a given i-expression as it occurs in a claim like this should not be interpreted in terms of INNER ESSENCE, i.e. as indicating origination from the assumed essence of an organism. Typically, scientific nativist claims deploying one or another i-expression lack at least some, and often many, of the implications associated with "essential traits" in the folk-psychological sense. For instance, that a trait is highly heritable does *not* imply that the development of a phenotypic version of the trait is determined by or independent of the relevant organism's developmental environment; instead, in many cases, the development of a phenotypic versions of a highly heritable trait (crystallized intelligence, height, body-mass-index, alcoholism) is highly contingent on the organism's developmental environment. Nor does the fact that a trait is highly heritable imply that the trait is difficult to change by environmental intervention; it does not imply that organism-internal causes of the trait are more relevant than organism-external causes (at least on many readings of 'relevant'); it does not imply that possessing a version of the trait marks inductively significant biological category differences. All this applies – in varying degrees – to claims to the effect that a trait is an adaptation, has genetic causes, is innate etc. Thus, interpreting scientific i-claims in essentialist terms is likely to result in misinterpretation of these claims.

Moreover, some argue that *if*, in a research program, an i-expression expresses but the folk concept of being caused by inner essence, then this research program and those theories within it that are expressed using the expression, are not in good epistemic standing. First, as a vague folk concept, INNER ESSENCE does not meet the standards of rigor and precision of a scientific concept. Theories in terms of this concept would thus be vague, uninformative and unverifiable. Secondly, some have vocally argued (Griffiths, 2002; Griffiths & Machery, 2008) that INNER ESSENCE and CAUSED BY INNER ESSENCE would fail to be explanatorily useful concepts even if made more rigid and clear-cut. Science is interested in trading in projectible categories, i.e., in categorizing things so that one could infer from the manifestation of one property associated with the category to the manifestation of many others. INNER ESSENCE and CAUSED

[10] This is a selection of words that Iris Berent (2020) uses interchangeably with or at least as a close semantic relative of 'innate'.

BY INNER ESSENCE would be projectible if the earlier listed features (fixity, typicality etc.) that the folk mind associates with these concepts were sufficiently well correlated so that inferences from a trait having one of the features to its having the others succeeded often enough. As said previously, with *some* traits across *some* contexts such inferences indeed succeed, meaning that the features associated with "essential traits" to some extent correlate. However, as many authors argue, they correlate far too weakly for INNER ESSENCE and CAUSED BY INNER ESSENCE to capture any explanatorily optimal categories. This has been stressed in particular in discussions of the scientific concept of innateness, where some argue that the concept of innateness in various sciences is but an undesirable relic of the folk concept of being caused by organisms essence (Bateson & Mameli, 2007; Griffiths, 2002; Machery et al., 2019; Mameli & Bateson, 2011; Matteo Mameli & Bateson, 2006).[11]

Given all this, the fact that some i-expressions have parallel usage in vernacular and scientific contexts creates two kinds of threat. Call these threats the threat of bad science and the threat of badly interpreted science.

The threat of bad science. That an i-expression is used in both lay and scientific contexts creates the threat that, together with the expression, the theoretically inadequate INNER ESSENCE sneaks into scientific theories and explanations. There, it generates lazy pseudo-explanations tainted with pretheoretical intuitions, distracts, inhibits explanatory progress etc. For example, using 'innate' and 'genetic program' in explanatory contexts has been heavily criticized on just this basis (Griffiths, 2017; Griffiths, 2002; Machery et al., 2019b)

The threat of badly interpreted science. Suppose that in a given theoretical context the relevant i-expression does *not* express INNER ESSENCE. Suppose, instead, that it expresses an explanatorily useful well-defined technical concept. Even then, since by definition i-expressions tend to be interpreted in terms of psychological essentialism by lay people, the risk remains that theories worded in terms of the expression would be systematically (mis)interpreted by broader audiences because of the latter's essentialist beliefs. This worry is often raised with regards to the communication of empirical findings about the genetic causes of human traits. For example, Heine (2016) gives plenty of examples of how common it is that upon hearing that a trait "has genetic causes" or "is heritable" one infers that behavioral choices make little difference to whether the trait develops (an inference that Heine thinks is mediated by the folk assumption that genes constitute the essence of an organism). Not only do the technical concepts of genetic causation and heritability fail to license such an inference; this inference is often false, as only a small fraction of all human traits whose genetic

[11] Which concept a given i-expressions expresses in a given scientific context, whether this concepts is sufficiently different from its explanatorily idle folk-counterpart, and whether it serves any explanatory purpose, are all topics of ongoing debate within philosophy of science (e.g., regarding 'innate', 'genetic information', 'genetically caused', 'heritable'). I will say more about the content of these discussions later on when giving background to the papers of this thesis

correlates/causes have been studied, detected and reported are in this way genetically determined. The "broader audience" whose interpretation of scientific findings might get led astray by equivocation between folk and technical concepts associated with a given i-expression does not exclude scientists themselves. A scientist never ceases to be a lay person sharing the all too human cognitive biases and speaking the same vernacular language. Therefore, it is plausible that at least sometimes essentialist biases, encouraged by i-expressions, influence how scientists themselves interpret their findings and which further inferences they draw from scientific i-claims (Griffiths, 2002; Knobe & Samuels, 2013; Machery et al., 2019; Samuels, 2016). These misguided interpretations and inferences would thus infiltrate the relevant scientific theory. As different disciplines provide input to one another, wrongly interpreted results of one discipline can contaminate theorizing in other disciplines as well.

Granted, most expressions used in science originate from and continue to have parallel usage in lay discourse (e.g., 'force', 'energy', 'species', 'belief'). Insofar as this is so, the two threats are not unique to i-expressions. But with some expressions, including i-expressions, the threats are elevated. First, bad science and bad interpretation of science are more likely to occur if the folk concept associated with a given word in lay discourse is an attractor concept. In this case, breaking free from folk concepts associated with the expression is cognitively more difficult, and slippages into pre-theoretical thinking patterns are more likely. As implied by the theory of psychological essentialism, INNER ESSENCE is an attractor concept and i-expressions are those expressions that tend to be pulled towards this attractor concept. For this reason, i-expressions are among the words that are particularly susceptible to the above-named threats (see also: Machery 2021). Second, with some expressions and scientific claims, the two threats are more significant because they involve higher risks, namely, of having practically perilous consequences. That one occasionally falsely interprets 'force' as a term in physics in terms of the lay concept of force may be an obstacle for making theoretical progress in physics, training physicists, communicating physical theory etc., but will hardly have far reaching practical or moral consequences. In contrast, misguided essentialist interpretation of the genetics of human differences can have tangible practical consequences. They can help induce negative stigmatizing attitudes against out-groups, bad social policy, detrimental life choices and health interventions strategies (Haslam et al., 2006; Heine et al., 2017; Keller, 2005; Mandalaywala et al., 2018).

The papers of this thesis in this context

In the face of these two threats of bad science and badly interpreted science, two projects are relevant. First, it is relevant to monitor and scrutinize the usage of i-expressions in the context of scientific explanations so as to check whether a given i-expression is used in connection with an explanatorily useful concept; to identify where this is not the case; if possible, to refine the concept associated with the i-expression so that it would better serve the explanatory purpose it is

meant to serve. Second, it is relevant to clarify and communicate the content of i-expressions in scientific contexts and thus the content of respective i-claims. Both projects have a venerable tradition in philosophy. The three articles and a manuscript that compose this thesis contribute to each project. “Elusive Vehicles of Genetic Representation” engages in the first type of project. It looks into whether ‘genetic information’ is used in connection with an ontologically motivated concept. “Innate mind need not be within”, and “Causal Social Construction” in combination with the manuscript “Socially Constructed and/or Genetically Caused” contribute to the second type of project. They clarify, respectively, the concept expressed by ‘innate’ in the context of cognitive science’s innateness hypotheses and the concept expressed by ‘has genetic causes’ in the context of research into the genetic causes of human traits. In the next section, I will clarify the context and background of each of the four papers in more detail.

3. Individual papers in context

The four papers of the thesis, and each paper individually, engage with and combine topics and literature from different subfields of philosophical research. Giving an overview of each of these sets of literature would be impossible in the current format. In the sections to come, I will give background to each of the four papers individually. My focus will be on the part of the background that deals with debates concerning the respective i-expression – as the project of clarifying the content and implications of different types of i-claims in which certain i-expressions occur is the linchpin of this thesis. I will *not* aim to provide exhaustive overviews of these debates. Such overviews have been given elsewhere. Rather, in the sections to come, my priority is to clarify and outline certain aspects of the dynamics and proper methodology of the relevant discussions as they matter for understanding the status and coordinates of my own papers. At the end of each thematic section I will quickly indicate where the relevant paper is positioned relative to this background and give a quick summary of the contents of the paper.

3.1. Innateness and "Innate Mind Need Not Be Within" (IMNW)

'Innate' is the most paradigmatic and most discussed of the expressions that in the vernacular mind are likely associated with INNER ESSENCE. The word has a long history of established usage within various empirical sciences and philosophy (see (Samet, 2019) for a historical overview). But it also has a long history and established presence in vernacular English (as do its equivalents in other languages). The latter makes 'innate' when used in explanatory contexts particularly susceptible to the kinds of threats that I pointed out at the end of the previous section. Partly because of that, what concept is expressed by 'innate' and whether it should figure in our theories about biological and human phenomena has been the topic of rich debates in 20th and 21st century philosophy of science. The paper "Innate mind need not be within" borders upon these discussions. The paper clarifies and maps out the relationship between two views: concept nativism and causal externalism about mental content. Concept nativism is the common view within cognitive science that concepts are innate. Causal externalism about mental content is the prominent view within philosophy of mind that the content of a concept is determined by causal relations between the organism that possesses the concept and its external environment. In IMNW I clarify why, given the meanings of 'innate' in cognitive science and related disciplines, there is no inferential shortcut from the thesis that a concept is innate to the conclusion that causal externalism is not true of the content of the concept. This clarification relies upon existing philosophical analyses of the scientific concept of innateness. In what follows, I will outline the general landscape of contemporary philosophical discussions around the concept(s) of innateness and position IMNW within it.

The concept of innateness

Contemporary philosophical discussions over the concept of innateness emerged in response to two developments in the behavioral and psychological sciences in the middle of 20th century. Both developments manifest a turn away from behaviorism which had dominated these sciences before and which had been committed to empiricist (i.e., anti-nativist) views on cognitive development. The first of these developments was within behavioral ecology where certain theories of animal behavior that explained much of animal behavior as instinctive gained prominence, where 'instinctive' was used as a near synonym to 'inborn' or 'innate' (e.g., in the work of Konrad Lorenz). The second development was the "cognitive turn" within linguistics and cognitive science, largely associated with, and driven by, Chomskian linguistics. It was characteristic of this turn and the consequent evolution of cognitive science that one posited rich innate mental, including representational, structures in explanations of linguistic and cognitive development and functioning. This trend lives on in many strands of cognitive science, developmental psychology and evolutionary psychology.

Central to both of these turns is the thesis that animal and human minds are equipped with relatively rich innate contents. This thesis assumes that 'innate' refers to a meaningful and epistemically useful category. Many existing philosophical discussions over innateness target precisely this assumption. These discussions center around two entangled questions.

1) What *is* the concept expressed by 'innate' (i.e., the concept of innateness) when this or that trait is described innate?

2) Does this concept play a legitimate role in scientific explanation?

These questions are motivated by the observation that although 'innate' is frequently employed in scientific contexts, it is not always explicitly defined – nor is it clear without defining – what the word is supposed to mean and which property or properties it is meant to pick out.[12] Philosophers have attempted to clarify all this.

It is quite possible that 'innate' expresses a well-defined concept without this concept playing any legitimate explanatory role. However, those who think that the concept of innateness plays no legitimate role in explanatory contexts typically think so because they think that there is *no* well-defined scientific innateness concept (Griffiths, 2002; Machery et al., 2019; Mameli & Bateson,

[12] This much is generally said and taken for granted that innate traits are those that are not acquired by learning (Carey, 2009; Samuels, 2004). However, this is generally not considered to suffice as a specification of the concept of innateness – it only amounts to rephrasing the challenge of articulating the content of 'innate' as one of spelling out the content of 'not acquired by learning' in a manner that would be at once precise to the extent of being informative, and extensionally adequate to actual innateness ascriptions by scientists. This has proven difficult to do (e.g., Mameli, 2008).

2011; Matteo Mameli & Bateson, 2006; Shea, 2012). These innateness-skeptics often conclude with the upshot that 'innate' should be expelled from scientific discourse (I will return to this point in section 4). Correspondingly, the view that there *is* a decent concept expressed by 'innate' in scientific contexts is typically defended hand in hand with the view that this concept also has a genuine epistemic role to play: that of capturing a relevant kind/property that underlies useful generalizations, of serving as a useful heuristic etc. (Ariew, 1999; Birch, 2009; Cofnas, 2017; Khalidi, 2016; Mallon & Weinberg, 2006; Margolis & Laurence, 2013; Northcott & Piccinini, 2018; O'Neill, 2015; Perovic & Radenovic, 2011; Samuels, 2004).

A number of different optimistic analyses of the scientific concept of innateness have been proposed. I will not say much more about the content of these different analyses.[13] Instead, I will make some comments about the nature of these analyses in order to clarify the commitments, context, premises and ambitions of INMW, but also of the other papers of the thesis.

Analysis of a scientific concept

Different kinds of project can be, and have been, pursued under the name 'analysis of a scientific concept' (see e.g., Kraemer, 2018; Stotz et al., 2004; Waters, 2007). The same is true of analyses of the scientific innateness concept. Most saliently, an analysis of the scientific innateness concept can aim at making explicit the concept that is actually used in science when a trait is deemed innate (call it 'descriptive analysis'), or at enhancing the concept that is actually in use so as to make the concept better serve the epistemic purposes it is meant to serve (call it 'ameliorative analysis' or 'recitfying analysis' as does Kraemer). Most existing analyses of the scientific concept of innateness combine both the descriptive and ameliorative components. To some extent this is unavoidable. An amelioration of a concept must comply with, and thus presupposes an understanding of, certain essential aspects of the actual concept that it purports to enhance – lest it amount to a creation of a new concept rather than an amelioration of an existing one. And as Waters (2004) argues, a descriptive analyses of what a concept *is* will always partly build upon an understanding of what the concept *should* be. However, analyses of innateness can – and do – significantly differ with regards to which project, descriptive or ameliorative, is prioritized. It is important to be explicit about which project one prioritizes because it has consequences for the appropriate methodology and success conditions of an analysis. [14]

[13] See (Griffiths, 2021) or (Gross & Rey, 2012) for overviews of existing analyses of the scientific usage of 'innate'.

[14] Often one is not explicit about whether one is pursuing an ameliorative or descriptive analysis of the concept of innateness. But it can sometimes be inferred from the methodology one has chosen. For example, Knobe and Samuels (2013) and Cofnas (2017) must be seen as foremost in the business of clarifying the *actual* concept of innateness. This shows itself in the fact that they use data about scientists' actual judgements about which features of a trait qualify it as innate as their sole evidence for identifying what their concept of innateness is. Samuels'

Which project *should* be prioritized depends upon the broader context in which the analysis is undertaken. For example, in philosophical discussions one often seeks to define the concept of innateness in order to defend using the concept in explanatory contexts. Given this aim, the ameliorative project is in place. Suppose we need to decide if wine is an appropriate ingredient of a sauce or, instead, the sauce would taste better without wine in it. The correct answer does not depend upon whether any of the worst kinds of wine of those available, or wines gone bad, would enhance the taste of the sauce. It depends upon whether some reasonably good wine of those available would do so. Likewise with the concept of innateness. The legitimacy of a concept of innateness in explanatory contexts does not depend upon the epistemic pros and cons of the version of the concept of innateness actually employed insofar as this version might be a bad version of the concept. Instead, it depends upon whether there is *some* version of the concept of innateness that would serve certain epistemic aims, where this version need not be the one actually employed (although it would have to reside in the semantic vicinity of the actually employed version). However, if the broader goal of analyzing the concept of innateness is to illuminate the evidential potency and implications of actual innateness ascriptions – as is the goal of IMNW – then the ameliorative analysis is of little value. Instead, an adequate description of the concept of innateness that a relevant theory actually employs is relevant.

But a descriptive analysis of an actual scientific concept of innateness could still be after different things. Namely, 'the actual scientific concept of innateness' is ambiguous. It can mean any of the following (and more):

- what scientists explicitly believe 'innate' means, i.e., which criteria they explicitly believe determine whether 'innate' applies to a trait.
- which criteria of falling under 'innate' scientists actually rely on in their innateness judgements.

---

and Ariew's (2006) accounts on the other hand appear to have a notable ameliorative orientation. This shows itself in the fact that both measure the adequacy of a candidate definition, first, by how well the definition serves the relevant purported epistemic aims of employing the concept of innateness and, second, by its fit with some core intuitions of native speakers about what it means to be innate. One might wonder: how is the latter supposed to indicate that one is proposing an ameliorative analysis? Does it not, instead, indicate that one is analyzing the actual *lay* concept (as Kraemer (2018) diagnoses). It does. However, consulting native speakers' intuitions in order to identify the core features of the lay concept can function as an instrument of figuring out what a scientific term should or shouldn't mean. For instance, if a scientific concept diverges too far from the lay concept of innateness, it might not anymore count as a concept of innateness (i.e., a concept properly expressed using 'innate'). Or one might want a good scientific concept of innateness to align with the core of the folk concept in order to prevent chronic misinterpretations of the scientific innateness hypothesis in terms of the folk concept. In my view, the charitable interpretation of Samuels and Ariew is that they consult lay intuitions precisely with this aim in mind.

- which property or kind 'innate' in fact tracks in the contexts of well-supported scientific theories according to which a trait is innate.[15]

These three "actual scientific concepts of innateness" need not coincide and getting knowledge about each requires different methods. For instance, if one is interested in what scientists explicitly think 'innate' mean (their semantic beliefs about 'innate'), a reasonable thing to do is to ask scientists what they think 'innate' means.[16] If one is interested in which concept in fact guides scientists' innateness judgements – regardless of whether or not they are themselves aware of it – then one way to go is to observe in which conditions scientists actually employ 'innate', i.e., which factors their innateness judgements are in fact sensitive to. For instance, Machery et al. (2019) and Knobe and Samuels (2013) employ this method to ascertain scientists' concept of innateness.

What is of interest in IMNW is the third of the options which I will call 'operative scientific concept of innateness'.[17] The aim of IMNW is to ascertain and articulate the implications of theories in cognitive science according to which a concept is innate for causal externalist theories of mental content. In other words, its aim is to answer the following question: assuming that a hypothesis according to which a concept is innate is confirmed with certain methods, what then follows from this hypothesis for the truth or falsity of causal externalist theories? In order to evaluate this, what needs clarification is which property or set of properties it is that a given method in fact reliably tracks when the word 'innate' is applied. Put in epistemic terms: we need to know which set of properties (*qua* innateness) one is justified in believing a trait to have, given that an innateness hypothesis has been confirmed with a given method. This property may coincide with what scientists as individuals *think* 'innate' picks out, or it may not. Here's why. Suppose that according to scientists' explicit concept of innateness, to be innate is to have a certain set of properties p_1, p_2, p_3 (call this set '{innate}'). The scientists then devise a method M, including an operationalization of {innate}, for finding out which traits instantiate {innate} and which do not. Now, once in application, M can either succeed in reliably detecting {innate} (i.e., M can get it right reliably enough that a trait has/has not p_1, p_2 and p_3) or M can fail to reliably detect {innate}. If M fails to track {innate} then M either tracks nothing at all or M tracks some different set of properties, {innate*} (e.g., a set of properties p_1, p_4 and p_5). If M tracks no set of properties at all then

[15] Sally Haslanger (e.g., 2005) makes very similar distinctions when discussing what an analysis of a lay concept may amount to.

