Causal Specificity, Biological Possibility and Non-parity about Genetic Causes

Paper presented at EPSA 2017, Section 3: Philosophy of the Life Sciences

Marcel Weber
Department of Philosophy
University of Geneva
5, rue De-Candolle
1211 Genève
Switzerland

Abstract

Several authors have used the notion of causal specificity in order to defend non-parity about genetic causes (Waters 2007, Woodward 2010, Weber 2017, forthcoming). Non-parity in this context is the idea that DNA and some other biomolecules that are often described as information-bearers by biologists play a unique role in life processes, an idea that has been challenged by Developmental Systems Theory (e.g., Oyama 2000). Indeed, it has proven to be quite difficult to state clearly what the alleged special role of genetic causes consists in. In this paper, I show that the set of biomolecules that are normally considered to be information-bearers (DNA, mRNA) can be shown to be the most specific causes of protein primary structure, provided that causal specificity is measured over a relevant space of biological possibilities, disregarding physical as well as logically possible states of the causal variables.

1. Introduction

The notion of causal specificity as it is discussed here was introduced by Waters (2007) and Woodward (2010) in an attempt to show that different classes of causes, for example in biology, differ with respect to whether they allow fine-grained control over their effect variable. The basic idea is that typical biological events such as the synthesis of a protein may have thousands of different causes, but for a large part of them the corresponding cause and effect variables take only a few different values, e.g., present or absent. Only a few select causes are such that their variable takes many different values, each value maps onto a different value of the effect variable and (almost) each value of the effect variable is caused by a distinct value of the cause variable.[1] In other words, a cause is highly specific to the extent in which its variable as well as its effect variable take many different values and there is a bijective mapping (or something close to a bijection) between the values of the cause and effect variables, respectively. This would appear to be the case for DNA and mRNA, for example: Each molecule of a given length may take a vast number of different nucleotide sequences and there is a mapping sufficiently close to a bijection between, e.g., sets of DNA sequences and their primary RNA transcripts or between mRNA sequences and the set of corresponding amino acid sequences into which the gene expression machinery translates these nucleotide sequences (disregarding the redundancy of the genetic code). Put differently, by manipulating the nucleotide sequence of DNA or mRNA, we could make any arbitrary polypeptide of a given length. The same is not true for other biochemical constituents of the gene expression machinery: RNA polymerase, for example, will obviously affect whether protein is made in a cell and how rapidly, but you cannot manipulate the polypeptide sequence of the proteins made by tampering with the RNA polymerase.

[1] See Ross (forthcoming) for a discussion of the role of causal control in biochemistry.

While Waters and Woodward initially portrayed causal specificity as a property that is either present or absent in a causal dependence relation, I argued in Weber (2006) that it admits of degrees, depending on how many distinct values of the cause and effect variables there are. I also pointed out that this allows us to replace the traditional dichotomy of information-bearing and ordinary causes of development, which is rejected by proponents of Developmental Systems Theory (e.g., Oyama 2000, Griffiths and Gray 2005), by a continuum without thereby accepting a causal parity thesis. More recently, Griffiths et al. (2015), Calcott (2017) and Bourrat (forthcoming) have used information theory in order to develop quantitative measures for different kinds of causal specificity.

While there is widespread agreement that causal specificity is an interesting concept, the question of whether it can be used to formulate a thesis of causal non-parity is still subject to debate. The minimal consensus seems to be that biological causes are not on a par in the sense that they differ in the degree of causal specificity, in whichever way it is measured. Disagreement remains in particular with respect to the question of whether there exists a way of determining causal specificity that makes a particular set of causes come out as more highly specific than all the others. Waters (2007) maintains that, at least within the set of actual-difference making causes in certain populations chosen by biologists for inquiry, DNA and mRNA are the most specific causes with respect to protein sequences. In my (2017, forthcoming) I argue that the same is true for the set of causes that can realize a high level of specificity by biologically normal interventions.

In this paper, I would like to propose an alternative approach to establishing non-parity with the help for causal specificity, namely a *modal* approach.[2] In a nutshell, I will argue that causal specificity should be measured over a suitable modal space of biological possibilities.

