Forthcoming in: C. Kenneth Waters and James Woodward (eds.), *Philosophical Perspectives on Causal Reasoning in Biology*. Minnesota Studies in Philosophy of Science. Vol. XXI. Minneapolis: University of Minnesota Press

CAUSAL SELECTION VS CAUSAL PARITY IN BIOLOGY: RELEVANT COUNTERFACTUALS AND BIOLOGICALLY NORMAL INTERVENTIONS

Final version (3 August 2017)

Marcel Weber

Department of Philosophy
University of Geneva
CH-1211 Geneva
Switzerland

e-mail: marcel.weber@unige.ch

Abstract

Causal selection is the task of picking out, from a field of known causally relevant factors, some factors as elements of an explanation. The Causal Parity Thesis in the philosophy of biology challenges the usual ways of making such selections among different causes operating in a developing organism. The main target of this thesis is usually gene centrism, the doctrine that genes play some special role in ontogeny, which is often described in terms of information-bearing or programming. This paper is concerned with the attempt of confronting the challenge coming from the Causal Parity Thesis by offering principles of causal selection that are spelled out in terms of an explicit philosophical account of causation, namely an interventionist account. I show that two such accounts that have been developed, although they contain important insights about causation in biology, nonetheless fail to provide an adequate reply to the Causal Parity challenge: Ken Waters's account of *actual-difference making* and Jim Woodward's account of *causal specificity*. A combination of the two also doesn't do the trick, nor does Laura Franklin-Hall's account of explanation (in this volume). We need additional conceptual resources. I argue that the resources we need consist in a special class of counterfactual conditionals, namely counterfactuals the antecedents of which describe biologically normal interventions.

1. Introduction

The causal parity thesis in the philosophy of biology (henceforth CPT) challenges the usual ways in which different causes of biological processes are foregrounded in biological explanations. This kind of critique has been elaborated by proponents of Developmental Systems Theory (DST), an intended alternative to received views about development and evolution (e.g., Oyama 2000; Griffiths and Gray 2005). Among other tenets, DST opposes dichotomous view of ontogeny, that is, views that distinguish between information-bearing parts of organisms (e.g., DNA or mRNA) and mere information-expressing machinery, or between genetic program and parts that execute the program. There is no sense of the term "information" that applies only to DNA or RNA and to no other parts of an organism, proponents of DST argue. DST also rejects the replicator/vehicle- or replicator/interactor-distinction, including the idea of an extended replicator (cf. Sterelny, Smith and Dickison 1996). Genes aren't the only things that replicate according to DST. Epigenetic modifications of the chromatin, cell organelles, cytoskeletal structures and morphogenetic gradients are also among the things that replicate when a cell divides. What is "passed on" when an organism reproduces is a whole developmental matrix, of which genes and DNA are merely parts. The latter clearly make a causal contribution to an organism's development, but so do zillions of other parts of the developmental matrix. Thus, the inherent gene centrism of much of current biology is unjustifiable, or so it is argued.

Attempts to meet the CPT challenge that have been offered so far differ in kind. I propose to distinguish between (1) *methodological*, (2) *information-theoretic* and (3) *causal selection* attempts.

(1) Methodological attempts basically accept the CPT as a metaphysical thesis (or at least reject its contrary thesis), but maintain that gene centrism has a heuristic value for research (e.g., Schaffner 1998; Waters 2006).

(2) Information-theoretic attempts try to work out a substantive sense for the term "information" or "coding" that makes this term applicable precisely to those entities that biologists describe as information-bearing. Some of these attempts read "information" in an intentional sense, which shifts the burden to explaining how parts of even a simple organism can bear intentional content. The great evolutionary theorist John Maynard Smith (2000) has used a kind of natural selection-based teleosemantic theory in order to do so (see Weber 2005 for a critique and Shea 2007 for a sophisticated defense of this idea). Taking a somewhat different approach, Ulrich Stegmann (2005) has developed a notion of instructional (as opposed to representational) content, which he views as a special kind of natural intentionality. Non-intentional accounts include Godfrey Smith's (2000) argument that the notion of genetic coding plays a specific, non-transferable theoretical role in models of protein synthesis, Sarkar's (2003) idea of formal information systems, Jantzen's and Danks's (2008) topological causation approach, as well as Stegmann's (2012b) conception of external ordering.