[16] I am not aware of studies of scientists' concept of innateness that would employ this method. But Stotz et al. (2004) provide an example of studying the scientists' explicit concept of a gene.

[17] My use of the term 'operative concept' differs from how Haslanger (e.g., 2005) uses the term. She calls the explicit lay concept 'manifest concept', the criteria that actually guide scientists innateness judgements 'the operative concept', and uses 'externalist concept' for something similar (though not identical) to what I label with 'operative concept'. For various reasons, I find 'operative concept' better suited for my purposes than 'externalist concept'.

an innateness hypothesis confirmed by M will have no systematic implications. If M reliably tracks a different set of properties {innate*}, then the implications of an innateness hypothesis hang from what {innate*} is, and so regardless of whether or not it coincides with {innate}. Of course, in either case, the scientist might still *believe* that M tracks and gives knowledge of the instantiation of {innate}; she might still *interpret* the results gained by M as showing that {innate} is instantiated; but she would be mistaken.

How do we find out what {innate*} is? First, although scientists' explicit and implicit judgements about what innateness is can be mistaken about which properties M in fact tracks, both serve as evidence for finding out what these properties are. Given that M is designed to track innateness, then what M in fact ends up tracking *qua* innateness depends upon what those (scientists) who design M believed about what it means to be innate. However, as said, such beliefs are but evidence of what M in fact tracks and not constitutive of it.

The second source of evidence is to see how M operationalizes being innate, that is, which features $f_1 \ldots f_n$ of a trait a given test employed in a study directly measures as indicating innateness. However, even though $f_1 \ldots f_n$ are surely among what M in fact reliably tracks, we should not prematurely assume that $f_1 \ldots f_n$ or any other operationalization of innateness exhausts {innate*}. For instance, a common empirical test in developmental psychology used to judge whether a cognitive ability is innate is to see if this ability is present early in life. But this does not mean that innateness *qua* what the innateness theories in developmental psychology track is nothing but being present early in life. To adopt, by default, that {innate*} is nothing but how innateness is operationalized seems to assume some version of operationalism according to which science only gives us knowledge of what its methods directly measure. This view is suspect. It leads to a far too extreme and absurd kind of skepticism regarding what empirical science enables us to know. A more reasonable approach is to be open to the possibility that existing empirical tests for detecting innateness do not merely give us knowledge of the fact that a trait has features $f_1 \ldots f_n$ that these tests directly measure but, in addition, knowledge of the fact that the trait has some further set of properties which $f^1 \ldots f^n$ indicate. It is these further properties that are the candidate for {innate*}. Of course, it might sometimes turn out that $f_1 \ldots f_n$ indicate nothing beyond themselves, in which case {innate*} is identical to $\{f_1 \ldots f_n\}$. But this is something that should not be assumed as a rule. In order to evaluate what this further set of properties is, and in order to correct for the possible mismatch between what this set really is and what scientists' explicitly or implicitly believe that it is, we need an epistemological inquiry into what a given method that is used to confirm an innateness hypothesis is fit to give knowledge of.

This is far from a worked out account of how to detect the operative scientific concept of innateness but it shall suffice here. Let it be said that in IMNW I do not propose any analysis of the operative scientific concept of innateness of my own. Instead I defer to existing philosophical analyses of the scientific concept

of innateness and assume that they do reasonably well at capturing the operative concept.

For convenience, I have been talking about *the* scientific concept of innateness. In reality, there is not much reason to assume that there is only one (or, alternatively, no) scientifically operative concept of innateness. 'Innate' is deployed in different fields and subfields of research (e.g., cognitive science, developmental psychology, cognitive linguistics, evolutionary psychology, behavioral ecology, various branches of biology). In these different fields, innateness hypotheses are established with different methods and for different purposes. Given this, it is plausible that in these different contexts different concepts are operative behind 'innate' and that different things are shown of traits when they are demonstrated to be innate. This kind of conceptual pluralism appears to be no exception with scientific terms in general (see e.g., Brigandt, 2020; Stotz et al., 2004) and is to be expected with the term 'innate' in particular. Even if it turns out to be possible to provide a descriptive analysis of the concept of innateness such that this analysis adequately captures the operative concept of innateness in a large array of innateness hypotheses across different fields, this analysis is likely to be so general as to be uninformative about the implications of such a hypothesis for any concrete question.

A certain degree of innateness pluralism is generally acknowledged among philosophers. It is common to consider separately the concept of innateness as it is used in the biological sciences and in the psychological sciences (Gross & Rey, 2012). How much more pluralist it is reasonable to go – should we expect there to be a multitude of innateness concepts *within* biological science and *within* cognitive science – is a further and more controversial question. For instance, Mameli and Bateson (2006; 2012) list six clusters of properties that in different empirical contexts 'innate' is intended to refer to. But the conclusion they and many others draw from this variation is that there is *no* coherent scientific concept of innateness – the different properties that 'innate' in fact refers to in different contexts may well be relevant properties to refer to, however, 'innate' is not the appropriate word for referring to these properties.[18] Whether they are right or not is surely an important matter, but does not matter where the aim is to illuminate the content of hypotheses that in fact predicate 'innate'.

Innate mind need not be within

So a reasonable strategy when it comes to the philosophical analysis of a scientific operative concept of innateness is to focus on narrower contexts and aim at *local* analyses of innateness which would illuminate the content of 'innate' in specific research programs. The research program of interest in IMNW is cognitive science and its adjacent fields. IMNW does not take a stand on whether the innateness-talk within the cognitive sciences is legitimate. It needn't do so in order to do what it purports to do. Instead, the paper stands on the following assumption.

[18] This sentiment is shared by Griffiths (2002).

Either there is an epistemically decent innateness concept operative in cognitive science's hypotheses that declare a concept to be innate or there is not. If there is not, then these innateness hypotheses have no, or have little, knowledge value and therefore no evidential bearing upon theories of mental content. If there is an epistemically decent notion or a set of notions of innateness operative in these hypotheses, then in order to assess the implications of these hypothesis, an analysis of this concept is needed. In IMNW I draw upon extisting widely discussed philosophical analyses of the concept of innateness designed specifically to capture the content of 'innate' within cognitive science: the account according to which a concept is innate only if it has not been acquired by a psychological process; the account according to which a concept is innate only if it is acquired in informationally impoverished conditions; and the account according to which a concept is innate only if its acquisition has not been caused by experience. None of these accounts has been explicitly devised as a descriptive analysis of what I call the operative concept of innateness in cognitive science's innateness hypotheses. However, in the paper I assume that at least one of them provides an accurate enough analysis of such a concept. In fact, I find it plausible that *within* the field of cognitive science there is a plurality of operative innateness concepts. If so, it is plausible that each of the analyses captures the operative meaning of 'innate' in some well supported innateness ascriptions within cognitive science; and that as a disjunction, they capture the operative meaning of 'innate' in all innateness ascriptions within cognitive science. I discuss the implications of concept nativism for causal externalism about conceptual content for all these accounts. Intuitively, causal content externalism may seem in tension with concept nativism. Both in lay and theoretical contexts the predicate 'innate' is strongly associated with the idea of organism-*internal* origin and determination. Thus the following reasoning is inviting: *Insofar as a concept is innate, and insofar as concepts are individuated by their content, the content of the concept must be determined by organism-internal factors and thus cannot be determined by organism-external factors, as causal externalism has it. Therefore, the scope of externalism must at least be limited to those concepts that are not innate*. A conclusion in these terms is explicitly drawn by Pitt (2000) and suggested by other philosophers. In IMNW I explain why this reasoning is not valid.[19] The weight of the paper is not so much on proving *that* causal externalism can be true of innate concepts. After all, in some circles this much is taken for granted (albeit only with regards to *some* versions of causal externalism).[20] The weight is on

[19] That a clarification of this sort is needed, and that the threat of leaping into the intuitive *innate-therefore-not-externally-determined* reasoning is lurking in the background, can be seen from the genealogy of the paper. The paper was initially meant to defend the thesis opposite to that which it actually defends and quite similar to the one that Pitt defends (before I stumbled upon his paper). I then soon realized that I had based my judgement on the pretheoretical assumption that what is innate must be determined by what is inside and *not* on consulting the concept of innateness operative in empirically well-grounded concept nativism.

[20] For example, Fodor (1997), Cummins (1981) and Gross and Rey (2012) all take for granted that innate concepts can be externalistically individuated and, moreover, that they

spelling out in detail the reasons *why* this is the case and highlighting the variability of these reasons. Not only are there plausibly different concepts operative in different instances of concept nativism. As I highlight in IMNW, the category "causal externalism" also comprises significantly different kinds of theories of mental content. Whether, and under which conditions, a concept can be both innate and externalistically individuated does not turn on what these theories share as *externalist* theories – the generic thesis that a concept's content is determined by causal relations between an organism and its external environment. Instead, it turns upon which *kinds* of causal relations to the external environment determine the concept's content, and this can vary from one type of causal externalist theory to another. I distinguish between three types of causal externalism based on how they specify this kind of relationship. I then show that regardless of which type of externalism and which analysis of innateness we assume (a) there is no conceptual contradiction between being innate and being externalistically individuated, albeit for different reasons; (b) whether a particular concept can be innate and externalistically individuated is an empirical rather than conceptual question.

3.2. Genetic information and genetic causation

The rise, growth and prominence of genetics in the 20th and 21st century has had a wide and deep impact on our scientific view of the living world. It has also greatly impacted our vernacular worldview and discourse. We think and talk in terms of genes. To some extent, knowledge gained from genetic research has genuinely changed and enriched folk-biological views. In some cases, however, the impact appears to be merely cosmetic or terminological in that gene-related glossary appears to be lain over old ingrained conceptual structures. The fact that expressions referring to the genetic origin of a trait such as 'in the DNA', 'genetically caused' etc. are often used interchangeably with, and have come to replace 'innate' in lay discourse (Griffiths, 2020; Linquist et al. 2011) suggests a phenomenon of the latter kind. It suggests that new labels have been glued on an old concept earlier associated with 'innate'. This is also corroborated by empirical findings to the effect that 'in the DNA', 'in the genes' etc. are, just like 'innate', associated with INNER ESSENCE (Dar-Nimrod & Heine, 2011). Dar-Nimrod and Heine ague that in the lay-mind genes are perceived as the material carriers of organism's inner essence and, correspondingly, traits believed to be genetically caused are seen as emerging from this essence. They call this phenomenon 'genetic essentialism' (Cheung et al., 2014; Dar-Nimrod et al., 2014, 2021; Dar-Nimrod & Heine, 2011; S. Heine, 2016; Heine et al., 2017; Linquist et al., 2011). Admittedly, available evidence is inconclusive regarding how pervasive genetic

cannot be inernalistically individuated. But all these authors assume specific versions of causal externalism and the reasons why they take externalism to be compatible with nativism do not generalize to other versions of causal externalism.

essentialism is. Even if essentialist biases do influence the way people interpret genetic information, they are definitely also other influences. The impact of essentialist tendencies may be mitigated, outweighed, screened off by other factors; essentialist interpretations of genetic information might only manifest in certain circumstances, etc. (Cheung et al., 2014; Dar-Nimrod et al., 2014, 2021; Dar-Nimrod & Heine, 2011; S. Heine, 2016; Heine et al., 2017; Linquist et al., 2011). However, this much is uncontroversial that sometimes, with some traits and in certain conditions, people interpret claims about genetic origins of traits as if genes constituted the inner essence of an organism.

Parallel to the lay usage, expressions indicating the genetic origin of traits are also employed in research contexts, where traits are said to be 'genetically caused', 'encoded in the DNA', 'genetically programmed', 'under genetic control 'etc. As with 'innate', the content of such expressions in explanatory contexts is not always clear and, as with 'innate', philosophers of biology have attempted to explicate and evaluate this content.[21] "Elusive Vehicles of Genetic Representation" and the tandem of "Causal Social Construction" and "Socially Constructed and/or Genetically Caused" contribute to these discussions. In the next sections, I will position these papers against the backdrop of earlier philosophical discussions on the concept of genetic information and genetic causation.

3.2.1. Genetic information and "Elusive Vehicles of Genetic Representation" (EVGR)

One important strand in the philosophical discussions on matters concerning genes is motivated by the observation that biologists often describe genes as carrying information about the phenotypic traits that they cause. A variety of informational vocabulary is used to describe genes' relationship to traits. Among other things, genes are said to code, program, instruct, provide blueprints or receipts for the development traits etc. For brevity, I will call claims like these in which genes are said (or implied) to carry information about traits 'genetic information ascriptions' (I use 'trait' and sometimes developmental effect' broadly to include all sorts of downstream effects of genes, be it proteins, gene expression traits, metabolic traits, whole organism traits of different kinds; I use 'gene' broadly to denote whichever genetic entity is at issue). Philosophers have been discussing the status and legitimacy of biologists' genetic information ascriptions. Such ascriptions might be legitimate – or not – for different reasons. For instance, informational predicates might provide a useful heuristic or a useful way to model the role of genes in development and/or in evolution (e.g., Levy (2011) defends biologists' talk of genetic information from a fictionalist and instrumentalist viewpoint; see (Griffiths & Stotz, 2013, Ch. 6) for discussion and overview). However, often in the philosophical discussions the legitimacy of scientific genetic information ascriptions is assessed by assessing whether or not these

[21] 'Innate' and expressions referring to the genetic origin of a trait are sometimes used interchangeably also in scienctific contexts (Cofnas, 2017; Mameli, 2008).

ascriptions are true.[22] Do genes *really* carry information about traits rather than just causing traits? Do biologists' information-predicates pick out a relation that genuinely and objectively differs from *merely* causal relations?

Let it be noted that these questions are often discussed in the broader context of discussing the assumed gene-centrism of genetic research. Namely, it is sometimes claimed that biologists often do not view genetic and environmental causes as on par. Instead, they tend to view genes as playing a qualitatively different role in the development of traits than environmental causes, and tend to view genetic causes as more relevant than environmental causes. Frequently, the hypothesis that genes bear information is discussed as one possible reason why such views on genes might be correct: genes are different and/or more relevant than environmental causes of traits in virtue of carrying information about traits. This corresponds to what appears to be a standard view among biologists themselves. Here I will be discussing the problem of genetic information independently of its bearing on gene centrism.

Views vary regarding the ontological status of biologists' information talk (Godfrey-Smith and Sterelny (2016) give an overview). Some think that informational predicates are indeed but metaphors and/or that biologists' talk of information in genes can be reduced to causal claims (Griffiths, 2001; Griffiths & Gray, 1994; Oyama, 1985; Šustar, 2007). Others defend one or another realist account of genetic information (e.g., Sarkar, 2004; Shea, 2007). I will say more on different types of such accounts later on. But first, I will outline the criteria (and variation in these criteria) that one commonly employs – or if not, then *should* employ – to decide if genes contain information about traits, and thereby the success conditions of a philosophical realist account of genetic information. Not always are these criteria explicitly mapped out. Mapping out these criteria will serve to locate and delineate the limits and premises of my own criticism of the teleosemantic account of genetic information in EVGR.

The truth conditions of 'genes carry information'

Under which conditions is it true – as many biologists seem to assume – that genes indeed carry information about some relevant type of trait? To ascertain these conditions, we need to recall the aim of asking whether genes contain information about traits. The aim, as here specified, is to assess if scientists' genetic information ascriptions (claims in which genes are said or implied to carry information about traits) are true. Thus, a philosophical defense of an affirmative answer to this question can be considered successful if it demonstrates the truth of scientists' genetic information ascriptions. We can parse the truth conditions of scientists'

[22] One option is that genetic information ascriptions in biology are merely a metaphoric way of speaking. It is a separate question then whether informational metaphors – if metaphors they are – are epistemically or practically useful metaphors (as per Levy (2011)) or detrimental as according to Griffiths (2002), Griffiths and Gray (1994), Oyama (1985), Keller (2000) and many others.

genetic information ascriptions into two. Each specifies a success condition of a philosophical account of genetic information *qua* a defense of the claim that genes bear information about traits. First, any claim that ascribes informational content to genes is true only if:

(1) Given some correct metaphysical account of carrying information, genes qualify as carrying information about traits.

Determining if (1) is the case is partly an empirical project that requires determining whether as a matter of empirical fact genes relate to certain traits in such a way that meets the metaphysical conditions upon carrying information about these traits. But it is not an *exclusively* empirical project. What constitutes information carrying in the first place is an open question, and a philosophical one for that matter. Partly, debates over the presence, absence and nature of genetic information consist precisely in debates over this issue. Correspondingly, disagreement over whether genes meet (1) can also boil down to diverging views about what the relevant criteria of bearing information are.

But the truth of (1), the metaphysical thesis, does not suffice for the truth of biologists' genetic information ascriptions. For hardly would the truth of (1) vindicate scientists' genetic information ascriptions if, given the notion of carrying information that makes (1) true, what biologists say and think *of* genetic information turned out to be false. So, a successful philosophical defense of the thesis that genes bear information about traits must not only demonstrate that there is a concept of information that makes (1) true but also that, given this concept of information, (2) is true.

(2) What scientists claim and assume *about* genetic information is true.

This is the second truth condition of biologists' genetic information ascriptions. Of course, (2) should not be read as meaning that positing genetic information is legitimate only if *everything* that biologists say and assume of this information is true. Scientists claim and assume various things about and in relation to genetic information in different biological contexts and surely it must be possible for some of this to be mistaken without this undermining the genetic information discourse *tout court*. Views vary as to which claims and assumptions of scientists about genetic information are those that would have to come out true if their information talk is to be vindicated. Here it suffices to point out two assumptions that are commonly seen as central to scientists' views on genetic information and as such taken to constrain philosophical defenses of view that genes carry information about traits.

One is the assumption that only genes, and not environmental causes of developmental outcomes, carry information about developmental outcomes. This assumption is often called 'the uniqueness assumption'. Correspondingly, a common counterargument to philosophical accounts of genetic information is that

given the notion of carrying information that this account assumes, environmental causes of development also turn out to count as bearing information about traits.