In the following Section, I will expose in somewhat greater detail the problem of how to measure causal specificity. In Section 3, I will introduce the concept of biological modality that I will use. In Section 4, I show how a modal restriction on causal specificity makes precisely those biological causes stand out that biologists normally refer to as information-bearers. In Section 5, I try to draw together some lose ends.

2. Causal Specificity: What Should We Measure?

Waters's and Woodward's accounts of causal specificity were modeled on Lewis's (2000) concept of influence, with some modifications:

> (INF) There are a number of different possible states of C (C_1… C_n), a number of different possible states of E (E_1… E_m) and a mapping F from C to E such that for many states of C each such state has a unique image under F in E (that is, F is a function or close to it, so that the same state of C is not associated with different states of E, either on the same or different occasions), not too many different states of C are mapped onto the same state of E and most states of E are the image under F of some state of C. This mapping F should describe patterns of counterfactual dependency between states of C and states of E that support interventionist counterfactuals. Variations in the time and

[2] My (2017, forthcoming) approach is also modal because it appeals to what could be caused by natural processes. However, biological normality is a difficult concept and it is therefore worth examining if there isn't a clearer modal criterion.

> place of occurrence of the various states of E should similarly depend on variations in the time and place of occurrence of states of C (Woodward 2010, p. 305).

With respect to Lewis (2000), Woodward (2010) has introduced two modifications: One change concerns the purpose. While Lewis introduced influence in order to define causality, Woodward only wants to distinguish between different kinds of causes, which are antecedently defined in interventionist terms.[3] The second modification is a simplification: Lewis required that the different values of the cause and effect variables be close to each other, thus requiring only minute changes to be actualized. This requirement doesn't appear in Woodward's account. The approach outlined here could be seen as reintroducing it, or at least something similar.

Woodward (2010) suggested that causal specificity distinguishes some biological causes from others. For example, among the many different causal factors that are involved in protein synthesis, the genes appear to be the most specific ones because there are many different possible DNA sequences that can give rise to zillions of different protein molecules. In a similar way, Waters (2007) has argued that some actual-difference making causes (a technical notion that he introduced himself, see Section 1) are highly specific while others are non-specific. For example, alternative splicing controlled by spliceosome complexes in eukaryotic cells (usually) has only a few possible states that map into a small set of different polypeptides. By contrast, genes have a vast amount of specificity with respect to protein sequence.

Griffiths et al. (2015) have called Waters's conclusion with respect to the comparative specificity of DNA and splice agents into question. They point out that there exist cases of alternative splicing with more than 38'000 different splice variants. In order to allow quantitative comparisons of causal specificity, they have introduced an information-theoretic measure, namely mutual information. Briefly, they define the specificity of a causal relation as

[3] For an approach that combines these two objectives see Strand and Oftedal (2017).

the difference in Shannon-entropy of the effect variable before and after the cause variable has been set to a specific value by an intervention. They then show that causal specificity so defined is about of the same order of magnitude for alternative splicing and for actual DNA variation in select biological systems.

As Weber (2017) shows, Griffiths et al. (2015) misconstrue Waters's (2007) notion of causally specific actual-difference maker. The way in which they apply their information measure to the examples, they take into account only the actual variation in cause and effect variables in specific systems. While, on Waters's account, actual variation in a population is indeed necessary for the existence of actual-difference makers in this population, causal specificity as Waters uses it (namely in Lewis's sense of influence) is a property of a range of *possible* values of the variables in question. Thus, what matters according to Waters is the specificity of the *potential* variation of the actual-difference makers, and not of the actual variation of the actual difference-makers, which is what Griffiths et al. are comparing.

While information theory may be a useful resource for studying causality in biology, for the purposes of my argument we don't really need it. In what follows, I will only need a quantity that Bourrat (forthcoming) refers to as "range of causal influence", which is the number of possible values of a variable for which there is an invariant causal dependence relation plus a bijective or near-bijective mapping of the possible states of the cause and effect variables. I shall henceforward refer to this quantity as “RCI specificity”. This is one of two components of causal specificity, the other being the closeness of an invariance relation to a bijection (Woodward 2010, Bourrat forthcoming). We could measure these quantities in terms of information as Griffiths et al. (2015) or Bourrat (forthcoming) do, but since it's not necessary for my argument I will stay clear of the concept of information and the whole technical apparatus of information theory altogether, as it would only introduce additional difficulties. Thus, I will take RCI to be simply a natural number that measures how many distinct values

the cause and effect variables take, provided that there is a near bijective mapping between them.