(3) The causal selection attempts have no truck with the concept of information, nor with intentionality. They confront the CPT head on, i.e., they try to elaborate principles for selecting causes[1] that highlight DNA (and perhaps messenger-RNA) as being a distinct kind of cause. One such attempt is due to Ken Waters (2007), who has worked out a distinction between *actual*- and *potential*- difference

making cause. In a nutshell, actual-difference making causes are those causes that account for the *actual* variation in some trait or property in a population. Any causally relevant factor that does not account for some actual variation in an existing population is merely a *potential*-difference making cause according to Waters's conception. Thus, trivially, traits that do not vary (or do not relevantly vary) have no actual-difference making causes, but they may have any number of potential-difference making causes. By the same token, causally relevant factors that do not vary are barred from being actual-difference making causes, but they may still be potential-difference making causes.

Waters's notion already goes some way in refuting the CPT. For many parts of an organism's developmental matrix are constant in actual populations. For example, many parts of the protein-synthesizing machinery in cells do not exhibit relevant variation in actual populations ("relevant" meaning basically that they satisfy Waters's definition of actual-difference making cause, more of which later). Only the DNA and mRNA (messenger-RNA) do. Waters argues that, in prokaryotic cells (i.e., bacteria), sequence variation in DNA and mRNA sequence are the only actual-difference making causes with respect to protein sequence, or, to be precise *the* actual-difference making cause.[2] All the other parts of the developmental matrix are merely potential-difference making causes, at least with respect to the amino acid sequence of the proteins made by a cell. In eukaryotic cells (i.e., animals, plants and fungi), by contrast, there are other cellular components that are also actual-difference making causes. An example is provided by the agents involved in the processing of primary transcripts that are precursors to mRNA. Many eukaryotic genes exhibit alternative splicing, i.e., the non-coding parts of the corresponding transcripts can be cut our ("spliced") in many different ways.

Waters suggests that even in such cases, DNA and mRNA are not causally on a par with other parts of the cell. The reason is that, while they are not the only actual-difference making causes with respect to a cell's population of protein species, they are the *causally most specific* actual-difference making causes.[3] The notion of causal specificity is akin to David Lewis (2000) notion of *influence*, which Lewis introduced as part of a general (reductive) theory of causation. This relation holds whenever there is a counterfactual dependence of different (but not too different) modifications of an event *E* on different modifications of an event *C* such that there is bijective mapping (or something near enough) of the *C*-states into the E-states. Woodward (2010) has refined this notion and reformulated it in terms of his interventionist theory of causation. According to Woodward, INF (as he abbreviates it), may be useful for countering claims of causal parity, as it might describe precisely the role of genes in protein synthesis.

Thus, on Waters's and Woodward's views, the CPT is only plausible given some crude conception of causation. There exist more sensitive causal notions that are responsive to subtle differences in causal role. Their respective accounts are attempts to explicate these subtle notions.[4]

I classify Waters's and Woodward's attempts under the heading "causal selection attempts" because they attempt to provide principles that foreground one specific kind of cause from a motley of causally relevant factors in the ontogeny of an organism. This is the problem of causal selection. It should be distinguished from another problem in the theory of causality, namely the problem of “actual” (or sometimes “token”) causation. The latter problem pertains to cases of singular or

token causation, including cases of causal overdetermination and pre-emption (e.g., Hitchcock and Knobe 2009). There, we sometimes want to know what factor from a set of causal relevant factors "actually" caused an event, or "made the difference" in the outcome, and the like. Our present problem is somewhat different, as it pertains mainly to causal regularities. It may apply also to token causation, however, this is controversial.[5] Nonetheless, there are similarities. In fact, the similarities are strong enough for solutions to one problem being at least helpful for solving the other, as I will show later.

This paper has two goals. First, I will show that, while both Waters's and Woodward's accounts are helpful and steps in the right direction, they do not quite succeed in answering the CPT challenge. Franklin-Hall's (in this volume) account of scientific explanation also doesn't fit the bill, as I will show. Second, I will present an account that does succeed in answering that challenge. The basic idea is similar to Waters's and Woodward's attempts: We must introduce fine-grained distinctions within our causal concepts. However, I will argue that the distinction between actual-difference makers and potential-difference makers cannot account for causal selection in the cases under consideration, even if augmented with the notion of causal specificity. Rather, what is doing the work is causal specificity measured over a special kind of potential variation, namely variation that could be caused by biologically normal interventions. This will often select actual-difference makers in Waters's sense, but it will sometimes select some potential-difference makers that are not actual-difference makers. Further, I will show that Woodward's notion of INF alone will not do the job of selecting the causes that biologists foreground; it turns out to be too permissive. What we need is a distinction between *relevant* and *irrelevant counterfactuals*, where relevant

counterfactuals are such that they describe *biologically normal possible interventions*.