The second common assumption is implicit in the fact that scientists often posit genetic information as explanans in explanations of the development of traits. Correspondingly, a metaphysical account of genes carrying information is successful and vindicates the truth of biologists' information-talk only if it maintains the truth of sentences of the following form:

S. That genes carry information about trait T explains the development of T.

I state S ambiguously enough to leave room for different options as to *how* genetic information is supposed to explain the development of T. This much is agreed in philosophical discussions on the topic that genetic information explains T in an interesting sense only if it adds something over and above non-informational, causal explanations, of the development of T. However, there can be disagreement over some further details as to how and what exactly about the development of T genetic information is to explain. I will come back to this in the end of this section.[23]

The criteria (1) and (2), often applied in a mixed manner, may pull in different directions. It can be true that genes carry information according to some metaphysical account of carrying information while it is not true that genes' carrying information in this sense plays the explanatory role that biologists ascribe to it (*mutatis mutandis* with the uniqueness assumption). Thus, depending upon which criterion one prioritizes, the answer to whether genes carry information (*qua* whether scientists' genetic information ascriptions are true) can come out differently. If one puts little weight on (2) and more weight on (1) then an account of genetic information can be considered successful even if, given the metaphysics of information carrying that this account appeals to, relatively few claims and assumptions of scientists come out true. The more weight one puts on (2), the more of scientists' claims and assumptions about genetic information would have to be rendered true by the account.[24]

Which of (1) and (2) *should* be prioritized depends partly, on my view, upon whether the particular biologists' assumption at issue when assessing the truth of (2) is *constitutive* of biologists' concept of genetic information – part of the scientists' concept of genetic information – or not. If the assumption is part of the biologists' concept of information, (2) sets stronger constraints on a successful account of genetic information. Here's an illustration of how and why.

Let INF be a correct concept of what it is to carry information and let it be true that genes relate to (some) traits as INF specifies. (1) is thus met. Now suppose

[23] E.g., Levy (2011) objects to Shea's teleosemantic account of genetic representation on account of failing to make true S, given the relevant interpretation of S.

[24] For instance, Stegmann's (2009) response to Sarkar (2003) reveals a difference between these two authors in how much relevance either attributes to (2). Sarkar treats (2) – including the uniqueness assumption and the explanatory relevance assumptions – as a central success criterion of an account of genetic information, while Stegmann does not.

that what biologists mean by carrying information when they say that genes carry information about traits is not INF but INF*. Not only is INF* not identical to INF but it is different enough from INF to not count as an ameliorated version of INF. This could be for various reasons. It could turn out that INF* is not really a concept of information at all – it does not capture what it really means to carry information – but a causal concept that biologists falsely or non-literally use in association with 'information'. Alternatively, INF* might not be a coherent concept but a non-cohesive conglomerate of different conceptions. Or INF* and INF conceptualize different types of such relationships that we properly call information-carrying relationships. In all of these cases, showing that genes carry information, given INF, does *not* prove biologists' genetic information ascriptions to be true. If INF* is not a coherent concept, then biologists' information ascriptions are nonsensical and thus lack truth conditions – no truth, likewise the truth of (1), could prove these ascriptions to be correct. If INF* is a concept all right, albeit not a concept of information, then the fact that genes carry information as per INF does not show biologists right when they claim that genes bear "information" because biologists never really claimed genes to bear information, despite the appearance to the contrary. If INF* is a concept of information, albeit one that is different from INF, then demonstrating that genes contain information given INF amounts to a change of topic rather than a demonstration that biologists are right about gene's carrying information. Because the proposition that was to be evaluated was *genes carry INF*ormation* and not *genes carry INFormation.*

In the philosophical discussions on the topic, it is not always clear if the uniqueness assumption, the explanatory role assumption or some other relevant assumption that biologists make about genetic information are taken to be constitutive or non-constitutive of biologists' concept of genetic information. However, as I have just suggested, this does make a difference. For example, if explaining the development of traits is taken to be part of what biologists mean by (genetic) information, then an account of genetic information that implies that genes do not explain development of traits would straightforwardly imply that the account is not an account of genetic information *qua* what biologists are talking about. If explaining the development of traits is but a contingent property of genetic information, then the fact that (1) implies the falsehood of this assumption on a given concept of information can be more easily sidelined in the face of other considerations. What identifies biologists' concept of information and under which conditions is this concept similar enough to some other concept of information is a further question that turns on theories of conceptual content and concept individuation, and one that I will not address here.

My own criticism of the teleosemantic account of genetic information in EVGR turns only on (1). For this reason, I now elaborate some more on the conditions of carrying information and how genes have been thought to meet these conditions.

Information bearing

What it means to carry information, i.e., under which conditions can genes be said to carry information about traits and thus meet (1), is no trivial question. Typically, arguments to the conclusion that genes indeed carry information about traits proceed by identifying the features that gene-trait relationships share with other, less controversial, paradigmatic instances of information bearing. However, there is a motley of phenomena that we acknowledge as 'information bearing', and a motley of accounts of what makes those phenomena paradigms of information bearing. In the discussions of genetic information, one commonly distinguishes between two broad kinds of information: correlational information (or Shannon information) and semantic information. That genes carry correlational information about traits is trivially true and thus not at issue. An entity X carries correlational information about another entity Y insofar as X correlates with Y. This is so whenever X and Y are variables that can assume different states, and the instantiation of a state y* of Y increases the likelihood of the instantiation of some state x* of X. Applied to genes and traits, a gene carries information about a trait whenever having a certain version (an allele) of the gene increases the likelihood of having a certain version of the trait. That genes carry information about some traits in this sense is trivially true, being implied by the undisputed thesis that genes cause some traits in the sense measured as statistical correlation. Moreover, correlation is symmetric: as much as genes carry information about a trait, the trait also carries information about genes on this conception. Thus, the concept of genes carrying information about traits in the thin sense of correlating with traits does not make sense of the common assumption that genetic information is typically conceived to be instructive, and that only genes are thought to instruct the production of traits and not the other way around.

So, what is at issue in philosophical debates is whether genes carry information about traits in some thicker, "semantic", sense of the word, a sense that has affinity with the concepts of meaning, aboutness (where 'semantic' is often used synonymously with 'representational' or 'intentional').[25] But this achieves little in the way of specifying what the interesting sense of carrying information is. The label 'semantic' draws together a heterogenous array of phenomena. First, we talk of semantic or meaningful phenomena in different spheres. On the one hand, we recognize and talk of meaningful phenomena (pictures, words, thoughts, traffic rules, symbols etc.) in pre-thoeretic contexts. On the other hand, semantic terms and concepts are employed in more or less technical contexts. There is the

[25] Godfrey-Smith (1999) argues that 'semantic information' within the debates over genetic information is a misnomer. This is because semantic phenomena like meaning cannot be analyzed via the concept of information if by 'information' one means correlational/Shannon information – meaning, despite a common view, is not a subspecies of information so understood (William Ramsey (n.d.) is on a similar view). I will continue using 'semantic information' regardless of whether the phenomenon typically so called is analyzable in terms of Shannon information or not, and whether it is properly called 'information' or not. This aligns with the standard practice.

talk of "inner representations" in cognitive science; of "codes" and "programs" in computer science etc. These special-purpose technical uses originate from vernacular usage. However they have taken up an aspect or other of some vernacular semantic concept which has then been tuned to serve some specific theoretical purpose. In consequence, semantic concepts in technical contexts may have little overlap with their pre-theoretic ancestor concept as well as with one another (see (Godfrey-Smith, 2004) for a discussion in relation to propositional attitudes). Second, within both of these spheres, we can observe a *variety* of semantic phenomena. This reflects in the diversity of semantic terminology used to denote different types of semantic phenomena: 'template', 'code', 'blueprint', 'program', 'recipe' etc. While all of these things may be said to bear content, they might bear it in different ways. Thus it is quite plausible that 'semantic information' hides a genuine pluralism or a disjunctive kind (Godfrey-Smith, 2004; Shea, 2013; Shea 2018). Moreover, it is a further question what defines something as a code, a representation, a map or a vehicle of semantic information, and thus, which properties should genes share with the relevant paradigmatic instances of information carriers in order to be of the same kind with them. Lack of consensus in such matters is evident in long-lasting disputes within philosophy of mind about the nature of intentionality.

Depending on which paradigm of carrying semantic information one assumes for comparison, and which features one assumes to be essential to this paradigm, the answer to whether genes bear information may again turn out differently. It also has consequences for *what* they carry information about. An information-bearing entity necessarily carries information *about* something. Any account of carrying information is an account of what the relation between two entities X and Y would have to be for this relation to be one of X bearing information about Y. Genes carry semantic information about exactly those traits that they relate to via this relation.

Existing accounts of genetic semantic information

Existing accounts of genes bearing information about traits differ both in terms of what they take to define genes as carriers of semantic information and what they take genes to carry such information about – gene products, developmental programs, whole organism traits or something else (for overview: Godfrey-Smith and Sterelny (2016), Griffiths and Stotz (2013)).[26] For purposes of presentation, I roughly divide such existing accounts into two clusters based on which mark of the X-Y relationship one takes to be characteristic of a semantic relationship: arbitrariness or the possibility of failure.[27] Accounts of both clusters share a

[26] Those who reject the view that genes bear information can and do likewise differ in which notion of information carrying they take to be the relevant one.

[27] Godfrey-Smith (1999) proposes a different twofold taxonomy of accounts of genetic semantic information: "analyses in terms of developmental role [of genes]" and "analyses in

general framework with regard to what is required of X to be a vehicle of semantic information. This framework is articulated by Godfrey-Smith who calls it 'the basic represetationalist model'. Godfrey-Smith presents this model as a model of *representing* shared by many naturalistic accounts of representation. However, the model generalizes to instances that some who prefer a narrower sense of 'representation' might not want to call so (Godfrey-Smith himself uses 'representation' rather inclusively). Godfrey-Smith sets off with the general idea which he takes to be the very core of the pre-theoretic concept of representation, of meaning, of having content – that for some X to represent Y is for X to stand in for Y (as a map stands in for a landscape, 'Maria Stuart' stands in for a concrete human individual etc.). He then articulates the stands-in-for relation as follows: X stands in for Y if the X–Y relation is exploited, or exploitable, by some separate system which one standardly calls the consumer or reader of X. X-Y relationship is exploitable by a consumer if the latter can act towards Y by way of acting on X, *viz.* by way of using X as an indicator that a particular behavior towards Y is appropriate (or as Godfrey-Smith puts it: if "X is consulted as a guide to behavior directed on Y"). For example, a map of Tartu represents Tartu because it is possible for a user of the map to find one's way through the streets of Tartu by way of consulting the map and adjusting one's behavior according to the features of the map.[28] One recognizes a system as operating with "stand-ins" in the described sense from its causal structure: the system can be identified to have components (X, Y, Z) such that one of these (X) can take different values (x', x'' etc.) and each of these values systematically causes Z to produce a specific output behavior which is appropriate given some specific value of Y (y', y'' etc.). With genes, this structure is thought to be implemented in that a developmental mechanism – the consumer – systematically produces a certain output in response to consuming a certain allele. If one type of allele is received, trait t' is produced; if a different allele is or would be received, a different trait t'' is or would be produced.

However, having this kind of causal structure is necessary but not sufficient for the system to count as one operating on representations. Systematic and specific causal relationship is simply that – a systematic specific causal relationship. There has to be something more that permits to characterize the relationship between X (e.g., a gene) and Y (e.g., a trait) as X representing Y. The two theory clusters mentioned earlier are distinguished by what they take this relevant "more" to be. Accounts of the first cluster rely on the idea of the symbolic nature of paradigmatic representations (words, signs etc.). Symbols mean, but what they mean is arbitrary in the sense that they might as well mean something different –

terms of evolutionary history" (p. 16). Extensionally, this taxonomy more or less overlaps with the one drawn here – however, the clustering criterion is different.

[28] In the map case, the relationship between a map and what it is a map of, e.g., a town, is exploitable by some consumer due to isomorphism between the map and the town. But the X-Y relationship can be exploitable in virtue of other features too. See Shea (2018) for a comprehensive discussion of some options.

by no law or necessity is a symbol tied to its particular meaning, to what it symbolizes. It is argued that this also characterizes the way in which a DNA sequence specifies the amino acid primary sequences in the process of protein synthesis – i.e., that the connection is arbitrary (Flament-Fultot, 2014; Godfrey-Smith, 1999, 2000; Wheeler, 2007).[29] Upon receiving an mRNA molecule transcribed from a given DNA sequence, the cell's protein synthesis mechanism (the consumer of a "genetic message") produces a particular protein. Which protein is produced upon receiving a given mRNA molecule is determined by a fixed mapping rule that maps specific ordered DNA base triplets, codons, onto specific amino acids. This mapping rule is argued to be arbitrary in the sense that there is no chemical or physical necessity why a given codon corresponds to a given amino acid – given the chemical and physical properties of the DNA triplets and the amino acids, the mapping might as well have been a different one. In this sense it resembles a conventional interpretation rule that ties symbols with their meaning (Godfrey-Smith (1999, 2000) and Griffiths and Stotz (2013, Ch.6) give a more detailed accessible description of the biological mechanism and its symbol-like nature). I shall refer to this cluster of accounts as 'symbolic accounts'.

Accounts of the second cluster take their paradigm of the semantic to be the concept of representation in the philosophy of mind and cognitive science. These accounts stress that essential to representation (content, meaning) is having success and, correspondingly, failure conditions (e.g., Shea 2007; Stegmann 2009). An indicative representation (e.g., the sentence "Mary Stuart is the queen of Scots") indicates or describes things as being in a certain way. It does so successfully if it describes them truthfully – if things in the world indeed are as described (e.g., Mary Stuart is the queen of Scots) – and fails to do so if it does not describe them truthfully. An instructive or imperative representation ("Execute Mary Stuart!") instructs that things be in a certain way. It does so successfully if the receiver of the instruction complies with the instruction, i.e., behaves so as to make things be in the way instructed. The content of a representation is defined (partly) by the conditions under which the representation is successful. Philosophical theories of content have wrestled with the challenge of telling a convincing naturalist story about what determines the success conditions, and thus the content, of a representation. If we assume that having success and failure conditions of the described kind defines semantic information, the same challenge is faced by those who argue that genes carry instructions – construing genes as carrying instructions is legit only if it is possible to naturalistically explain what determines the conditions under which a genetic instruction is/is not successfully followed by the consumer of the instruction (development or some developmental subsystem).

One prominent response to this challenge has been teleosemantics (Jablonka, 2002; Maynard Smith, 2000; Shea, 2007, 2013b; Sterelny et al., 1996). Teleosemantic accounts of representational content, first proposed and developed to

[29] What arbitrariness exactly amounts to – causal, chemical, or evolutionary arbitrariness – is debated and contested (see e.g., Stegmann, 2004).

explain the nature and content of mental representation, derive the success and failure conditions of a representation from the selection function of this representation.[30] According to teleosemantic accounts of genetic information, an allele carries an instruction to develop the trait that the allele was selected for causing. This instruction is complied with if the development or some developmental submechanism indeed produces the trait. It is not complied with if the development fails to produce the trait. Most teleosemantic accounts identify the intentional object of genetic representations in naturally selected whole-organism phenotypes (Maynard-Smith 2000, Shea 2007, 2013). However, Godfrey-Smith (1999) also considers the option that, given the teleosemantic criteria, DNA segments represent the amino acids that they map onto in protein synthesis. Which traits genes represent on the teleosemantic account depends on what it is that genes are naturally selected for, which in turn depends on broader issues like what function as interactors in the process of natural selection.

Both symbolic and teleosemantic accounts are proposed as vindications of the claim that genes bear semantic information about traits. They are both presented as accounts of "genetic information". But it is important to note that they are nevertheless accounts of significantly different things. Both assert genes to have semantic properties, but they assert them to have *different* properties – being arbitrarily related to what they are about and having satisfaction conditions. These different properties also differ in extension. On the symbolic account, all and only those DNA segments that are transcribed in the process of protein synthesis and specify the primary sequence of a protein (known as the coding segments of the DNA) contain instructions and function as representation vehicles. According to the teleosemantic account, only and exactly those DNA segments carry instructions that have been selected for causing a trait. And the DNA units that have been selected for causing a trait and those that code a protein need not coincide (I address this in more detail in EVGR).

The accounts also differ in which explanatory role they ascribe to genetic instructions. Recall the criterion (2) used to evaluate an account of genetic information *qua* vindication of biologists' genetic information ascriptions. Namely, according to (2), an account of genetic information is successful if it renders the relevant parts of what scientists claim *about* genetic information true. I also suggested that at the core of what biologists claim about genetic information are claims of the form S:

S. That genes carry information about trait T explains the development of T.

But S is ambiguous. One commonly distinguishes between at least two kinds of explanations of trait development: explanation via proximate and ultimate causes, for short, proximate and ultimate explanations (distinction originally introduced

[30] Views vary about whose function determines the content of a (genetic) representation: the representation itself, the consumer of the representation, the producer of the representation, or the representation consuming system as a whole.

by Mayr (1961)). To provide a proximate explanation of trait development is to answer the question: *How* did the developmental process that produced a given trait (here: trait variant) unfold? An answer to this question would describe the causal pathway along which the development of the trait unfolds. To provide an ultimate explanation of the development of a trait is to answer the question: *Why* did the developmental process produce a particular trait? In cases where the trait of interest is adaptive in its environment, and this adaptive match cannot be explained by environmental input during ontogenesis, one answer is – the trait was produced because in the past generations, this trait made organisms possessing the trait fitter than organisms lacking the trait and thus got selected. The factors that played a causal role in the selection of the trait are the ultimate causes of the development of the trait in contemporary organisms. Given this taxonomy of explanations, S can be interpreted to mean either S* or S**.

> S*. Genetic information explains *how* the developmental process that produced a given trait unfolds.
>
> S**. Genetic information explains *why* the developmental process produces a given (adaptive) trait?

Symbolic accounts and teleosemantic accounts both render S true, or at least have been argued to do so by their proponents. However, they make true different interpretations of S. According to the symbolic accounts, genetic instructions are part of the causal mechanism that produces proteins which are then used to build or regulate further downstream traits. As such, they play the role of explanans in proximal ontogenetic explanations of traits. These accounts preserve the truth of S*. In contrast, genetic instructions understood in terms of teleosemantic accounts feature in ultimate explanations of traits, namely, in evolutionary explanations of why the members of a lineage of organisms manifest a particular trait. They *cannot* feature in proximal explanations. According to teleosemantic accounts, genes' representational properties are historical properties – properties they have due to their (selection) history and would not have if this history was absent. And historical properties are standardly thought to lack proximate causal powers.[31] Teleosemantic accounts of genetic information therefore preserve the truth of S** but not S*.

We can ask: are S* and S** both *relevant* interpretations of S? If a successful account of genetic information is one that makes S true, then which interpretation of S should we prefer? Insofar as the project of scrutinizing whether genes have semantic properties aims at clarifying scientific ontology, this depends upon in which kinds of explanation biologists in fact adhere to genetic information. I will not address this empirical matter here. Let's just note that how these empirical facts turns out to be has consequences for whether a particular account of genetic

[31] See Griffiths & Stotz (2013 Ch.6.) for a more fine-grained discussion of the possible disambiguations of S, and Griffiths and Stotz (ibid.) and Shea (2007) for a discussion of the explanatory limits of teleosemantic accounts of genetic representation.

information counts as a success as per (2). For example, if biologists only (or mostly) adhere to genetic information in proximate ontogenetic causes of trait development, then teleosemantic accounts are off-topic and fail to vindicate biologists' information talk.[32] For instance, Levy (2011) objects to Shea's teleosemantic account of genetic representation on such grounds. However, a more plausible alternative is that biologists ascribe informational properties to genes in *both* proximal (developmental mechanistic) explanations and ultimate (evolutionary) explanations. For example, Godfrey-Smith & Sterelny (2016) list five explanatory contexts in which scientists make use of genetic information as explanans, ultimate and proximate explanations among them. In this case, teleosemantic and symbol theories are appropriately seen as explicating different segments of biologists' information discourse. A third option is that an account of genetic information identifies an altogether new explanandum that biologists do not yet explain in informational terms but should. This is a legit contribution. However, in this case, the account should be recognized, and argued for, as making a contribution of this kind.