The crucial question is what range of values of a causal variable should be taken into account for measuring RCI specificity. At first glance there are two options:

(1) All realized values in some actual population

(2) All values possible in any population

Let us consider these two options.

Option (1). If we count only the values the are actualized in some real population, there will be lots of populations for which the specificity is higher for processes such as alternative splicing etc. than it is for the molecules that biologists refer to as information-bearers (DNA and mRNA). Griffiths et al. (2015) make that case convincingly. If the goal is to find a special causal role played by genes then we must take a different approach. Perhaps option (2)?

Option (2). This option consists in counting all the *possible* values that a causal variable my take in any population when determining its range of causal influence. For the case of DNA-protein, this could mean all the possible DNA-sequences that can give rise to a polypeptide of a given length. (I assume that this corresponds to the INF property used by Woodward and Waters). Obviously, this will return gigantic numbers for DNA sequence variation (e.g., there are 4^{1000} possible DNA strings of 1kb length). However, here we must be careful what exactly we mean by "possible values" for a causal variable. Without any constraints on possibility, we risk counting a lot of highly remote or even weird and gerrymandered possibilities. Remote possibilities would be such possibilities that could never be realized or would take too much time to evolve (see Griffiths et al. 2015, 545).

For a gerrymandered case, consider a scenario involving protein synthesis where the assignment of transfer-RNA (tRNA) and amino acid and hence the genetic code changes after each elongation step of the ribosome, i.e., after each new peptide bond formed by the ribosome. In other words, we take as possible values of the cause variable complex time-dependent

functions that indicate different tRNA populations at different times during the ribosome elongation cycle. In this way, we could generate any arbitrary amino acid sequence from any given mRNA sequence (even degenerate sequences such as poly-U or poly-A). In other words, the causal specificity of tRNA would match that of mRNA. Alternatively, we could imagine a change by intervention on the substrate specificity of aminoacyl-tRNA synthase after each elongation step in protein synthesis, which would also amount to a modification of the genetic code. In this way, too, we end up with a vast causal specificity for a molecule that is not normally considered as an information-bearer.

It should be noted that such weird and gerrymandered scenarios are ruled out if we apply Waters's criterion of actual-difference making. This notion requires that there be actual variation in a cause variable in some given population and that this actual variation account for the actual variation in the effect variable. However, this move will also rule out such populations where there is no actual genetic variation but where genes are playing the exact same role as in populations where there is variation. For this reason, we need a way of describing the causal role of genes that is less sensitive to the choice of a population.

I am going to argue that what is needed here is a restriction on the notion of possibility. We should calculate specificity not over any set of actualized values but over possible values of the causal variables, where the relevant sense of possibility should be neither logical possibility nor physical possibility – as they both permit the remote possibilities as well as weird and gerrymandered cases – but *biological* possibility instead. In other words, I am going to defend causal non-parity for DNA and mRNA by arguing that they have the highest RCI specificity at some relevant level of biological possibility. While this notion may appear elusive in light of the scant philosophical literature on the subject, there are attempts to explicate it. I will briefly introduce such an account in the following section.

3. A Primer on Biological Modality

It seems that most philosophers and biologists would accept the two following claims: (1) Every state of affairs that is biologically possible is physically possible, (2) not everything that is physically possible is biologically possible. An elephant with feathers would be physically possible but biologically impossible while a flying elephant would be both physically and biologically impossible (I guess). An immortal elephant is only biologically possible if it is physically possible. Similar claims could be made about necessity: It might not be physically necessary that an elephant show aging and senescence, but (so far as we know) it is biologically necessary.[4]

How should we think about such modalities? Are they features of organisms or of our knowledge? Do biological modalities come in degrees? Are they absolute or relative to specific lineages? And can there be a systematic theory about the validity of inferences involving such modalities and about the conditions under which modal claims are true or false? In what follows, I will ignore the question of whether modalities are about the objects of science or about our knowledge about them. By contrast, I will take a stance on the other questions: Yes, yes and yes: biological modalities come in degrees, they are lineage-relative and a logic of biological modalities is possible. To my knowledge, the first formal logic of biological modalities is due to Huber (2017). My considerations in this paper will be based on Huber's modal logic, which I will slightly adapt to my purposes.

Huber's logic refines and formalizes an idea presented by Dennett (1995). In order to conceptualize biological possibility, Dennett invented the "Library of Mendel", a library that

[4] According to some theories of the evolution of life histories, senescence has evolved due to selection pressures that limit the fitness contribution of genes that enhance survival late in life. There are also models that view senescence as an adaptation, however, most of them invoke constraints on natural selection to eradicate genes responsible for senescence or trade-offs with other life history traits such as fecundity (Stearns 1992).

contains all the genomes that can be constructed from the four DNA bases A, T, C and G. (Inspired from Jorge Luis Borges's "Library of Babel"). A "reader-constructor" maps genomes from the Library of Mendel to phenotypes. Biological possibility is then defined by Dennett in terms of an accessibility relation for genomes:

> X is biologically possible iff X is an instantiation of an accessible genome or a feature of its phenotypic products.

It is clear in Dennett's account that biological possibility is always relative to a given genome, *g*. A biological organism is possible at *g* to the extent in which it is the phenotypic product of a genome *g′* that is accessible from *g* (e.g., by a series of point mutations or sequence rearrangements). The more accessible *g′* is from *g*, the more possible is its phenotypic product at *g*.

Dennett did not really specify the relevant accessibility relation. This is where Huber's (2017) account comes in. He first reformulates the Library of Mendel as a relational structure (61):

> The Library of Mendel is a relational structure $\langle \Sigma_M, R_M \rangle$ where the domain is the language of the Library of Mendel *M* and the binary relation is the accessibility relation R_M.

The language of the Library of Mendel consists of an alphabet containing the four nucleotide bases A, G, C and T. Biological possibility is then defined in terms of satisfaction of the binary relation (61):

> Some x is biologically possible at $g \in \Sigma_M$ if and only if there is some $g' \in \Sigma_M$ such that gR_Mg' and x is an instance of g' or a feature of the phenotypic products of g'

Finally, Huber provides an interpretation of the accessibility relation R_M:

> For $g, g' \in \Sigma_M$, gR_Mg' if and only if there is a solution to a string editing problem with respect to g, g'

A string editing problem is the problem of obtaining some string of symbols from another string by the least costly set of edit operations. For example, the string 'AACTTC' can be obtained from the string 'GGCTTC' by an edit operation that replaces all Gs in the string by As. The same sequence could also be obtained by first replacing all Cs by As, then change back all As to Cs, and finally replacing all Gs by As. Obviously, the latter edit operation would be more costly. For most cases, we can identify the number of edit steps needed with the cost, in order words, the cost of each step is identical. However, there might be cases where the cost varies with the kind of change introduced. For example, we could consider operations that cannot be brought about by an existing biological mechanisms as being more costly. Alternatively, we could make the edit cost depend on the amount of metabolic energy needed or on the fitness landscape. In any case, the solution to a string edit problem depends on the assumption of a cost function and space of biological possibility is going to be relative to such an assumption.

This formulation in terms of string editing allows Huber to use the string edit distance as a measure of possibility. Several such measures exist; for example, the Levensthein distance is defined as the minimal number of operations needed to transform one string into another. As new genomes arise by mutation and recombination of existing genomes, the

number of such changes needed to obtain a new genome from an existing one seems like a biologically relevant measure of accessibility.

Thus, in brief, according to Huber the less mutational or recombinational steps are required to create a genome g' from an existing g, the more accessible and hence the more possible the latter is with respect to the former. This seems well in line with the intuition that, whatever biological possibility is, it must be relative to a given organism or lineage and it must admit of degrees.