This was to clarify that whether genes carry information about traits depends on a variety of background decisions, for example about what is the prototypical information carrying relationship and which features are essential to this prototypical information bearing relationship. Different philosophical defenses of the thesis that genes to contain information about traits can involve diverging views on any of these issues. Until these background decisions are made explicit, discussions on whether or not genes really carry information remain prone to merely verbal disputes due to equivocations at various levels of analysis.

Criticism of the teleosemantic account of genetic representation in EVGR

Most previous criticisms of teleosemantic accounts of genetic information in the literature have turned on criterion (2). For instance, teleosemantic accounts have been criticized for failing to account for the right explanatory role of genetic information (e.g., Levy 2011). In EVGR, I test the account against criterion (1). I focus on Nicholas Shea's defense of the teleosemantic account. Shea (2007, 2013b) adopts a modification of Ruth Millikan's teleosemantic account that has been articulated and applied in the philosophy of mind to capture the nature of the content of simple representations: simple mental representations, animal signals and others. Shea argues that on this account, the genetic inheritance system – the mechanism by which naturally selected whole-organism traits are passed down generations – qualifies as a representational system with genes representing the traits that they were selected to cause. I argue that in light of what we know of the role of the DNA in the process of natural selection, the genetic inheritance system does not comply with the teleosemantic criteria so as

[32] In his later work, Shea (2013b) attempts to demonstrate that genetic representations occasionally figure also in developmental explanations. This has been judged problematic (Griffiths & Stotz, 2013; Planer, 2016)

to support the thesis that genes represent whole-organism traits. My criticism is similar to, complements, and is complemented by that of Godfrey-Smith (2008) and Planer (2014) (the latter of whom I regrettably failed to cite in my own article). Both point out that the structure of the system in which genes are supposed to function as representations does not comply with the teleosemantic model of representation that Shea applies. I point out an additional and more comprehensive reason why this is so.

3.2.2. Genetic causation. "Socially Constructed and/or Genetically Caused" (SCGC) and "Causal Social Construction" (CSC)

Regardless of whether genes contain information about traits, that genes among other factors *cause* traits is uncontroversial. Yet, although *that* genes cause traits is uncontroversial, many issues related to the concept and phenomenon of genetic causation are not, and continue to feed lively philosophical discussions.

One salient segment of such discussions concerns the parity or non-parity of genetic and non-genetic (henceforth 'environmental') causes of traits. Recall that the hypothesis that genes carry information about traits is often discussed as a plausible hypothesis explaining why genes would be qualitatively different form environmental causes of development. But there is no necessary link between carrying information, being qualitatively different and being more relevant. For example, the proponents of the teleosemantic theory of genetic information make no issue of the implication of their theory that some environmental causes too can represent the traits that they cause (e.g., Shea, 2013b; Sterelny et al., 1996). Also, genes can qualitatively differ from environmental causes of traits in virtue of being certain *kinds of* causes without this amounting to being an information carrier. Whether or not this is so is something that philosophers have often explored in relation to genetic causation.

The question of parity/non-parity of genetic and environmental causes can be discussed from an ontological, epistemic or pragmatic perspective**.** For example, Kitcher (2001) supports the thesis of parity of genetic and environmental causes as an ontological thesis. But he maintains that, epistemically, genes and environment *qua* causes of development are not on par: in many contexts, privileging genetic causes is justified because doing so serves heuristic purposes.[33] However, predominantly, the issue of parity is discussed as an ontological issue. Ever since the developmental systems theory came into prominence, and especially over the past couple of decades when epigenetic research and different brands of systems biology have been prominent in biological theory, the burden of proof in philosophical discussions has been on the proponents of the non-parity of genetic and non-genetic causes. The most common way to argue for the view that genes

[33] Gannet (1999) and Longino (2013) are skeptical about this epistemic non-parity thesis. They argue that often the privileged attention on the genetic causes of traits in scientific practice is for practical – e.g., economic – rather than theoretical reasons, and often for bad practical reasons.

indeed are ontologically different from other developmental causes, and in some sense more important, is to argue that genes are the most – or at least *among* the most – specific causes of the linear sequence of the amino acids that make a protein (this being the relevant trait). X is a specific cause of Y if it is possible to bring about an array of fine-grained changes in Y by bringing about fine-grained changes in X. Put somewhat differently, X is the more specific cause of Y the more there are such changes $x_1 \ldots x_n$ in X that would bring about changes $y_1 \ldots y_n$ in Y so that there is mapping from X to Y such that this mapping is close to being a function (e.g., Woodward, 2010, p. 203). This kind of relationship is argued to hold almost exclusively between the coding segments of the DNA (the "molecular genes") and the linear sequence of the primary structure of proteins that are synthesized in a cell (Calcott, 2017; Waters, 2007; Weber, 2017; Woodward, 2010; see Griffiths & Gray 1994, Moss 2003; but see (Stotz 2006) for an opposing view).[34] It is worth stressing that specificity is not transitive. That genes are specific causes of protein structure does not imply that genes are specific causes of any of the downstream traits that proteins feed into. For example, with most whole-organism traits, it is safe to assume that many environmental factors are much more specific causes of these traits than any molecular gene. Thus even if the specificity of the DNA-proteins relationship makes DNA different from, and perhaps more important, as a causal factor in protein synthesis, it does not make it different and more important a cause of more complex traits.

Irrespective of whether genes are in some relevant sense different from other causes of traits, there is a more fundamental question of philosophical interest: what does it mean for genes to cause a trait in the first place? This is far from clear. First, there's all reason to believe that 'x causes y' is used ambiguously in both lay and scientific contexts to speak of different kinds of phenomenon (causal relevance, causal explanation, causation, different types of causation). Secondly, the nature of causation (given some disambiguation 'x causes y'), and different aspects thereof, is a topic of perennial philosophical dispute. The various questions that arise in these general discussions also arise in relation to specifically genetic causation. *What does it mean for genes to cause a trait?* is the question that I am concerned with in SCGC. Before turning to the focus of SCGC, notice that, as it stands, the above question is ambiguous. For instance, it can be interpreted as any of the following:

1) What is it for a trait T to have genetic causes? This is a metaphysical question about the nature of genes *qua* causes of traits.

[34] All cited authors concede that genes are not the only specific causes of the primary sequence of proteins. What they argue is that there are few other causes (if any) that are *as specific* as genes. In addition, none of these authors think that being a specific cause of proteins suffices for making genes special. E.g., according to Waters (2007) genes are among the most specific *actual* difference making causes of protein structure. According to Weber (2017), genes are the most specific *potential* difference making causes of protein structure, given that the changes in DNA sequence that would bring about changes in protein sequence are "biologically normal".

2) What does 'T has genetic causes' mean in the context of scientific research? I argued in 3.1 (when the notion of 'scientific concept of innateness' was discussed) that this question is itself ambiguous. It might concern either the scientists' explicit beliefs on what it means for a trait to have genetic causes. It might concern the criteria that scientists in fact tend to employ when judging a trait to have genetic causes. Or it might be about what I called the scientifically operative concept of having genetic causes – the property of traits that the methods of empirical science in fact track when what they track is called 'has genetic causes'.

3) What does 'T has genetic causes' mean, given the lay concept of having genetic causes? This question is likewise ambiguous in the manner just described.

Answering (1), (2) and (3) requires different methods and answers will likely differ. For example, Waters (2007) argues that when biologists demonstrate that genes are (among) the causes of T then what they demonstrate is that genes are (among) "the actual difference making causes" of T, that is, causes that explain actual differences in T in some target population where being a cause is analyzed in terms of Woodward's interventionist account of causation. This analysis is primarily intended as an answer to (2), and, in particular, to *What is the operative concept of having genetic causes in scientific contexts?*.[35] In addition, Waters himself thinks that his account also provides a correct answer to (1) – to be an actual difference making cause is to be a special type of causes, and whether something is one is a matter of objective ontological fact. So, on Waters' view, what biologists track *qua* genetic causes *are* in fact genetic causes metaphysically speaking.[36] But he *might* be wrong about the latter without this undermining his account as an account of the former. Lay people's concept, implicit or explicit, of having genetic causes is in turn likely to correspond with neither what empirical methods in fact track under the label 'genetic causation' nor what genetic causation really is. For instance, in section 2 I hypothesized that both the explicit

[35] Waters himself does not explicitly draw these kinds of distinctions, nor does he use the term 'operative concept'. However, from what he says I infer that his main aim is indeed to capture what I call the operative concept of having genetic causes. To give more examples, Bourrat (2019, 2020), Russo and Wiliamson (2007), Joffe (2013) also elucidate the concept of genetic causes operative in different biological sciences, although all of them, just like Waters, think that this operative concept coincides with what it *really* means to have genetic causes.

[36] That what empirical studies show *qua* genes causing a trait can diverge from what it *really* is for genes to cause a trait is shown by the lively discussions on whether heritability estimates are indeed estimates of genetic causation. For discussion and overview: Lynch and Bourrat (2017), Sesardic (2005, pp. 23–27), Tabery (2014 Ch.3), Bourrat (2019). My own view is that a trait's heritability is nothing but the extent to which genes cause actual differences in the trait in a population (see (Bourrat, 2019) for an interpretation of heritability that supports this view). Therefore, the question of whether heritability is indeed an estimate of genetic causation boils down to whether being an actual difference making cause (in Waters' sense) of a certain kind amounts to being a cause.

and the implicit lay concept of having genetic causes is imbued with essentialist conceptions. However, if Waters is right and the content of scientific ascriptions of genetic causes boils down to the thesis that genes are (among) the actual difference making causes of a trait, then scientific ascriptions of genetic causes lack most of the essentialist implications that lay people commonly associate with genetic causation.

The question I am concerned with in SCGC is (2). In particular, I am interested in the scientifically operative concept of having genetic causes. My aim in SCGC is to clarify what follows from the empirical finding that a human trait has genetic causes for social constructionist accounts of that trait. For example, assuming that schizophrenia is shown to have significant genetic causes, then what follows from this for the view that schizophrenia is, at least to some extent, socially constructed? Given this aim, what needs to be clarified is the concept of having genetic causes that determines the empirical content of such findings irrespective of how this concept fares in light of some metaphysical theory of genetic causation. In SCGC, I argue that this concept and thus the content of empirical findings to the effect that a human trait has genetic causes is best analyzed in terms of a contrastive counterfactual dependence account of causation (defended for instance by Hitchcock (1996), Northcott (2008), Schaffer (2005) and (2003)). And I demonstrate how this account proves useful in social constructionist discussions for assessing the compatibility of genetics findings and social constructionist accounts. I will sketch out the core idea of the contrastive notion of causation after some remarks on the motivation for and the context of SCGC.

Socially constructed and/or genetically caused?

The claim that a human trait or a fact about a human trait is socially constructed is common in the social sciences, social and critical philosophy. To argue that a trait is socially constructed is to argue for some sort of contingency of T on social factors, where this contingency on social factors is typically thought to have further political, social, or moral importance. What *kind* of contingency, and on *which* social factors – this can vary, and significantly so, from one social constructionist account to another.[37] However, amidst this variability, most social constructionist theories share the following assumption: arguing that a trait is socially constructed aims, among other things, to establish that the trait is not natural or biological in some relevant sense of the word. Opinions differ as to whether not being biological is part of the very concept of being socially constructed (as Diaz-Leon (2015) suggests) or strong evidence of not being socially constructed, but this difference does not concern me here. As a flip side of the constructed-therefore-not-biological assumption, a common strategy to argue against a social constructionist account of a trait is to present evidence for the claim that the trait *is* biological. This typically amounts to presenting evidence for the claim that the trait has

[37] Moreover, this is at times left vague. I address issues like these in my paper "Causal Social Construction".

important genetic causes (e.g., Sesardic, 2010; Shiao et al., 2012). Findings to the effect that genes cause a trait are seen as a *prima facie* challenge, or at least a constraint, for the social constructionist accounts of the trait by both social constructionists themselves and their opponents. With the ever-growing amount of genetics findings about a variety of human traits, and the perceived epistemic authority of such findings, social constructionists and social scientists more broadly are increasingly more compelled to respond to this challenge and engage with genetics.

SCGC aims at providing theoretical tools for taking up this challenge. This project is in part motivated by my judgement that some frequently pursued strategies of dealing with the challenge are unsatisfactory. One such strategy is the following. A recurring way to push back against the pressure from genetics on social constructionist accounts is to appeal to the non-dichotomous nature of being socially and genetically caused. It is stressed that the development of all traits has *both* genetic and environmental, often social, causes; and that, as a rule, genetic and environmental causes interact – the effect of a gene depends upon which environmental causes are operative, and *vice versa*.[38] Taking this as a premise, one then argues along the following lines (to summarize and simplify)*: Indeed, T (a human trait) has genetic causes; but as everyone knows, all traits have both genetic and environmental causes, many of which are social by nature; therefore, that T has genetic causes does not imply that T does not have social causes nor therefore that a social constructionist analysis of T is false.*[39] Call the thesis that all traits are caused by both environmental and genetic causes which interact 'the interactionist thesis'. And call the above-described strategy to respond to the finding that a trait has genetic causes 'the interactionist pushback'.

Why is the interactionist pushback unsatisfactory? One thing that might at first glance seem wrong with this pushback is that it seems to be in tension with the aforementioned *socially-constructed-therefore-not-biological* assumption. According to the interactionist pushback, T's having genetic causes does not rule out that T is socially constructed (here: in the sense of being socially caused) because every trait that has genetic causes also has environmental causes, some of which might well be social. But if this is the case, then demonstrating that T is socially constructed shouldn't suffice for a demonstration that T is not biological in the sense of having genetic causes either – because all traits that have social causes, e.g., socially constructed traits, also have genetic causes. But this apparent tension need not amount to a genuine inconsistency. As mentioned above, 'genes cause T' can mean different things and thus also the claim that T has genetic causes. It might well be the case that the interactionist thesis together with the

[38] Note that 'gene-environment interaction' is ambiguous. For example, Tabery (2007) points out that the term is being used in at least two senses: in the sense of biometric or of developmental interaction. Here, what is meant is the latter. However, as Ferreira Ruiz and Umerez (2021) point out, this latter sense also involves vaguenesses and ambiguities.

[39] For example, I take Gannet (2010) and Fausto-Sterling (2005) to employ this kind of reasoning, albeit in far more elaborate versions.

other claims embedded in the interactionist pushback is true on one meaning of 'genes cause T',[40] while the socially-constructed-therefore-not-genetically-caused assumption is also true, albeit on a different meaning of 'genes cause T'.

However, while acquitting the proponent of the interactionist pushback of the accusation of inconsistency, this reasoning reveals the real issue with the interactionist pushback. Namely, that it is off target. Let's take for granted that in the context of the interactionist pushback and in the context of the socially-constructed-therefore-not-genetically-caused 'genes cause T' mean different things. Now, as stressed, when the aim is to evaluate whether an empirical claim to the effect that T has genetic causes undermines the thesis that T is socially constructed, then the relevant concept of 'genes cause T' is the concept that in fact contributes to the content of this empirical claim, that is, the operative concept of having genetic causes. The relevant task is to see if, given *this* sense of having genetic causes, T's having genetic causes speaks against social constructionism about T. The interactionist pushback assumes that the operative concept of having genetic causes that determines the evidential bearing of genetics against social constructionist accounts is the one that makes the interactionist thesis true. But this assumption cannot be correct, for at least three reasons. First, one typically appeals to the interactionist thesis as if it was a trivial empirical fact. Let's suppose that it is. If so, then it is very uncharitable to assume that the extensive and expensive empirical research undertaken to identify if this or that trait has significant genetic causes is in the business of identifying if the trait has genetic causes merely in the sense of the word that makes the interactionist thesis trivially true. For why take pains to identify trivia? Second, it would be uncharitable to assume that those (assumedly informed people) who in fact do bring findings from genetics to bear against social constructionist accounts would do so if T's having genetic causes in this sense – the sense in which the fact that genes cause a trait is obvious and does obviously little harm to social constructionist accounts – exhausted the content of these findings. Third, I take it that all parties consider it absurd to claim that empirical findings to the effect that genes cause a human trait *never* conflict with social constructionist accounts of the trait. But this is precisely what would follow if the concept of being a cause that is operative in these findings was the one that makes the interactionist thesis trivially true. Therefore, the concept of having genetic causes that is operative in empirical findings must be some less inclusive concept than the one that makes the interactionist thesis true.

[40] For example, given Woodward's (2003) interventionist definition of being a cause, widely agreed to provide an accurate characterization of what scientists track when providing causal models, it is indeed trivially true that all traits have both genetic and environmental causes. According to Woodward's account, "X is a cause of Y" means the following: there are background circumstances B such that if, in B, we changed the value of X (and no other variable) without changing the value of any other cause of Y, then Y or the probability distribution of Y would change. This much is trivially true of all X and Y such that Y is a trait and X is some genetic variable, and of all X and Y such that Y is a trait and X is an environmental variable.

In SCGC I argue, relying on Northcott (2012), that the contrastive counterfactual dependence account of causation provides useful tools to articulate and represent this less inclusive concept. The general idea of counterfactual dependence accounts of causation is this: 'X causes Y' means *if X had not occurred then Y would not have occurred.* The contrastive supplement adds that a causal claim always specifies, even if implicitly, some relevant causal and effect *contrasts*. That is, a causal claim specifies a range of relevant alternatives to X and Y: which alternative non-actual scenarios X* and Y* we need to consider when judging whether Y had, or had not, occurred if X had not occurred. According to the contrastive account, ascriptions of genetic causes to traits can be represented in the following way.

A gene G causes an individual i to have trait version T^a iff either (i) or (ii) is the case:

(i) If i, given E^a, had possessed G* rather than G^a, then i would possess T* rather than T^a

(ii) If i, given G^a, had experienced E* rather than E^a, then i would not have possessed T* rather than T^a

where G^a is i's actual genome (or some allele of a concrete gene or whichever genetic entity is the putative cause of possessing T^a in a research context), E^a is the entirety of i's actual developmental environment or some specific feature of it (depending on the research context), T^a is the version of T that i actually possesses, and G*, E* and T* are respectively some non-actual alternatives to E^a, G^a and T^a or classes of such alternatives. The claim that genes cause a trait is true just in case: if i had had the particular relevant alternative G* (assuming that (i) is the relevant condition), i would have the specific relevant alternative trait T*. For simplicity, I chose to articulate the idea of the contrastive counterfactual dependence account on the example of causal claims about token traits. Typically, human genetic research (e.g., GWAS, twin studies) is interested in causal relations between gene and trait distributions in a population. The above schema can be adjusted to express the content of causal claims with different ontological kinds of causal relata, including gene/trait distributions.