In order to apply Huber's modal theory to our problem, we need to widen its scope a little. In particular, we must allow strings that are not representations of DNA nucleotide sequences but of something else, e.g., tRNA populations or different states of an enzyme complex such as the spliceosome. Huber's formalism doesn't cover such cases and I will not attempt to provide a modified formalism, I will simply assume that this could be worked out formally.

4. Causal Non-parity of DNA Mutation and Alternative Splicing as Causes of Protein Sequence Variation

Using Huber's theory of biological possibility, we can easily show that there exists a kind of biological variation with high causal specificity that is (highly) biologically possible and that any variation with the same causal specificity is less biologically possible.

In order to show this, Huber's account of biological modality has to be adapted to structures other than genomes, that is, to such structures as they feature in the gerrymandered scenarios. In order to adapt it to the gerrymandered tRNA case, we need to replace the DNA strings by tRNA populations. Instead of a string editing problem (how can we obtain string g' from g?) we have the problem of how we have to modify a tRNA population such as to make a certain set of protein sequences. This is going to be costly because the tRNA population might

have to be changed for each new amino acid added in order to obtain a causally highly specific mapping.

In order to be clear about what exactly we are comparing, let us consider two DNA strings that code for a protein of 300 amino acids length. Furthermore, let us assume that these two strings are separated by an edit distance (Levensthein) of two, i.e., we allow two mutation steps. Let us count only those mutations that will cause single amino acid substitutions in two different positions in the corresponding proteins. This will yield $1/2(300 \times 299)20^2 = 22{,}111{,}200$ different polypeptides. Thus, at this level of biological possibility – which appear quite high as two point mutations can easily occur in nature – we already have an RCI value corresponding to more than 22 million variants.

Could we make a matching number of polypeptides in the gerrymandered tRNA scenario? Recall that the scenario is this: We keep the DNA and corresponding mRNA sequence constant and manipulate the genetic code (i.e., the codon or amino acid specificity of the tRNA population) wherever necessary to obtain all the possible polypeptide variants that show two amino acid substitutions in two different positions. In order to do so, we might have to replace the cell's entire tRNA population several times as the elongation of the nascent polypeptide chain proceeds. It seems to me that, in this way, the entire 22 millions or so variants can be generated. Thus, the RCI specificity is the same. But note that this is more costly than if the same variety is generated by interventions on the DNA. For starters, it requires at least four interventions per polypeptide variant as opposed to two in the DNA case, because the tRNAs have to be changed back after each intervention such as to allow the rest of the sequence to be read off normally. Furthermore, we have to either replace all the tRNA molecules in the cell or, in a Maxwell's demon kind of way, catch exactly those individual molecules that are about to

deliver an amino acid moiety to the nascent chain. Even though this is not exactly the same kind of change as a series of string editing steps, it clearly is more difficult to bring about.[5]

Thus, if we presuppose a ranking of biological possibility based on the number and cost of changes necessary to bring about a certain kind change, we can conclude that the interventions that would be necessary to realize the same range of causal influence as is inherent in genes and mRNA are *less biologically possible*.

So far we have shown that there exists a level of biological possibility at which the most specific causes for protein sequence are DNA and mRNA. Now, in order to establish a kind of causal non-parity based on graded biological possibilities, we still have to show that the variability that can be generated at this possibility level by interventions on DNA/mRNA is actually greater than that of other specific causes of protein sequence, in particular alternative splicing. Let us consider the case of *Drosophila* DSCAM, also discussed by Griffiths et al. (2015) and Weber (2017). This locus is organized into 24 exons that can be combined in different ways such as to produce 38,016 different protein isoforms. In order to respect parity of reasoning, we might have to take into account not only the actual variants of this protein, but also potential variants that could result from interventions in the splice mechanism.

However, it should be noted that we cannot generate additional diversity by altering the splice signals, i.e., the sequences that direct the spliceosome complex to the splice sites and mark the intron/exon boundaries (see Alberts et al. 2015, 310-320, for details about splicing mechanisms). The reason is that, in order to count the potential variation of the splice

[5] So far as I know, there isn't even a cellular mechanism that could carry out such an intervention, however, this isn't relevant for the kind of biological possibility that I am after here. The point here is that, even if there were such a biological mechanism, it would be costlier than the mechanism required for generating the same variation by interventions at the DNA level.

mechanism itself we have to keep the DSCAM gene constant. Now, this gene is unlikely to contain another set of *repeated* signals that would be suitable for alternative splicing. Thus, if we wanted to generate more diversity by intervening on the splice mechanism, we would have to change the splice signals recognized by the spliceosome during the splicing process, i.e., in between different cuts made by the spliceosome. This could potentially produce a vast number of additional protein variants.

I think it is clear that this would be another one of the weird cases analogous to the gerrymandered tRNA cases. As in the latter case, we could argue that bringing about the necessary changes in this way would be more costly than by simply mutating the DNA- or mRNA sequences by introducing two independent mutations. Thus, such modifications are less biologically possible than modifications of genes or mRNA.

The upshot is that there exist levels of biological possibility at which genetic variation at the DNA or mRNA level has the greatest RCI specificity. In order to find variety generators of comparable specificity we always had to resort to scenarios that, on an account of biological possibility such as Huber's, are much less biologically possible. This establishes a new kind of causal non-parity in the modal realm.

5. Conclusions

Recently, Griffiths et al. (2015) have developed a quantitative measure of causal specificity based on information theory and used it to criticize existing attempts to show that DNA and mRNA are causally more specific with respect to protein primary structure than, for example, alternative splicing. In particular, they argue that a measure that they call SAD and (erroneously) take to be a quantitative version of Waters's notion of specific actual-difference making cause, gives similar values for genes and for some extreme cases of alternative splicing. In this paper, I have proposed a modal approach to defending causal non-parity against this challenge.

According to the measure developed by Griffiths et al., causal specificity is defined as the decrease in entropy or uncertainty about the value of an effect variable by setting the cause variable to a specific value by an intervention. There are two sets of issues that must be addressed separately in order to apply this measure to real cases. First, causal specificity is always relative to a *probability distribution* on the values of the two variables. Second, we must be explicit what the relevant *domain* of the cause variable (and thereby also of the effect variable) should be. This is basically the question of what kinds of *interventions* should be considered when measuring causal specificity. Obviously, the choice of domain can have as much an impact on the entropy of the variables and therefore on the causal specificity measure as the assumed probability distribution over the values of the variable.

In this paper, I have ignored issues having to do with the probability distribution of different values of causal variables in biology because these are hard to estimate. Instead, I have concentrated on what Bourrat (forthcoming) calls range of causal influence or RCI for short, where I take RCI to be a natural number. One advantage of RCI measured in this way is that it is independent of the probability distribution of the causal variables.

However, RCI is sensitive to different domain choices. I have focused on the following options for selecting a domain for determining RCI: (1) the set of all *physically possible values*, for example, on the coding sequence of a gene or the enzymatic specificity of the spliceosome complex (which mediates alternative splicing), (2) the set of all *values that have a certain level of biological possibility in a given context*, where biological possibility is understood along the lines discussed in Section 3.

(1) In the case of genes, the set of all physically possible values will include all the possible permutations of A, G, T and C in a DNA sequence that defines a certain protein-coding gene. The resulting RCI with respect to protein sequence is vast, as without any constraints on the number of possible changes the RCI is given by the number of possible proteins of a given length, i.e., 20^N if N is the number of amino acids. While this looks impressive, the set of all

physically possible values is not a good choice of domain if the goal is defending causal non-parity. The reason is that, as we have seen, this domain choice will also yield comparatively high values for biomolecules that are not usually considered to be information-bearers, such as tRNA (transfer-RNA) or the enzymes that charge tRNAs with amino acids in protein synthesis (aminoacyl tRNA-synthase). These are among the molecules that define the genetic code. Because it is *physically* possible to change that code after each chain elongation event due to an intervention, any arbitrary polypeptide could be made from any arbitrary DNA sequence. Thus, the causal specificity of these molecules matches that of DNA and mRNA. Thus, the set of all physically possible values will fail to single out just those molecules considered to be information-bearers with respect to protein sequence according to contemporary biology (usually DNA and mRNA).