Given this contrastive understanding of the concept of having genetic causes, it is not trivially true that all traits have genetic causes. Here's why. Most philosophers agree that if a typical study into the genetic causes of a human trait reveals that the trait has genetic causes, then what it reveals is that the trait counterfactually depends upon genes or does not depend upon environment: if the genes of an organism (or genetic distribution in a population) were different, then the trait of the organism (or distribution of the trait in a population) would be different as well *(mutatis mutandis* for developmental environment) (Joffe 2013; Russo and Williamson 2007; Waters 2007; Woodward, 2010). On this basic version of the counterfactual account, all traits have genetic causes as matter of trivial fact. For every actual trait of an organism there will be *some* conceivable alternative

to the organism's actual genes such that if the organism had had these alternative genes instead of its actual genes then the organism would not have its actual version of the trait. And for every actual trait of an organism there will be *some* possible environment such that had the organism experienced this alternative environment instead of its actual environment, the organism would nevertheless have its actual trait version. It is also trivially true that all traits have environmental causes: for each actual trait of an organism there will be *some* non-actual alternative environment such that had the organism experienced this alternative environment instead of its actual environment, the organism would *not* have its actual version of the trait.

In contrast to the basic version of the counterfactual account, the contrastive version of this counterfactual dependence view highlights that empirical findings from genetics don't merely show that in case of *some* alternative genes or environments the organism would have, or not have, a different trait. What these findings show is that swapping the actual genes or environment of an organism for some specific alternative G* or E* would have the effect that the organism would, or would not, have some specific alternative trait T*. And not all possible alternatives are relevant for whether T counts as having genetic causes in the sense scientists of a given field are interested in. Whether or not an individual would have a different version of some trait, had she had some *specific* alternative genome or had she experienced some *specific* alternative environment is far from being a trivial question.

I want to emphasize that I do not commit to the view that this contrastive schema provides a correct metaphysical theory of gene-trait causality. I only assume that the truths revealed in the context of much of empirical research into the genetic causes of human traits has this kind of structure, regardless of whether these truths count as metaphysical truths about genetic causation. However, I also want to stress that this contrastive account need not be the only correct representation of such truths.[41] I treat the contrastive account as a useful *model* that illuminates certain aspects of the empirically operative concept of having genetic causes that are particularly relevant for the project of assessing the implications of scientific ascriptions of genetic causation on social constructionist accounts. In SCGC, assuming this model, I outline a general framework for assessing such implications. The contrastive model emphasizes that the content and thus implications of a finding that T has genetic causes depends among other things upon how the contrasts E* and G* are filled in (which is often not explicitly clarified). In different research contexts the relevant contrasts can be filled in in a variety of ways. The implications of a finding to the effect that T has genetic causes – e.g.,

[41] For instance, the contrastive account is very much compatible with Waters' (2007) account of genetic causes as actual difference making causes and Bourrat's (2019; 2020) causal analysis of the concept of heritability and SNP-heritability. In fact, I think that at core they all correctly express the same operative concept of having genetic causes but in different ways and with emphasis on different aspects of this concept. This latter thought needs, of course, further argument.

whether or not it conflicts with a social constructionist account of T – vary correspondingly. The main original contribution of the paper lies precisely in stressing that the implications of the claim that a trait has genetic causes can, and should, be ascertained in part by identifying the causal contrasts involved in this claim. I illustrate this on the example of heritability studies.

"Causal Social Construction" (CSC) defends a premise of SCGC. In meta-discussions on the concept of social construction one standardly distinguishes between two ways in which something can be socially constructed: constitutively and causally (Diaz-Leon, 2015; Haslanger, 2003; Mallon, 2016). A trait is constitutively socially constructed if social factors constitute/define the trait. A trait is causally socially constructed if social factors cause the trait. In SCGC I discuss the compatibility of genetics claims with both kinds of constructionist claims. However, in order to do this, an obstacle needed to be overcome. Namely, no general analysis existed of how social factors would have to cause a trait in order for the trait to be (causally) socially constructed. In "Causal Social Construction" I provide such an analysis. I propose an explication of the notion of causal social construction in terms of the contrastive counterfactual dependence account of causation. [42] In part, this explication serves as an operationalization of the concept of causal social construction for the clarifications undertaken in SCGC (The core idea of SCGC is present in the last section of CSC). But it also makes a self-standing contribution to filling a gap in the discussions of social constructionism.

3.3. Summary points

I defined i-claims to be those scientific nativist claims that tend to be interpreted through folk essentialist conceptions and tend to be so interpreted partly because they contain an i-expression, an expression that in folk discourse is associated with the concept of inner essence. Although not all scientific nativist claims need to be i-claims, many of them are. The nativist claims that the papers of this thesis target are, on my judgement, also i-claims. In concert, the papers accentuate two things about such i-claims, and by extension about scientific i-claims in general.

First, they accentuate that different i-expressions that figure in different nativist claims can be hooked onto different operative concepts, and that the same i-expression can be hooked onto different operative concepts in different research programs and, sometimes, even within the same research program broadly construed (a point that Griffiths (2001) and Mameli and Bateson (2007; 2011) make about 'innate'). According to some criteria, these differences in operative concepts may be slight. However, they are significant insofar as they make a difference regarding what a given nativist claim implies for some further issue at hand. This possibility of pluralism in the meaning of i-expressions should function

[42] In a sense, I owe this paper to Karola Stotz who drew my attention to the fact that it is far from obvious that causal constructionist claims are to be modeled in terms of a counterfactual dependence accounts of causation, as I at first assumed.

as a kind of regulative idea in the project of assessing the implications of different nativist claims. It underscores the need for specific case studies, especially ones that pay attention to the differences rather than cross-contextual commonalities of the concepts involved. Many of the existing philosophical analyses of the concepts of innateness and genetic information and others have aimed at general applicability, and as such, they might not be well-fitted for this kind of project.

Second, the papers accentuate that scientific claims to the effect that a trait is innate, genetically caused etc. are compatible with various, sometimes unexpected forms of environment-dependence of the trait. Both IMNW and SCGC make explicit the reasons why traits deemed innate or genetically caused in scientific contexts can be both environmentally *caused* as well as environmental *by nature*, i.e., constituted by environmental factors. If this is the case, then it follows that an individual (or a population) that instantiates an innate or genetically caused trait (or a trait distribution) need not have instantiated the trait in some relevantly different environment. Moreover, assuming the contrastive counterfactual dependence account of genetic causation, a trait's property of having genetic causes is itself an environmentally supervening property – whether a trait counts as one having significant genetic causes can change even if no changes occur in the trait itself or the individuals possessing it. If teleosemantic theories of genetic representation were to be correct, this much would also be true of genes bearing representational information with a particular content. Teleosemantic theories of content are externalist theories – what a gene represents, and whether it represents at all, is determined by the history and historical environment of the genetic lineage. Thus, it is possible that if the phylogenetic histories of two genetically identical organisms O1 and O2 differ in relevant ways, then the genes of O1 and O2 carry *different* information, or even that the genes of O1 do carry information while the genes of O2 do not. But let me emphasize again that the ways in which, say, being genetically caused, or possessing an innate, genetically caused, or genetically encoded trait is a function of the environment, will vary depending on the specific concept expressed with 'innate', 'genetically caused' or 'genetically encoded'. Case-specific inquiry is required to ascertain what these ways are.

That many scientific nativist claims, including innateness- and genetic causation-claims, have these features might not be that newsworthy. But it is important to emphasize them and put them into the spotlight because these are the features that we are prone to overlook. First, many nativist claims contain i-expressions. By definition, different i-expressions are drawn towards the same (vague) concept INNER ESSENCE. For this reason, the contextual differences in the meaning of these expressions tend to be obscured. The likelihood that different i-expressions are interpreted to express the same concept is exacerbated by the fact that one frequently uses different i-expressions (loosely) interchangeably. In some confined contexts, this assumption of synonymy is legitimate – sometimes different i-words indeed express the operative concept. However, this cannot be *assumed* to be the case, especially if the expressions are used in different research contexts. Second, i-expressions *qua* expressions associated with INNER

ESSENCE tend to be perceived to indicate endogenous origin. Moreover, Berent (2020) makes the case that traits perceived to originate from an organism's inner essence, to be innate, also tend to be perceived as endogenous by constitution. To counterweigh these tendencies, the environment-dependence of "innate", "genetic" etc. traits is in need of constant restatement, attention, and clarification. Getting things right in this respect is relevant not least of all because it has consequences for the nature and scope of explanations that appeal to innate, genetically caused, naturally selected traits to explain some more complex human phenomenon. There is a widespread intuition (as suggested by Berendt's findings) that explanations of, say, complex behaviour and social phenomena with an innate trait as explanans are individualistic explanations – explanations in terms of features that supervene on the intrinsic properties of an individual. However, if an innate trait is one supervening on some relation of the organism to its broader environment, then such explanations have affinity with environmental, structural, explanations, with all their associated characteristics and limitations.

4. The prospects of eliminativism

In the beginning of this introduction, I motivated my dissertation as a response to two threats emanating from the fact that scientific claims are expressed using i-expressions: the threat that essentialist biases influence the content of science in inappropriate ways and the threat that essentialist biases influence the interpretation of scientific findings in inappropriate ways**.** I argued that these threats make relevant the task of monitoring, checking, clarifying, and communicating the content of scientific claims that contain i-expressions. To conclude this introduction, I want to address an alternative strategy that has been proposed to counter the threat of badly interpreted science. Namely, if using, say, 'innate' or 'genetic program' in science contexts induces the risks of bad science and badly interpreted science, then why not just stop using these words altogether rather than clarify their technical meaning? This is what some philosophers in fact argue in relation to some i-expressions: that these expressions should be expelled from scientific discourse altogether. For example, Griffiths (2002), Griffiths and Machery (2008), Machery et al. (2019), Machery (2021) argue for this with regard to 'innate' and Griffiths (2001), Keller (2000), Oyama (1985) reason along these same lines regarding 'genetic program' and its cognates. This alternative strategy might be seen as a superior alternative to the one of clarifying the meaning of different i-expressions in explanatory contexts for a method of preventing the misinterpretation of these expressions. I now want to suggest that it is not – that eliminating 'innate', 'genetic' or the like is not likely to be as effective an instrument of doing away with the threat of misinterpretation of scientific nativist claims in terms of i-expressions as its proponents think. At least it does not remove the need for continuing with the clarification project. I present my reply by using the example of 'innate'.

Call the position that 'innate' should be expelled from scientific explanations "innateness eliminativism". Innateness eliminativism is typically supported with a two-headed argument. One reason why 'innate' is said to have no legitimate place in scientific explanations is because the word allegedly invokes no coherent explanatorily useful concept. This reason will not interest us here. The second reason is that using 'innate' in scientific nativist theories is likely to trigger the folk concept INNER ESSENCE and thus give rise to misguided interpretations of these theories in terms of this concept, which is undesirable. This reason can be pursued independently of the first and one could argue that it is a strong enough reason to eliminate 'innate' *even* if it has a well-defined explanatorily useful technical meaning in some research context. For example, Griffiths (2002) seems to be of this view. He argues that if 'innate' happens to come equipped with a (semi)technical epistemically useful meaning in some of its scientific contexts of use, and probably comes equipped with *different* meanings in different contexts, then one should express this meaning in words that makes this technical meaning transparent, rather than by saying 'innate'. This would help prevent essentialist interpretations of i-claims as well as cross-contextual equivocation mistakes (*ibid.*, p. 82).

It is this second reason in support of innateness eliminativism that I take issue with. The reasoning stands on the premise that eliminating 'innate' and substituting it with a different fitting phrase would successfully block essentialist interpretations of the relevant scientific theories. This premise, however, is unwarranted given what was said in section 2: that (1) in (English) lay discourse, 'innate' is associated with INNER ESSENCE and (2) INNER ESSENCE is a universal developmentally fixed feature of human psychology that functions as an attractor concept. Let's take (1) and (2) for granted, as do the named proponents of innateness eliminativism. Granting (1) and (2), the association between 'innate' and INNER ESSENCE is surely contingent in the following two senses. First, given that INNER ESSENCE is a developmentally fixed human universal, possessing and deploying it does not presuppose having 'innate' – or any other word – in one's vocabulary. Second, there are, have been, can be and will be other expressions besides 'innate' associated with INNER ESSENCE as is evident if only from the observation that even in the English language alone certain expressions (e.g. 'in the DNA', 'evolved by natural selection', 'genetic') are often used synonymously with 'innate' (Linquist et al., 2011; see also section 2). This suggests two things. First, even if one lacks or ceases to employ 'innate', this does not yet mean that one lacks INNER ESSENCE or ceases to deploy this concept in one's thinking and reasoning about the world. Second, even if an expression (e.g., 'innate') that happens to be associated with INNER ESSENCE at some point in time stops being used, there will be other expressions that either continue to be used, or are recruited to be used in its place. Now, suppose that we do what the eliminativist calls for: we swap 'innate' for some other, more transparent and arguably more accurate expression in a scientific theory. The innateness eliminativist envisages that by doing so we make it significantly less likely that the theory is interpreted via INNER ESSENCE. But assuming (1), (2) and the contingency of 'innate'-INNER ESSENCE association, this is not very likely. Any plausible alternative to 'innate' would be an expression that already resides in the semantic vicinity of 'innate' and, with all likelihood, would itself already be associated with INNER ESSENCE as per *Association* (see section 2).[43] The new expression might be associated with INNER ESSENCE more weakly than 'innate'. However, being in the semantic "gravitation field" of INNER ESSENCE, in the long run the expression is likely to gravitate closer to INNER ESSENCE and become no less likely than 'innate' to trigger interpretations of the relevant theory in terms of INNER ESSENCE and essentialist beliefs – *unless* the content of these new expressions continues to be monitored and explicated.

[43] For instance, it is quite plausibly true of the expressions that Griffiths proposes as substitutes for 'innate': "Substituting what you actually mean whenever you feel tempted to use the word 'innate' is an excellent way to resist this slippage of meaning. If a trait is found in all healthy individuals or is pancultural, then say so. If it has an adaptive-historical explanation, then say that. If it is developmentally canalized with respect to some set of inputs or is generatively entrenched, then say that it is. If the best explanation of a certain trait differences in a certain population is genetic, then call this a genetic difference. If you mean that the trait is present early in development, what could be simpler than to say so?" (Griffiths 2002, 82)

This prediction is consistent with empirical findings. First, there is evidence for the negative thesis that a change merely in the words that are used to express scientific content does not significantly change how this content gets interpreted. For instance, Condit & Condit (2001) and Condit et al. (2002) demonstrate that the way the lay audience interprets claims about the genes-phenotype relationship is not predicted by whether 'genetic blueprint' or 'genetic receipt' is employed in these claims. Instead, regardless of which expression is used, how subjects interpret such claims is predicted by their pre-existing understanding of how genes relate to the phenotype. Grossi (2017) offers an illustration of how verbal change does not bring about conceptual change in a scientific context. Grossi observes that the usage of 'innate' has significantly decreased in psychological and brain sciences, attributing this shift to the fact that the concept of innateness has been broadly discredited as unscientific. She also observes that the decreased use of 'innate' is paralleled by the increased use of 'hardwired' which according to here is considered to express a concept more precise and scientific than the concept of innateness. However, she demonstrates that, in reality, the concept used in association with 'hardwired' appears to be nothing but the same concept earlier associated with 'innate' under the guise of a new word. Second, there is some direct evidence that clarifying the operative concept behind an i-expression *does* change how the expression is interpreted (Condit & Condit, 2001; Condit et al., 2002; Dar-Nimrod et al., 2021; Knobe & Samuels, 2013; Parrott et al., 2004; Samuels, 2016). That this is so is also consistent with abundant philosophical and empirical evidence for the more general proposition that much of how one interprets a word is explained by the context (broadly construed) of the interpreter. It is also consistent with the empirical fact, discussed in section 2, that whether one forms and entertains essentialist beliefs about a particular category is sensitive to the subject's background attitudes (see section 2 for references).

It would be going too far to claim that the kinds of shifts in terminology that the eliminativists propose have no effect whatsoever on the extent to which essentialist beliefs mediate our interpretation of scientific nativist claims.[44] What I do predict is that these effects will be modest and are likely to fade in time as the association with INNER ESSENCE is likely to reestablish itself. This prediction surely requires more empirical backing and philosophical argument. Empirical study into the mechanisms and factors that affect the interpretation of i-claims is still scarce (albeit growing). Even less have these mechanisms been investigated through the lens of philosophical theories, for instance, the relevant theories we find in the philosophy of language. It is a study of precisely this kind that I see myself undertaking in the future.

[44] E.g., Parrot and Smith (2013) provide evidence that sometimes, to some extent, swapping one plausible i-expression for a different one does effect interpretation. In addition, there is plenty of evidence for the more general view that linguistic choices do make a difference to how likely one is to form an essentialist representation of a category. For example, many studies show that using generic language when referring to a category (e.g., 'girls') can foster essentialist perceptions of the category in children (Gelman & Roberts, 2017). However, note that here the efficacious linguistic intervention concerns logical form rather than terminology. This and other considerations suggest that even if linguistic variations do make a difference to interpretation, these variations are not at the term level.

REFERENCES

Ariew, A. (1999). Innateness is Canalization: In Defense of a Developmental Account of Innateness. In A. Ariew (Ed.), *Philosophy of Science* (pp. 19–27). MIT Press, Cambridge, MA.

Astuti, R., Solomon, G. E. A., & Carey, S. (2004). Constraints on conceptual development: A case study of the acquisition of folkbiological and folksociological knowledge in Madagascar. *Monographs of the Society for Research in Child Development*, *69*(3), 1–135, vii–viii; discussion 136–161.

Bateson, P., & Mameli, M. (2007). The innate and the acquired: Useful clusters or a residual distinction from folk biology? *Developmental Psychobiology*, *49*(8), 818–831.

Berent, I. (2020). *The Blind Storyteller*. Oxford University Press. New York.

Birch, J. (2009). Irretrievably Confused? Innateness in Explanatory Context. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *40*(4), 296–301.

Bourrat, P. (2019). Heritability, Causal Influence and Locality. *Synthese*, *198*(7), 6689–6715.

Bourrat, P. (2020). Causation and Single Nucleotide Polymorphism Heritability. *Philosophy of Science*, *87*(5), 1073–1083.

Brigandt, I. (2020). How Are Biology Concepts Used and Transformed? In K. Kampourakis & T. Uller (Eds.), *Philosophy of Science for Biologists* (pp. 79–101). Cambridge University Press.

Cacchione, T., Hrubesch, C., Call, J., & Rakoczy, H. (2016). Are apes essentialists? Scope and limits of psychological essentialism in great apes. *Animal Cognition*, *19*(5), 921–937.

Calcott, B. (2017). Causal Specificity and the Instructive–Permissive Distinction. *Biology and Philosophy*, *32*(4), 481–505.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press. Oxford, New York.

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.

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *The Behavioral and Brain Sciences*, *37*(5), 461–480.

Cofnas, N. (2017). Innateness as genetic adaptation: Lorenz redivivus (and revised). *Biology & Philosophy*, *32*(4), 559–580.

Condit, C., & Condit, D. (2001). Blueprints and Recipes: Gendered Metaphors for Genetic Medicine. *Journal of Medical Humanities*, *22*, 29–39.

Condit, C. M., Bates, B. R., Galloway, R., Givens, S. B., Haynie, C. K., Jordan, J. W., Stables, G., & West, H. M. (2002). Recipes or blueprints for our genes? How contexts selectively activate the multiple meanings of metaphors. *Quarterly Journal of Speech*, *88*(3), 303–325.

Cummins, R. (1997). The Lot of the Casual Theory of Mental Content. *Journal of Philosophy*, *94*(10), 535–542.

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.

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

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.

Diaz-Leon, E. (2015). What Is Social Construction? *European Journal of Philosophy*, *23*(4), 1137–1152.

Diesendruck, G., & Menahem, R. (2015). Essentialism promotes children's inter-ethnic bias. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01180

Fausto-Sterling, A. (2005). The Bare Bones of Sex: Part 1—Sex and Gender. *Signs*, *30*(2), 1491–1527.

Ferreira Ruiz, M., & Umerez, J. (2021). Interactionism, Post-interactionism, and Causal Complexity: Lessons From the Philosophy of Causation. *Frontiers in Psychology*, *12*.

Flament-Fultot, M. (2014). On Genic Representations. *Biological Theory*, *9*(2), 149–162.

Fodor, J. A. (1981). The Present Status of the Innateness Controversy. In J. Fodor (Ed.), *RePresentations: Philosophical Essays on the Foundations of Cognitive Science* (pp. 257–316). Cambridge, MA: MIT Press.

Fox Keller, E. (2000). *The Century of the Gene*. Harvard University Press.

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

Gannett, L. (2010). Questions Asked and Unasked: How by Worrying Less About the "Really Real" Philosophers of Science Might Better Contribute to Debates About Genetics and Race. *Synthese*, *177*(3), 363–385.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought* (pp. x, 382). Oxford University Press.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409.

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.

Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7900–7907.

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs: I. Introduction. *Monographs of the Society for Research in Child Development*, *69*(1), 1–14.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38*(3), 213–244.

Godfrey-Smith, P. (1999). Genes and Codes: Lessons from the Philosophy of Mind? In *Biology Meets Psychology. Constraints, Conjectures, Connections* (V. Hardcastle (ed.), pp. 305–331). MIT Press.

Godfrey-Smith, P. (2000). On the Theoretical Role of "Genetic Coding." *Philosophy of Science*, *67*(1), 26–44.

Godfrey-Smith, P. (2004). On Folk Psychology and Mental Representation. In H. Clapin (Ed.), *Representation in Mind* (pp. 147–162). Elsevier.

Godfrey-Smith, P. (2008). Innateness and Genetic Information. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind, Volume 3*. Oxford University Press.

Godfrey-Smith, P., & Sterelny, K. (2016). Biological Information. *The Stanford Encyclopedia of Philosophy (Summer 2016 Edition), Edward N. Zalta (Ed.)*. https://plato.stanford.edu/archives/sum2016/entries/information-biological/

Gould, W. A., & Heine, S. J. (2012). Implicit Essentialism: Genetic Concepts Are Implicitly Associated with Fate Concepts. *PLoS ONE*, *7*(6), e38176.
Griffiths, P. (2017). Communicating genomic complexity. *Journal & Proceedings of the Royal Society of New South Wales*, *150*(1), 93–103.
Griffiths, P. (2021). The Distinction Between Innate and Acquired Characteristics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/innate-acquired/
Griffiths, P. E. (2001). Genetic Information: A Metaphor in Search of a Theory. *Philosophy of Science*, *68*(3), 394–412.
Griffiths, P. E. (2002). What is Innateness? *The Monist*, *85*(1), 70–85.
Griffiths, P. E., & Gray, R. D. (1994). Developmental Systems and Evolutionary Explanation. *Journal of Philosophy*, *91*(6), 277–304.
Griffiths, P. E., & Machery, E. (2008). Innateness, Canalization, and "Biologicizing the Mind." *Philosophical Psychology*, *21*(3), 397–414.
Griffiths, P., Machery, E., & Linquist, S. (2009). The Vernacular Concept of Innateness. *Mind & Language*, *24*(5), 605–630.
Griffiths, P., & Stotz, K. (2013). *Genetics and Philosophy: An Introduction*. Cambridge University Press.
Gross, S., & Rey, G. (2012). Innateness. In *The Oxford Handbook of Philosophy of Cognitive Science*. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195309799.001.0001/oxfordhb-9780195309799-e-14
Grossi, G. (2017). Hardwiring: Innateness in the Age of the Brain. *Biology and Philosophy*, *32*(6), 1047–1082.
Hampton, J. A. (2012). Thinking Intuitively: The Rich (and at Times Illogical) World of Concepts. *Current Directions in Psychological Science*, *21*(6), 398–402.
Haslam, N., Bastian, B., Bain, P., & Kashima, Y. (2006). Psychological Essentialism, Implicit Theories, and Intergroup Relations. *Group Processes & Intergroup Relations*, *9*(1), 63–76.
Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, *39*(1), 113–127.
Haslam, N., & Whelan, J. (2008). Human Natures: Psychological Essentialism in Thinking about Differences between People. *Social and Personality Psychology Compass*, *2*(3), 1297–1312.
Haslanger, S. (2003). Social Construction: The "Debunking" Project. In M. Gilbert & K. Kimble (Eds.), *Socializing Metaphysics* (pp. 301–325). Rowman & Littlefield Publishers.
Haslanger, S. (2005). What Are We Talking About? The Semantics and Politics of Social Kinds. *Hypatia*, *20*(4), 10–26.
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.
Hitchcock, C. (1996). The Role of Contrast in Causal and Explanatory Claims. *Synthese*, *107*(3), 395–419.
Jablonka, E. (2002). Information: Its Interpretation, its Inheritance, and its Sharing. *Philosophy of Science*, *69*(4), 578–605.
Joffe, M. (2013). The Concept of Causation in Biology. *Erkenntnis*, *78*(2), 179–197.

Keil, F. C. (1989). *Concepts, Kinds, and Cognitive Development*. A Bradford Book.
Keller, J. (2005). In genes we trust: The biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology*, *88*(4), 686–702.
Khalidi, M. A. (2016). Innateness as a natural cognitive kind. *Philosophical Psychology*, *29*(3), 319–333.
Khalidi, M. A., & Mugg, J. (2014). The inherent bias in positing an inherence heuristic. *The Behavioral and Brain Sciences*, *37*(5), 493–494.
Kitcher, P. (2001). *Science, Truth, and Democracy*. Oxford University Press.
Knobe, J., & Samuels, R. (2013). Thinking like a scientist: Innateness as a case study. *Cognition*, *126*(1), 72–86.
Kraemer, D. M. (2018). Philosophical analyses of scientific concepts: A critical appraisal. *Philosophy Compass*, *13*(9), e12513.
Levy, A. (2011). Information in Biology: A Fictionalist Account. *Noûs*, *45*, 640–657.
Linquist, S., Machery, E., Griffiths, P. E., & Stotz, K. (2011). Exploring the folkbiological conception of human nature. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1563), 444–453.
Longino, H. (2013). *Studying Human Behavior*. The University of Chicago Press.
Lynch, K. E., & Bourrat, P. (2017). Interpreting Heritability Causally. *Philosophy of Science*, *84*(1), 14–34.
Machery, E. (2021). A new challenge to conceptual engineering. *Inquiry*. https://doi.org/10.1080/0020174X.2021.1967190
Machery, E., Griffiths, P., Linquist, S., & Stotz, K. (2019). Scientists' Concepts of Innateness: Evolution or Attraction? In R. Samuels & D. Wilkenfeld (Eds.), *Advances in Experimental Philosophy* (pp. 172–201). Bloomsbury Academic.
Mallon, R. (2016). *The Construction of Human Kinds*. Oxford University Press.
Mallon, R., & Weinberg, J. M. (2006). Innateness as Closed Process Invariance. *Philosophy of Science*, *73*(3), 323–344.
Mameli, M. (2008). On Innateness: The Clutter Hypothesis and the Cluster Hypothesis. *The Journal of Philosophy*, *105*(12), 719–736.
Mameli, M., & Bateson, P. (2011). An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1563), 436–443.
Mandalaywala, T. M. (2020). Does essentialism lead to racial prejudice? It is not so Black and White. *Advances in Child Development and Behavior*, *59*, 195–245.
Mandalaywala, T. M., Amodio, D. M., & Rhodes, M. (2018). Essentialism Promotes Racial Prejudice by Increasing Endorsement of Social Hierarchies. *Social Psychological and Personality Science*, *9*(4), 461–469.
Margolis, E., & Laurence, S. (2013). In defense of nativism. *Philosophical Studies*, *165*(2), 693–718.
Matteo Mameli, & Bateson, P. (2006). Innateness and the Sciences. *Biology and Philosophy*, *21*, 155–188.
Maynard Smith, J. (2000). The Concept of Information in Biology. *Philosophy of Science*, *67*(2), 177–194.
Mayr, E. (1961). Cause and effect in biology. *Science (New York, N.Y.)*, *134*(3489), 1501–1506.
Medin, D., & Ortony, A. (1989). Psychological essentialism. *Vosniadou and Ortony*, *1989*, 179–195.

Newman, G. E., Herrmann, P., Wynn, K., & Keil, F. C. (2008). Biases towards internal features in infants' reasoning about objects. *Cognition*, *107*(2), 420–432.
Norenzayan, A., & Heine, S. J. (2005). Psychological Universals: What Are They and How Can We Know? *Psychological Bulletin*, *131*(5), 763–784.
Northcott, R. (2008). Causation and Contrast Classes. *Philosophical Studies*, *139*(1), 111–123.
Northcott, R. (2012). Genetic Traits and Causal Explanation. In K. Plaisance & T. Reydon (Eds.), *Philosophy of Behavioral Biology* (pp. 65–82). Springer.
Northcott, R., & Piccinini, G. (2018). Conceived This Way: Innateness Defended. *Philosophers' Imprint*, *18*(18), 1–16.
O'Connor, C., & Joffe, H. (2014). The social aetiology of essentialist beliefs. *The Behavioral and Brain Sciences*, *37*(5), 498–499.
O'Neill, E. (2015). Relativizing Innateness: Innateness as the Insensitivity of the Appearance of a Trait with Respect to Specified Environmental Variation. *Biology and Philosophy*, *30*(2), 211–225.
Oyama, S. (1985). *The Ontogeny of Information: Developmental Systems and Evolution*. Cambridge University Press.
Parrott, R., Silk, K., Weiner, J., Condit, C., Harris, T., & Bernhardt, J. (2004). Deriving Lay Models of Uncertainty About Genes' Role in Illness Causation to Guide Communication About Human Genetics. *Journal of Communication*, *54*(1), 105–122.
Parrott, R., & Smith, R. (2013). Defining Genes Using "Blueprint" Versus "Instruction" Metaphors: Effects for Genetic Determinism, Response Efficacy, and Perceived Control. *Health Communication*, *29*.
Perovic, S., & Radenovic, L. (2011). Fine-Tuning Nativism: The "Nurtured Nature" and Innate Cognitive Structures. *Phenomenology and the Cognitive Sciences*, *10*(3), 399–417.
Pitt, D. (2000). Nativism and the Theory of Content. *ProtoSociology*, *14*, 222–239.
Planer, R. J. (2016). Are Genetic Representations Read in Development? *The British Journal for the Philosophy of Science*, *67*(4), 997–1023.
Prentice, D. A., & Miller, D. T. (2007). Psychological Essentialism of Human Categories. *Current Directions in Psychological Science*, *16*(4), 202–206.
Rakoczy, H., & Cacchione, T. (2019). Comparative metaphysics: Evolutionary and ontogenetic roots of essentialist thought about objects. *WIREs Cognitive Science*, *10*(5), e1497.
Ramsey, W. (n.d.). *Defending representational realism*.
Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531.
Rhodes, M., & Moty, K. (2020). Chapter One—What is social essentialism and how does it develop? In M. Rhodes (Ed.), *Advances in Child Development and Behavior* (Vol. 59, pp. 1–30). Elsevier Academic Press Inc.
Russo, F., & Williamson, J. (2007). Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science*, *21*(2), 157–170.
Samet, J. (2019). The Historical Controversies Surrounding Innateness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019). Metaphysics Research Lab, Stanford University.
Samuels, R. (2004). Innateness in Cognitive Science. *Trends in Cognitive Sciences*, *8*(3), 136–141.
Samuels, R. (2016). The Concept of Innateness as an Object of Empirical Enquiry. In *A Companion to Experimental Philosophy* (pp. 504–519). John Wiley & Sons, Ltd.

Sarkar, S. (2003). Genes Encode Information for Phenotypic Traits. In D. Hull (Ed.), *Contemporary Debates in Philosophy of Science* (pp. 259–272). Blackwell.
Sarkar, S. (2004). How genes Encode Information for Phenotypic Traits. In *Molecular Models of Life: Philosophical Papers on Molecular Biology*. MIT Press.
Schaffer, J. (2005). Contrastive Causation. *Philosophical Review*, *114*(3), 327–358.
Sesardic, N. (2005). *Making Sense of Heritability*. Cambridge University Press.
Sesardic, N. (2010). Race: A Social Destruction of a Biological Concept. *Biology and Philosophy*, *25*(2), 143–162.
Shea, N. (2007). Representation in the genome and in other inheritance systems. *Biology & Philosophy*, *22*(3), 313–331.
Shea, N. (2012). Genetic Representation Explains the Cluster of Innateness-Related Properties. *Mind and Language*, *27*(4), 466–493.
Shea, N. (2013a). Naturalising Representational Content. *Philosophy Compass*, *8*(5), 496–509.
Shea, N. (2013b). Inherited Representations are Read in Development. *The British Journal for the Philosophy of Science*, *64*(1), 1–31.
Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.
Shiao, J. L., Bode, T., Beyer, A., & Selvig, D. (2012). The Genomic Challenge to the Social Construction of Race. *Sociological Theory*, *30*(2), 67–88.
Sousa, P., Atran, S., & Medin, D. (2002). Essentialism and folkbiology: Evidence from Brazil. *Journal of Cognition and Culture*, *2*(3), 195–223.
Stegmann, U. E. (2004). The arbitrariness of the genetic code. *Biology and Philosophy*, *19*(2), 205–222.
Stegmann, U. E. (2009). DNA, Inference, and Information. *The British Journal for the Philosophy of Science*, *60*(1), 1–17. https://doi.org/10.1093/bjps/axn041
Sterelny, K., Smith, K. C., & Dickison, M. (1996). The Extended Replicator. *Biology and Philosophy*, *11*(3), 377–403.
Stotz, K. (2006). Molecular Epigenesis: Distributed Specificity as a Break in the Central Dogma. *History and Philosophy of the Life Sciences*, *28*(4), 533–548.
Stotz, K., Griffiths, P. E., & Knight, R. (2004). How Biologists Conceptualize Genes: An Empirical Study. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *35*(4), 647–673.
Šustar, P. (2007). Crick's notion of genetic information and the 'central dogma' of molecular biology. *The British Journal for the Philosophy of Science*, *58*(1), 13–24.
Tabery, J. (2007). Biometric and developmental gene-environment interactions: Looking back, moving forward. *Development and Psychopathology*, *19*(4), 961–976.
Tabery, J. (2014). *Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture*. MIT Press.
Uhlmann, E. L., Zhu, L., Brescoll, V. L., & Newman, G. E. (2014). System-Justifying Motives Can Lead to Both the Acceptance and the Rejection of Innate Explanations for Group Differences. *Behavioral and Brain Sciences*, *37*(5), 503–504.
Waters, C. K. (2004). What concept analysis in philosophy of science should be (and why competing philosophical analyses of gene concepts cannot be tested by polling scientists). *History and Philosophy of the Life Sciences*, *26*(1), 29–58.
Waters, C. K. (2007). Causes That Make a Difference. *The Journal of Philosophy*, *104*(11), 551–579.
Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, *43*(2), 294–308.

Weber, M. (2017). *Causal Selection vs Causal Parity in Biology: Relevant Counterfactuals and Biologically Normal Interventions* [Preprint].
Wheeler, M. (2007). Traits, Genes, and Coding. In M. Ruse (Ed.), *Philosophy of Biology* (pp. 369–401). Prometheus Books.
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.
Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318.

EESTIKEELNE KOKKUVÕTE

Nativistlike väidete sisu ja järelmid. Filosoofiline analüüs

Väitekiri koosneb kolmest eelretsenseeritud artiklist, artiklikäsikirjast ja sissejuhatavast ülevaateartiklist. Kõiki artikleid ühendab eesmärk selgitada ja hinnata erinevates empiirilistes teadustes tehtud nativistlike väidete sisu ja nende väidete järelmeid teatud filosoofilistele teooriatele. Nativistlikeks nimetan väiteid, mis väidavad mingil organismi tunnusel olevat olulised seesmiseid, sünnipäraseid ja bioloogiliselt päritavad põhjused. Sellisteks väideteks on näiteks väide, et tunnus on kaasasündinud; et tunnus on geneetiliselt põhjustatud; et tunnus on DNA-s kodeeritud; et tunnus on päritav jmt. Nativistlikke väiteid kohtab erievates teaduskontekstides: psühholoogias, geneetika erinevates harudes, arengubioloogias, kognitiivteadustes, käitumisökoloogias, evolutsiooniteadustes jm. Nende mõju on aga laiem. Nativistlikud väited väljendavad hüpoteese, teadmisi, mis on sisendiks ja tõendmaterjaliks teistes teoreetilistes ja praktilistes valdkondades. Nad kujundavad üldisemalt meie arusaama inimese ja teiste elusolendite olemusest, nende suhtest oma keskkonnaga. Seetõttu on oluline, et teaduslikke nativistlikke väiteid tõlgendataks õigesti.

Ometi on oht, et teaduslikke nativistlikke väiteid tõlgendatakse süstemaatiliselt vääriti. Et väärtõlgendusi ennetada ja korrigeerida, on oluline selliste väidete täpsemat sisu ja järelmeid uurida ja selgitada. Seda oma väitekirjas teengi. Igas väitekirja artiklis võtan ette kindlat tüüpi nativistliku väite ning avan ja hindan seda tüüpi väite sisu. Kahel puhul selgitan, mil moel piirab, või ei piira, seda tüüpi nativistlik väide vastust konkreetsele filosoofilisele küsimusele. Artiklis "Innate Mind Need Not be Within" (artikkel I) selgitan, mis järeldub empiirilisest hüpoteesist, et mingi vaimne mõiste on kaasasündinud, filosoofias laialt pooldatud põhjuslikele vaimse sisu teooriatele. Artiklis "Elusive Vehicles of Genetic Representation" (artikkel I) kritiseerin teatud tüüpi teleosemantilisi geneetilise informatsiooni teooriaid kui teleosemantilise sisuteooria rakendamist kohas, kus see tegelikult ei rakendu. Artiklites "Socially Constructed and/or Genetically Caused" (artikkel III) ja "Causal Social Construction" (artikkel IV) selgitan, mis tingimustel on teaduslik väide, et tunnusel on geneetilised põhjused, tõend sellest, et see tunnus ei ole sotsiaalselt konstrueeritud. Väitekirja artikleid saadab ülevaateartikkel. See koosneb kahest suuremast osast. Esimeses osas motiveerin oma ettevõtmise üldist tausta. Teises osas selgitan eraldi iga konkreetse artikli sisulisi ja metodoloogilisi eeldusi ning kohta olemasolevates temaatilistes aruteludes. Iga üksik artikkel seob filosoofilisi teemasid ja arutelusid, mida traditsiooniliselt on käsitletud teineteisest eraldi. Artiklite tausta avades keskendun artikli konteksti sellele aspektile, mis puudutab artikli fookuses olevat nativistlikku mõistet.