(2) What about the set of all values that enjoy a certain level of biological possibility in a given context? Clearly, this domain choice will yield much lower RCI specificity figures for the DNA variable, because not all permutations of a gene are equally biologically possible to a given population. For example, biological possibility as understood in this paper is relative to evolutionary timescales. When we consider just a few generations, evolutionary processes such as point mutation can introduce considerable genetic variation into a population (actual or merely possible), but certainly not all combinatorially possible permutations of gene. But this is not necessary. Even if we allow only two independent amino acid substitutions in a polypeptide of 300 amino acids length, we already have an RCI value > 22 millions, as I have shown. This is considerably larger than the 38’000 or so protein isoforms that are possible due to alternative splicing. A greater causal specificity for alternative splicing is imaginable but not biologically possible, at least not at the same level of possibility as in the two DNA-point mutation scenario.

As I have argued, in calculating the causal specificity of alternative splicing over the some biologically possible domain, we have to disregard variation generated by modified splice

recognition sequences. Alternative splicing depends on there being *repeated* splice signals in a primary transcript, and a given gene is unlikely to contain more than one set of repeated sequences that could serve as splice recognition signals.

To conclude, measuring the causal specificity of linear biomolecules over the set of the biologically possible instead of the physically possible interventions can yield non-parity of genetic causes compared to alternative splicing and other cellular mechanisms. This also confirms that causal explanations in biology sometimes implicitly appeal to biological modalities (Beatty 2016).

References

Alberts, Bruce, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts and Peter Walter. 2015. *Molecular Biology of the Cell.* Sixth Edition. New York: Garland Science.

Beatty, John. 2016. What Are Narratives Good For? *Studies in History and Philosophy of Biological and Biomedical Sciences* 58: 33-40.

Bourrat, Pierrick. Forthcoming. Variation of Information as a Measure of One-to-One Specificity. *European Journal for the Philosophy of Science*

Calcott, Brett. 2017. Causal Specificity and the Instructive-Permissive Distinction. *Biology and Philosophy* 32: 481-505.

Dennett, Daniel C. 1995. *Darwin's Dangerous Idea: Evolution and the Meaning of Life*. New York: Simon and Schuster.

Griffiths, Paul E., Arnaud Pocheville, Brett Calcott, Karola Stotz, Hyunju Kim, and Rob Knight. 2015. Measuring Causal Specificity. *Philosophy of Science* 82: 529-55.

Griffiths, Paul E. and Russell D. Gray. 2005. Three Ways to Misunderstand Developmental Systems Theory. *Biology and Philosophy* 20: 417-425.

Huber, Maximilian G. 2017. *Biological Modalities.* Ph.D. thesis, University of Geneva.

Lewis, David. 2000. Causation as Influence. *The Journal of Philosophy* XCVII:182-197.

Oyama, Susan. 2000. *The Ontogeny of Information: Developmental Systems and Evolution.* Second Edition, Revised and Expanded. Durham: Duke University Press.

Ross, Lauren N. Forthcoming. Causal Selection and the Pathway Concept. *Philosophy of Science*

Stearns, Steven. 1992. *The Evolution of Life Histories.* Oxford: Oxford University Press.

Strand, Anders, and Gry Oftedal. 2017. Restricted Causal Relevance. *British Journal for the Philosophy of Science.* https://doi:10.1093/bjps/axx034

Waters, C. Kenneth. 2007. Causes That Make a Difference. *The Journal of Philosophy* 104: 551-579.

Weber, Marcel. 2006. The Central Dogma of Molecular Biology as a Thesis of Causal Specificity. *History and Philosophy of the Life Sciences* 28: 565-580.

Weber, Marcel. 2017. Which Kind of Causal Specificity Matters Biologically? *Philosophy of Science* 84: 574–585.

Weber, Marcel. Forthcoming. Causal Selection versus Causal Parity in Biology: Relevant Counterfactuals and Biologically Normal Interventions. In *Philosophical Perspectives on Causal Reasoning in Biology,* ed. C. Kenneth Waters and James Woodward. Minnesota Studies in the Philosophy of Science, Vol. XXI. Minneapolis: University of Minnesota Press.

Woodward, James. 2010. Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation. *Biology and Philosophy* 25: 287-318.