Iga artikli keskmes oleva mõiste ja vastavate nativistlike väidete sisu ja sisukuse üle on teadus-filosoofilises kirjanduses palju arutatud. Väitekiri jätkab neid arutelusid ent erineb neist vähemalt kahe aspekti poolest. Esiteks, see mitte ei keskendu ühele neist mõistetest, nagu on tüüpiline, vaid käsitleb korraga erinevaid

nativistlikke mõisteid võrdlevalt ühtses raamis. Teiseks, iga väitekirja artikkel kui ka väitekiri tervikuna täidab eeskätt silla-ehituse ja teooriate ja uurimisprogrammide vahelise tõlkimise eesmärki. Nii näiteks ei esita ma algupärast analüüsi sellele, mida tähendab olla kaasasündinud või omada geneetilisi põhjuseid, või mida tähendab kanda millegi kohta informatsiooni. Ma kasutan olemasolevaid teadusfilosoofilisi käsitusi, neid vajadusepõhiselt organiseerides, uudsel viisil küsimuste lahendamisel, mis tüüpiliselt jäänud väljapoole sedalaadi teooriate huviorbiiti.

Järgnevas annan detailsema ülevaate väitekirja sisust selle osade kaupa.

Ülevaateartikli esimeses osas motiveerin oma ettevõtmist empiirilise väite taustal, et seda, kuidas tavainimesed tõlgendavad nativistlikke väiteid, seletab muu hulgas psühholoogiline essentsialism. Psühholoogiliseks essentsialismiks nimetatakse inimesele universaalset psühholoogilist kalduvust näha bioloogilisi olendeid otsekui omaksid nad seesmist „olemust" või „essentsi" – mittevaadeldavat tuuma, mis on liigile tüüpiline, mis püsib samana isegi kui organismi välised tunnused muutuvad, mis on päritud bioloogilistelt vanematelt ja mis põhjustab teisi selliseid organismi omadusi, mida see organism teiste omasugustega jagab. Mitmed autorid on väitnud, et psühholoogilise essentsialism avaldub ka selles, kuidas tõlgendatakse nativistlikke väiteid – väidet, et tunnus on näiteks geenide põhjustatud või kaasasündinud, kiputakse tõlgendatakse kui väidet, et tunnus pärineb organismi „olemusest" ning et seega on sel tüüpilisele olemuslikule tunnusele omased omadused. Selline teaduslike nativistlike väidete tõlgendamine on üldjuhul väär. Asjaolust, et nativistlikke väiteid kaldutakse tõlgendama psühholoogilise essentsialismi vaimus, tuleneb kaks ohtu. Esiteks, kuivõrd psühholoogiline essentsialism on kognitiivne kalduvus, mis iseloomustab ühtviisi nii teadlast kui tavainimest, on oht, et teaduslikkuse näilises kuues hiilib rahvalik essentsialistlik mõtteskeem sisse teadusteooriatesse ja genereerib nii viletsat teadust. Teiseks, on oht, et teaduslikke nativistlikke väiteid *tõlgendatakse* vääralt psühholoogilise essentsialismi vaimus ning järeldatakse neist midagi, mis neist ei järeldu. Väitekirja artiklid püüavad neid ohte ennetada ja võimalikke tõlgendusvigu korrigeerida.

Väitekirja artiklid lähtuvad eeldusest (seda ühtlasi illustreerides ja kinnitades) et eri tüüpi ja eri teaduskontekstides tehtud nativistlike väidete järelmid võivad oluliselt erineda. Seetõttu tuleb neid hinnata juhtumipõhiselt, pöörates tähelepanu nimelt kontekstuaalsetele erinevustele ja eeldamata, et üht tüüpi nativistliku väite järelmid üldistuvad teist tüüpi nativistliku väite järelmiteks. Väitekirja artiklid on igaüks taoline juhtumiuuring.

Artiklis **„Innate Mind Need Not Be Within"** selgitan ja hindan kognitiivteadustes ja sellega piirnevais vaimufilosoofia harudes levinud väite, et mõned mõisted on kaasasündinud, järelmeid põhjuslikule eksternalismile. Põhjuslik eksternalism on vaimufilosoofias populaarne vaade, mille järgi mingi vaimse mõiste sisu ei määra mitte seda mõistet omava organismi seesmised omadused, vaid see, milline nähtus välises maailmas kas selle mõiste omamist või esinemist põhjustab. Kuivõrd meil on tugev intuitsioon, et kaasasündinud tunnused on

määratud organismi *seesmiste* omadustega, võib esmapilgul tunduda, et kaasasündinud mõiste sisu ei saa olla sel viisil määratud. Seega, kui eksternalism on mõeldud olema universaalne sisuteooria, siis on mõistenativism ja mõisteeksternalism vastuolus. Artiklis näitan, et sellist vastuolu ei ole.

Ülevaateartikli sektsioonis 3.1. avan „Innate Mind Need Not Be Within" tausta ja metodoloogilisi eeldusi. Rõhutan, et kui eesmärk on hinnata teatud tüüpi nativistliku väite – näiteks väite, et tunnus on kaasasündinud – järelmeid mõne muu valdkonna jaoks, siis mõiste, mis analüüsimist vajab, on kaassasündivuse *operatiivne* mõiste: mõiste, mis kindlas teaduskontekstis, panustab *selle* propositsiooni sisusse, mis huvialuses kontekstis ka tõepoolest on tõendatud. Rõhutan, et tõenäoliselt on kognitiivteadustes selliseid operatiivseid kaasasündivuse mõisteid mitmeid ning nii võivad ka väite, et mõiste on kaasasündinud, järelmid varieeruda.

Sellelt eelduselt lähtub „Innate Mind Need Not Be Within". Artiklis eeldan, et *kui* kognitiivteadustes leidub sisukas ja legitiimne kaasasündivuse mõiste, siis vähemalt üks järgmistest selle mõiste prominentsetest filosoofilistest analüüsidest on adekvaatne: mõiste on kaasasündinud siis, (a) kui see pole omandatud psühholoogilisi teid pidi, (b) kui see on omandatud tingimustes, kus keskkonnas leiduv informatsioon ei ole piisav selle mõiste omandamiseks, (c) kui selle mõiste omandamist ei ole põhjustanud kogemus. Seejärel eristan kolme tüüpi põhjuslikku sisuteooriat: informatsioonilist sisuteooriat, omandamisteooriat ja põhjusliku ajaloo teooriat. Analüüsin kõikide nende teooriate kombinatsioone eraldi. Iga kombinatsiooni puhul selgitan, mis asjaoludest sõltub, kas konkreetne mõiste saab olla korraga kaasasündinud ja selle sisu määratud sellest, mis seda mõistet põhjustab. Näitan, et ühegi kombinatsiooni puhul pole põhjuslike sisuteooriate ja mõistenativismi vahel vastuolu.

Artiklis **„Elusive Vehicles of Genetic Representation"** tegelen küsimusega, kas geenid kannavad organismi fenotüübiliste tunnuste kohta nn semantilist informatsiooni. Bioloogiateadustes on levinud praktika kirjeldada genoomi just sel viisil: organismil arenevat teatud fenotüüp seetõttu, et tema geenides peitub informatsioon, et just selline tunnus peab arenema. Paljud filosoofid arvavad, et selline kõnepruuk on kas metafooriline või väär. Teised aga kaitsevad väidet, et geenid tõepoolest kannavad teatud tunnuste kohta informatsiooni. Teleosemantilised geneetilise informatsiooni teooriad on üks prominentne viis sellist väidet kaitsta. Teleosemantilise teooria järgi kannavad geenid informatsiooni nende tunnuste kohta, mida nad on looduslikult valitud põhjustama. Artiklis võtan üksipulgi lahti need tingimused, mida geenid teleosemantilise sisuteooria järgi peaksid rahuldama, representeerimaks organismi kompleksseid tunnuseid. Väidan, et selle valguses, mida teame geenide rolli kohta evolutsiooniprotsessis ja viisist, kuidas nad organismi tunnuseid põhjustavad, ei rahulda geenid neid tingimusi, ja seda ükskõik millist levinud geeni mõistetest me eeldame.

Ülevaateartikli sektsioonis 3.2.1 selgitan, mida artiklis „Elusive Vehicles of Genetic Representation" adresseeritud küsimus „Kas geenid kannavad fenotüübi kohta semantilist informatsiooni?" üleüldse küsib, kuidas sellele küsimusele vastata ja millest vastus sellele küsimusele sõltub. Eristan kaht tarvilikku tingimust, mille korral on tõesti nii, et geenid kannavad fenotüübi kohta informatsiooni.

Esiteks, geenide suhe fenotüüpi peab vastama metafüüsilistele informatsiooni kandmise tingimustele. Teiseks, oluline hulk sellest, mida bioloogid geneetilise informatsiooni kohta ütlevad, peab olema tõene. Oma teleosemantilise teooria kriitikas apelleerin esimesele tingimusele. Seetõttu avan mõneti pikemalt seda, mida üldse tähendab olla semantilise informatsiooni kandja. See on vaidlusalune küsimus. On olemas erinevad semantilise informatsiooni kandmise paradigmad ja neist lähtuvalt erinevaid arusaamu sellest, mis on sellise informatsiooni kandmiseks konstitutiivselt tarvilik. Eristan kaht tüüpi filosoofilisi teooriaid, mille kohaselt geenid on semantilise informatsiooni kandjad. Need teooriad erinevad muu hulgas selle poolest, millist semantilistele nähtustele omaseks peetud tunnust nad peavad keskseks: sümbolilisust või tõeväärtuslikkust. Rõhutan, et kuigi need eri tüüpi teooriad kaitsevad mõlemad teesi, et geenidel on semantilised omadused, kaitsevad nad ometigi oluliselt erinevaid teese. Nad omistavad geenidele *erinevaid* „semantilisi" omadusi, millel omakorda oluliselt erinev seletuslik potentsiaal.

Artikleid **„Causal Social Construction"** ja käsikirja **„Socially Constructed and/or Genetically Caused"** tuleb vaadelda ühe tervikuna. Sotsiaalteadustes ja sotsiaalfilosoofias sageli väidetakse, et üks või teine inimkategooria või -tunnus on sotsiaalselt konstrueeritud. Ning tüüpiline on eeldada, et *kui* on näidatud, et tunnus on sotsiaalselt konstrueeritud, *siis* on ühtlasi näidatud et mingis olulises ja täpsustatud tähenduses ei ole see tunnus bioloogiline, kusjuures „bioloogiline" all peetakse aina enam silmas seda, et tunnusel on geneetilised põhjused. Kooskõlas selle eeldusega on üha sagedasem, et mõne sotsiaalkonstruktivistliku teooria õigsust või väärust hinnatakse teaduslike leidude valguses huvialuse tunnuse geneetiliste põhjuste kohta. „Socially Constructed and/or Genetically Caused" ja „Causally Socially Constructed" püüavad üheskoos seda praktikat mõisteliselt organiseerida. Selgitan filosoofilise põhjuslikkuseteooria abil, mis tingimustel on või ei ole empiiriline väide, et tunnusel on geneetilised põhjused, vastuolus väitega, et tunnus on kas konstitutiivselt või põhjuslikult sotsiaalselt konstrueeritud. Artikli teises osas rakendan näite korras seda üldist raami pärilikkusuuringute järelmite hindamiseks.

Artiklis „**Causal Social Construction"** pakun üldise definitsiooni sellele, mida tähendab olla põhjuslikult sotsiaalselt konstrueeritud. Rakendan selleks kontrastiivset kontrafaktuaalse põhjuslikkuse teooriat. Aruteludes millegi sotsiaalse konstrueerituse üle on tavaks eristada kahte viisi, kuidas nähtus saab olla sotsiaalselt konstrueeritud – nii, et sotsiaalsed tegurid seda nähtust kas põhjustavad või konstitueerivad. Ometigi ei olnud neis aruteludes seni selgesõnaliselt määratletud, mil viisil ja mis tähenduses peavad sotsiaalsed tegurid nähtust põhjustama, et see nähtus oleks sotsiaalselt põhjuslikult konstrueeritud. Sest mitte iga sotsiaalsete põhjustega nähtus ei saa olla sotsiaalselt konstrueeritud. Artiklis „Causal Social Construction" püüan seda auku täita. Väidan, et nähtus X on sotsiaalselt konstrueeritud järgmisel tingimusel: kui teatud tegelike sotsiaalsete faktorite asemel oleksid olnud alternatiivsed sotsiaalsed faktorid SF*, siis X-i ei oleks, ja SF* on kas sellised alternatiivsed sotsiaalsed faktorid, mis on manifesteerunud mõnes tegelikus (mineviku või oleviku) ühiskonnas või mida on praktiliselt võimalik esile kutsuda. See määratlus teenib sotsiaalkonstruktivistlike

teooriate eesmärke ja on kooskõlas sotsiaalkonstruktivistliku diskursusega. Seda määratlust eeldab „Socially Constructed and/or Genetically Caused“.

Sektsioonis 3.2.2 selgitan nende kahe artikli tausta. Palju teadusfilosoofilisest geenide kui tunnuste põhjuste ümber toimuvast arutelust on keskendunud selle hindamisele, kas see, kuidas geenid tunnuseid põhjustavad, erineb mingil olulisel moel sellest, kuidas põhjustavad tunnuseid mitte-geneetilised ehk keskkondlikud tegurid (eeldades triviaalsed tõsiasja, et kõikidel tunnustel on nii geneetilised kui keskkondlikud põhjused). Ent olulisel kohal on ka fundamentaalsem küsimus: mida üldse tähendab, et geenid mingit tunnust põhjustavad? Artiklis „Socially Constructed and/or Genetically Caused“ väidan, et seda geneetilise põhjustanusse operatiivset mõistet empiirilistes teadustes on kasulik mudeldada kontrastiivse kontrafaktuaalse põhjuslikkuseteooria abil. See võimaldab selgesti näha, millist laadi tunnuse sõltuvust keskkondlikest teguritest teadusfakt, et sel tunnusel on geneetilised põhjused, välistab ja millist mitte. See aspekt omakorda on oluline hindamaks, kas ja kuivõrd on konkreetne empiiriline väide, et tunnusel on geneetilised põhjused, tõend sellest, et see tunnus ei ole sotsiaalselt konstrueeritud.

Ülevaateartikli **sektsioonis 4** vastan võimalikule kõhklusele ühe oma väitekirja motiveeriva eelduse suhtes. Mõned teadusfilosoofid väidavad – mis on minugi projekti eeldus – et kuna sõna ’kaasasündunud’ (*innate*) on tavakeeles seotud ähmaste ja väärate essentsialistlike kontseptsioonidega, ärgitab selle sõna kasutamine teaduslikes kontekstides teaduslike hüpoteeside väärtõlgendusi. Seetõttu oleks kõige parem loobuda selle sõna kasutamisest teadushüpoteeside sõnastamisel. Sama on väidetud ka teist tüüpi nativistlikes väidetes esinevate sõnade, näiteks geenidega seotud väljendite kohta. Kõnealust nn eliminativistlikku agendat võib näha alternatiivina siinse väitekirja ühele juhtmõttele – et tulenevalt ohust, et teaduslikke nativistlikke väiteid kiputakse krooniliselt ekslikult tõlgendama, tuleb erinevate nativistlike väidete ja neis esinevate sõnade sisu lahata, kontrollida ja selgitada. Spekuleerin, et eliminativistlik strateegia ei oleks kuigi tõhus alternatiiv väitekirjas ette võetud selgitustööle.

PUBLICATIONS

CURRICULUM VITAE

Name: Riin Kõiv
E-mail: riin.koiv@ut.ee

I. Education

2021. PhD candidate in philosophy. University of Tartu. Thesis topic: *"*The Content and Implications of Nativist Claims. A Philosophical Analysis". Supervisors: Daniel Cohnitz, Bruno Mölder.
2017–2018. Visitor at the Macquarie University, Sidney. Supervisor: Karola Stotz.
2016. Visitor at the Universidad de los Andes (Bogota, Colombia).
2015. Researcher at *Arbeitestelle Feuerbach* at the University of Münster (Wilhelms-Universität Münster).
2012–2013. Visitor at the University of California, San Diego. Mentor: Rick Grush.
2011. University of Tartu, MA *cum laude* in philosophy. Thesis title: "Naturalism kui idealism eitus. Ludwig Feuerbachi „tuleviku filosoofia programm"". Supervisors: Eduard Parhomenko, Jüri Lipping.
2009–2010. Philipps-Universität Marburg
2008. University of Tartu, BA in philosophy. Thesis title: „Eksperimendist kui filosofeerimise meetodist Nietzschel". Supervisor: Eduard Parhomenko. Minors: art history, classical studies.

II. Employment

2019–2021 Assistant with the project "Philosophical analysis of interdisciplinary research practices" (PRG462). Principal investigator: Endla Lõhkivi.
2016–2021. Lecturer at the Estonian Academy of Arts.

III. Teaching and supervision experience

2019–2021. MA thesis co-supervisor (title: "Is 'fitness' a primitive or a propensity? Diagnosing the role of explanatory reductionism on differing standards of scientific definitions"). University of Tartu
2021. Does being biological exclude being socially constructed? MA and PhD level. University of Tartu.
2020. Problems of the 20th and 21st Century Philosophy. E-course. MA level. Estonian Academy of Arts.
2016–2020. Problems of the 20th and 21st Century Philosophy. MA level. Estonian Academy of Arts.
2019–2020. 20. ja 21. sajandi filosoofia probleeme. MA level. Estonian Academy of Arts.
2017. Philosophical Perspectives on the Naturalistic Explanation of Religious Belief. With Uku Tooming. MA course, University of Tartu.
2016. Introduction to the Philosophy of Mind (in Estonian). MOOC e-course (with Uku Tooming and Vivian Puusepp).

IV. Research

Research interests

Philosophy of science, philosophy of biology, social philosophy (social construction), philosophy of mind, nature-nurture debate

Publications

Peer reviewed publications

2021. "Innate Mind Need Not Be Within". Acta Analytica. 36, 101–121.

2020. "Elusive vehicles of genetic representations". Biology and Philosophy, 35 (24), https://doi.org/10.1007/s10539-020-9741-8

2019. "Causal Social Construction". Journal of Social Ontology, 5 (1), 77−99.

Other academic publications

2020. "Mis, kes, kuidas ja millal on sotsiaalselt konstrueeritud?" (What, Who, How and When is socially constructed?"). Akadeemia, 1, 54−76.

2016. "Ludwig Feuerbach. Saatesõna" (afterword to the Estonian translation of Ludwig Feuerbach's selected writings). Avatud Eesti Raamat, Ilmamaa, 285−325.

2012. "Ludwig Feuerbach ja uus praktiline filosoofia" (Ludwig Feuerbach and the new practical philosophy). Akadeemia, 276(3), 407–423.

Philosophy translations

2016. Feuerbach, Ludwig. "Tuleviku filosoofia alused" (annotated translation of selected writings by Ludwig Feuerbach from German to Estonian). Ilmamaa: Tartu.

2017. Saul, Jennifer. "Koeraviled, poliitiline manipulatsioon ja keelefilosoofia" (original: *Dogwhistles, Political, Manipulation and Philosophy of Language.*) Akadeemia. 11, 1955–1966.

2016. Gabriel, Gottfried. "Gustav Teichmüller ja mõisteajaloo süstemaatiline tähendus" (original: Gustav Teichmüller and the Systematic Significance of Studying the History of Concepts). In *Tagasi mõteldes: Töid filosoofia ajaloost Eestis*. Ü. Matjus, J. Sooväli (Eds.). Tartu: TÜ Kirjastus, 135–145.

Conference presentations

2021. Omada geneetilisi põhjusi on interaktiivne omadus. XVI Estonian Annula Philosophy Conference. 23.–24. September, University of Tartu.

2021. Caused by genes is an interactive property. Reactivity and Categorization in the Human Sciences. A NOS-HS Nordic exploratory workshop series. May 27–28. Copenhagen University.

2019. Genetic Causes and Social Construction. Science, Politics, and Philosophy. Irish Philosophical Society Annual Conference. 6–7 December, University College Cork, Cork (Ireland)

2019. Philosophical (De)construction of Genetic Causes. Gender and Philosophy. 7th Annual Conference and General Meeting of the Society for Women in Philosophy-Ireland, 26–28 April, Dublin City University, Ireland.

2018. Sellest, kuidas filosoofia loomust vormib. XIV Estonian Annual Philosophy Conference, 5–6 October. Tallinn University, Tallinn.

2017. Why Human Traits Can Be Both Biological (Genetic) and Socially Constitutively Constructed. 9th European Congress of Analytic Philosophy. Munich, 21–26 August.

2016. In Search of the Vehicles of Genetic Representation. PBCS6, the Research Workshop on Philosophy of Biology and Cognitive Science. Barcelona, University of Barcelona, 28–29 April.

2016. On How Biological Need Not Be Biological. The annual meeting of The Nordic Network for Philosophy of Science (NNPS). Pärnu, April 21–23.

2016. On Why Innate Mind Need Not Be Within. PHILOGICA IV, IV Colombian Conference on Logic, Epistemology, and Philosophy of Science. 17–19 February, Bogota, Universidad de los Andes.

2015. The Idleness of Truth Relative to Shared Milieu (Or How not to Make Sense of Ideology Critique). Society for Women in Philosophy – Ireland. 4th Annual Conference and Annual General Meeting Ways of Knowing: Feminist Philosophy of Science and Epistemology. Dublin, 27–28 November.

2015. On Why Innate Is Not Within. SOPhiA, Salzburg Conference for Young Analytic Philosophy. 2–5 September.

2014. Scrutinizing and deflating the role of ‘shared milieu’ in Sally Haslanger’s account of ‘genuine critique’. X Estonian Annual Philosophy Conference. Tartu, September.

2014. Individuating the Vehicles of Genetic Representation. X Estonian Annual Philosophy Conference. Tartu, September.

2014. Shaping humans by conceptualizing Nature. Framing Nature. The European Association for the Study of Literature, Culture, and the Environment (EASLCE) biennial conference. Nordic Network for Interdisciplinary Environmental Studies IX conference. 29 April – 4 May.

2013. Mis see on, kui ‘geenid põhjustavad’...? IX Estonian Annual Philosophy Conference. Tallinn, October.

2012. Sisunativism versus (?) sisueksternalism. VIII Estonian Annual Philosophy Conference. Tallinn, 26–27 October.

V. Science popularization

Public lectures and interviews.

2019. „Kes on sotsiaalselt konstrueeritud?“ (Who is socially constructed?). *Vaba Akadeemia* lecture series.

2018. „Mis on loomus/loomulik?“ (What is nature/natural?). *Vaba Akadeemia* lecture series.

2017. An interview on the translation of Ludwig Feuwebach's selected writings. Estonian national radio.

Articles (selection)

2018. "Sugu: mis ja millest me räägime?"(*Sex/Gende: Why and What are We Talking About?*). (opinion piece in *Eesti Päevaleht,* Estonian daily newspaper)

2016. "Colombiast läbi filosoofi silmade" (*Colombia through the eyes of a philosopher*). An interview with Catalina González. *Sirp*.

2016. "Meediakangelane Zika viirus hoiab varju". Opinion story about the Zika-virus epidemics. Postimees, Estonian daily newspaper.

2015. "St. Ghislaini porised pilved" (*The dirty clouds of St. Ghislain*). Review of Ivar Veermäe's art exhibition "St-Ghislaini pilved". *Vikerkaar*.

2015. "Teadusfilosoofia raskekaaluline klassik kerges kuues. Carl Gustav Hempel. Loodusteaduse filosoofia.". Book review of the Estonian translation of C.G.Hempel's "Philosophy of Natural Science. *Sirp*.

2012. "Dekoloniseeri!. Review of Tanel Rander's art exhibition "Decolonize this". *Sirp*.

VI. Professional service and administration

2020–present. Editor of the electronic repository of the department of philosophy, University of Tartu.

2019. Member of the organization committee of "From ϕ -science to practical realism: an international conference in honour of Rein Vihalemm (1938–2015)." Aug 13–14, 2019, University of Tartu, Tartu, Estonia.

2019. Editorial work on "A Story of a Science: On the Evolution of Chemistry", Rein Vihalemm. *Acta Baltica Historiae et Philosophiae Scientiarum, Vol. 7, No. 2.*

2018. Member of the organization committee of *Applications in Cultural Evolution: Arts, Languages, Technologies.* June 6–8, University of Tartu, Tartu.

2015. Member of the organization committee for *23rd Annual Meeting of the European Society for Philosophy and Psychology*. University of Tartu, Tartu, July 14–17

2015–2016. Student representative in the philosophy department liaison committee.

Peer-reviewed for *Studia Philosophica Estonica* and *Acta Baltica Historiae et Philosophiae Scientiarum.*

ELULOOKIRJELDUS

Nimi: Riin Kõiv
E-post: riin.koiv@ut.ee

I. Haridus

2012–2021. Tartu Ülikool, doktoriõpingud filosoofias. Doktoritöö pealkiri: "The Content and Implications of Nativist Claims. A Philosophical Analysis". Juhendajad Daniel Cohnitz ja Bruno Mölder.

2017–2018. Külalisdoktorant Macquarie ülikoolis Sydneys. Juhendaja Karola Stotz.

2016. Külalisdoktorant Los Andes ülikoolis (Universidad de los Andes) Colombias

2015. Külalisuurija Münsteri Ülikoolis (Wilhelms-Universität Münster allüksuse Arbeitestelle Feuerbach juures).

2012–2013. Külalisdoktorant University of California San Diego filosoofia osakonnas. Mentor Rick Grush.

2011. Tartu Ülikool, magistrikraad *cum laude* filosoofia erialal. Magistritöö Pealkiri: "Naturalism kui idealismi eitus – Ludwig Feuerbachi "tuleviku filosoofia programm"". Juhendajad Eduard Parhomenko ja Jüri Lipping.

2009–2010. Marburgi Ülikool (filosoofia ja klassikaline filoloogia, ERASMUS ja DoRa magistrantide õpirände projekti raames).

2008. Tartu Ülikool, bakalaureusekraad filosoofia erialal. Bakalaureusetöö teema: "Eksperimendist kui filosofeerimise meetodist Nietzschel". Juhendaja Eduard Parhomenko. Kõrvaleriala: kunstiajalugu, klassikaline filoloogia.

II. Akadeemiline töökogemus

2019–2021. Projekti "Philosophical analysis of interdisciplinary research practices" (PRG462) kaastöötaja

2016–2021. Filosoofia lektor Eesti Kunstiakadeemias

III. Õpetamiskogemus

2019–2021. Anna Elise Rohtmetsa magistritöö "Is 'fitness' a primitive or a propensity? Diagnosing the role of explanatory reductionism on differing standards of scientific definitions" kaasjuhendaja. Tartu Ülikool.

2021. "Does being biological exclude being socially constructed?". Magistri- ja doktoriõpe. Tartu Ülikool.

2020. "Problems of the 20th and 21st Century Philosophy". E-kursus rahvusvahelistele magistrantidele. Eesti Kunstiakadeemia..

2019–2021. "20. ja 21. sajandi filosoofia probleeme". Magistriõpe. Eesti Kunstiakadeemia.

2016–2020. “Problems of the 20th and 21st Century Philosophy”. Magistriõpe. Eesti Kunstiakadeemia.
2017. “Philosophical Perspectives on the Naturalistic Explanation of Religious Belief” (koos Uku Toominguga). Magistriõpe. Tartu Ülikool.
2016. kaasosaline MOOC-kursuse “Põnev vaimufilosoofia” sisu loomisel. Tartu Ülikool.
juuni 2017: magistritöö retsensent
2012. “Filosoofiline kirjutamine”. Õppeassistent.

IV. Teadus-, arendus- ja loometegevus

Peamised uurimisvaldkonnad

Teadus- ja eeskätt bioloogiafilosoofia: geneetilise põhjuslikkuse, geneetilise informatsioon, kaasasündivuse mõiste, *nature-nurture* debatt
Vaimufilosoofia: vaimse sisu teooriad, eksternalistlikud sisuteeoriad, kaasasündinud mõisted.
Sotsiaalfilosoofia: sotsiaalkonstruktivism, ideoloogiakriitika

Eelretsenseeritud publikatsioonid

2021. “Innate Mind Need Not Be Within”. Acta Analytica. 36, 101–121.
2020. “Elusive vehicles of genetic representations”. Biology and Philosophy, 35 (24)
2019. “Causal Social Construction”. Journal of Social Ontology, 5 (1), 77−99.

Muud filosoofiaalased publikatsioonid

2020. “Mis, kes, kuidas ja millal on sotsiaalselt konstrueeritud”. Akadeemia, 1, 54−76.
2016. “Ludwig Feuerbach. Saatesõna”. Avatud Eesti Raamat, Ilmamaa, 285−325.
2012. “Ludwig Feuerbach ja uus praktiline filosoofia”. Akadeemia, 276(3), 407–423.

Filosoofiatõlked

2016. Feuerbach, Ludwig. “Tuleviku filosoofia alused” (Ludwig Feuerback valitud teoste kommenteeritud tõlge saksa keelest). Ilmamaa: Tartu.
2017. Saul, Jennifer. “Koeraviled, poliitiline manipulatsioon ja keelefilosoofia“. Akadeemia. 11, 1955–1966.
2016. Gabriel, Gottfried. “Gustav Teichmüller ja mõisteajaloo süstemaatiline tähendus”. In *Tagasi mõteldes: Töid filosoofia ajaloost Eestis*. Ü. Matjus, J. Sooväli (Toim.). Tartu: TÜ Kirjastus, 135–145.
2016. Feuerbach, Ludwig. “Tuleviku filosoofia alused” (Ludwig Feuerback valitud teoste kommenteeritud tõlge saksa keelest). Ilmamaa: Tartu.

Konverentsiettekanded

2021. *Omada geneetilisi põhjuseid on interaktiivne omadus*. Eesti Filosoofia Aastakonverents XVI. 23.–24. September, Tartu.

2021. *Caused by genes as an interactive property.* Reactivity and Categorization in the Human Sciences. A NOS-HS Nordic exploratory workshop series. 26.–28. Mai, Kopenhaagen.

2019. *Genetic Causes and Social Construction*. Science, Politics, and Philosophy. Irish Philosophical Society Annual Conference. 6.–7. detsember. University College Cork, Cork.

2019. *Philosophical (De)construction of Genetic Causes.* Gender and Philosophy. 7th Annual Conference and General Meeting of the Society for Women in Philosophy-Ireland, 26.–28. aprill, Dublin.

2018. *Sellest, kuidas filosoofia loomust vormib*. Eesti Filosoofia Aastakonverents. XIV Estonian Annual Philosophy Conference, 5.–6. oktoober, Tallinn.

2017. *Why Human Traits Can Be Both Biological (Genetic) and Socially Constitutively Constructed*. 9th European Congress of Analytic Philosophy. München, 21.–27. august.

2016. *In Search of the Vehicles of Genetic Representation*. PBCS6, the Research Workshop on Philosophy of Biology and Cognitive Science. 28.–29. aprill, Barcelona.

2016. *On How Biological Need Not Be Biological*. The annual meeting of The Nordic Network for Philosophy of Science (NNPS). Pärnu, 21.–23. Aprill.

2016. *On Why Innate Mind Need Not Be Within*. PHILOGICA IV, IV Colombian Conference on Logic, Epistemology, and Philosophy of Science. 17.–19. veebruar, Bogota.

2015. *The Idleness of Truth Relative to Shared Milieu (Or How not to Make Sense of Ideology Critique)*. Society for Women in Philosophy – Ireland. 4th Annual Conference and Annual General Meeting Ways of Knowing: Feminist Philosophy of Science and Epistemology. Dublin, 27.–28. november.

2015. *On Why Innate Is Not Within*. SOPhiA, Salzburg Conference for Young Analytic Philosophy 2015. 2.–5. September, Salzburg.

2014. *Scrutinizing and deflating the role of 'shared milieu' in Sally Haslanger's account of 'genuine critique*'. Eesti Filosoofia Aastakonverents X. Tartu, september.

2014. *Individuating the Vehicles of Genetic Representation*. Eesti Filosoofia Aastakonverents X. Tartu, september.

2014. *Shaping humans by conceptualizing Nature*. Framing Nature. The European Association for the Study of Literature, Culture, and the Environment (EASLCE) biennial conference. Nordic Network for Interdisciplinary Environmental Studies (NIES) IX conference. 29. aprill – 4. mai.

Mis see on, kui 'geenid põhjustavad'...? Eesti Filosoofia Aastakonverents IX. Tallinn, oktoober.
Sisunativism versus (?) sisueksternalism. Eesti Filosoofia Aastakonverents VIII Tallinn, 26.– 27. oktoober.

Stipendiumid ja uurimistoetused

2017–2018. Kristjan Jaagu välisõpingute stipendium
2014. Academica stipendium Ludwig Feuerbachi valitud teoste tõlke lõpetamiseks
2013. Keeleteaduse, filosoofia *ja semiootika* doktorikooli rahastus õpingute jätkamiseks välismaal
2012–2013. ESF DoRa programmi tegevus 6 stipendium õppimiseks välismaal
2012. Eesti Kultuurkapitali stipendium Ludwig Feuerbachi teoste tõlkimiseks
2011. Stipendiaat Göttingeni Ülikooli juures (Feuerbachi lühitekstide tõlkimise ja uurimise eesmärgil)
2010. DoRa magistrantide õpirände stipendium
2009–2010. ERASMUS üliõpilaste õpirände stipendium

Teoreetiliste teadmiste rakendamise ja populariseerimise kogemus

Avalikud loengud ja intervjuud

juuni 2019: Vaba Akadeemia avalik loeng „Kes on sotsiaalselt konstrueeritud?"
august 2018: Vaba Akadeemia avalik loeng „Mis on loomus/loomulik?"
mai 2017: ettekanne "Filosoofia gümnasistidele" loengusarja raames.

Populaarfilosoofilised tekstid

2018. "Sugu ja sugu. Kuidas ja milest me räägime?", *Eesti Päevaleht*.
2016. "Colombiast läbi filosoofi silmade". Intervjuu filosoofi Catalina Gonzálezega. *Sirp.*
2015. "Teadusfilosoofia raskekaaluline klassik kerges kuues. Carl Gustav Hempel. Loodusteaduse filosoofia". *Sirp*
2015. "St. Ghislaini porised pilved. Ivar Veermäe näitus "St-Ghislaini pilved"", *Vikerkaar* (juuni 2015).
2015. "Isiklik ja avalik virtuaalses ja reaalses ruumis. Ivar Veermäe, Karel Koplimets "Personal Record"". *Sirp*
2012. "Dekoloniseeri! Tanel Randeri näitus "Decolonize this", *Sirp*
2011. "Õnnekonverents – õnneta, aga õnnestunud", *Sirp*

Muud laadi osavõtt eriala rahvusvahelise kogukonna tööst

2019 august: konverentsi *From ϕ -science to practical realism: an international conference in honour of Rein Vihalemm (1938–2015)* korraldusmeeskonna liige

2018 juuni: konverentsi *Applications in Cultural Evolution* (Tartus) korraldusmeeskonna liige
2015 juuni: konverentsi *23rd Annual Meeting of the European Society for Philosophy and Psychology* (Tartus) korraldusmeeskonna liige.

Retsenseerinud artiklit ajakirjale Studia Philosophica Estonica ja Acta Baltica Historiae et Philosophiae Scientiarum

Osalemine ülikooli juhtimises ning institutsionaalses arendamises
2020–… Tartu Ülikooli filosoofia osakonna digivaramu toimetaja
2016–2017. filosoofia osakonna lõimekogu doktorantide esindaja

DISSERTATIONES PHILOSOPHICAE UNIVERSITATIS TARTUENSIS

1. **Jüri Eintalu.** The problem of induction: the presuppositions revisited. Tartu, 2001.
2. **Roomet Jakapi.** Berkeley, mysteries, and meaning: a critique of the non-cognitivist interpretation. Tartu, 2002.
3. **Endla Lõhkivi.** The sociology of scientific knowledge: a philosophical perspective. Tartu, 2002.
4. **Kadri Simm.** Benefit-sharing: an inquiry into justification. Tartu, 2005.
5. **Marek Volt.** The epistemic and logical role of definition in the evaluation of art. Tartu, 2007.
6. **Aive Pevkur.** Professional ethics: philosophy and practice. Tartu, 2011.
7. **Toomas Lott.** Plato on Belief (*doxa*) *Theaetetus* 184B–187A. Tartu, 2012, 208 p.
8. **Jaanus Sooväli.** Decision as Heresy. Tartu, 2013, 153 p.
9. **Ave Mets.** Normativity of scientific laws. Tartu, 2013, 217 p.
10. **Vivian Bohl.** How do we understand others? Beyond theories of mind-reading and interactionism. Tartu, 2014, 184 p.
11. **Uku Tooming.** The Communicative Significance of Beliefs and Desires. Tartu, 2014, 208 p.
12. **Andrus Tool.** Objektiivsuse teema Wilhelm Dilthey vaimuteadustefilosoofias. Tartu, 2014, 329 p.
13. **Janar Mihkelsaar.** Giorgio Agamben and Post-Foundational Political Ontology. Tartu, 2015, 169 p.
14. **Jaana Eigi.** The Social Organisation of Science as a Question for Philosophy of Science. Tartu, 2016, 282 p.
15. **Tarmo Tirol.** Subjektiivsus Jean-Paul Sartre'i eksistentsialismis. Tartu, 2017, 253 p.
16. **Eve Kitsik.** Revisionary Ontology: Improving Concepts to Improve Beliefs. Tartu, 2018, 196 p.
17. **Juhan Hellerma.** Mapping Time: Analysis of Contemporary Theories of Historical Temporality. Tartu, 2020, 138 p.
18. **Henrik Sova.** The Structure of a Consistent Global Pragmatism. Tartu, 2020, 87 p.