In the next Section, I critically examine Waters's account, in Section 3 the account of Woodward. This will reveal the shortcomings of both accounts. I will also show why a combination of the two accounts also doesn't quite do the work of refuting the CPT. In Section 4, I consider and reject two further possibilities, namely Lewis's (2000) original notion of influence as well as Laura Franklin-Hall's (this volume) causal economy account of scientific explanation. In Section 5, I present my alternative, which uses the concept of biologically normal interventions in order to draw a distinction between relevant and irrelevant counterfactuals in causal attributions. Section 6 draws out some more general implications for the philosophy of biology and perhaps general philosophy of science.

2. Waters's Conception of Actual-difference making Causes

Ken Waters (2007) argues that philosophical theories of causation, while putting much effort into distinguishing between causal and non-causal relations, have not tried hard enough in distinguishing between causal factors and background conditions, which is the problem of causal selection. Background conditions are also causally relevant, but are nonetheless classified as background. For example, we rarely cite the presence of oxygen as the cause (or even a cause) of a fire, even though it is causally highly relevant for fires. There seems to be a philosophical consensus that this factor versus background-distinction has no ontological basis at all but is only a matter of pragmatic interests (see, e.g., Mackie 1980, p. 35).

Waters opposes this consensus and suggests that we need to distinguish *actual*- from *potential*-difference making causes. The former are causal factors that actually vary in a population and that are responsible for the actual variation in some effect variable. The latter, by contrast, while also being causally connected to the effect variable, do not *actually* vary, nor do they account for some actual variation in the effect variable, they only *potentially* do so. Waters thus offers a solution to the problem of causal selection in the context of causal regularities. This account is based on Woodward's (2003) general theory of causation.[6] The result is the following account:

Here is Waters's definition of "*the* actual-difference making cause" (henceforth "*the* ADMC"):

> X is the actual difference maker with respect to Y in population p iff:
>
> (i) X causes Y (Woodward).
>
> (ii) The value of Y actually varies among individuals in p.
>
> (iii) The generalization 'X causes Y' is invariant with respect to the variables that actually vary in p.
>
> (iv) Actual variation in the value of X fully accounts for the actual variation of Y values in p.

In this account, the expression "fully accounts" should be understood in the following way:

> (a) Individuals with the same X values in p have the same Y values.

(b) An intervention on X with respect to Y that changed the X-value of all individuals in p to the value that one and the same individual had without intervention would change Y-values in p such that they no longer differed.

(c) There is no variable Z, distinct from X, such that an intervention on Z with respect to Y that changed Z values in one or more individuals in p to the Z value that one of the individuals had without intervention would change Y values in p.

Next, Waters defines his concept of *an* actual-difference making cause (henceforth *a* ADMC):

X is *an* actual difference maker with respect to Y in population p iff:

(i) X causes Y (Woodward).

(ii) The value of Y actually varies among individuals in p.

(iii) The generalization 'X causes Y' is invariant over at least parts of the space of values that other variables actually take in p.

(iv) Actual variation in the value of p partially accounts for the actual variation of Y values in population p

where the expression "partially accounts" is to be understood as follows:

An intervention on X with respect to Y that changed the X values in one or more individuals in p to the X value that one of the individuals had without intervention would change Y values in p ("difference changer").

Any cause that satisfies Woodward's criteria for "X causes Y" but fails to satisfy any of Waters's criteria for actual-difference making causes is a *potential*-difference making cause.

Waters argues that his concept of *the* ADMC allows a precise distinction of the causal role played by DNA and RNA compared to other cellular machinery, which amounts to a rejection of the causal parity thesis. At least in prokaryotic cells, DNA sequences are the ADMC with respect to a cell's population of mRNA molecules, and mRNA molecules are the ADMC with respect to a cell's population of proteins.

Waters is aware that causal relations are more complex in eukaryotic protein synthesis. There, other ADMCs with respect to a cell's population of protein exist. A case in point is splicing agents that remove different introns (non-coding intervening sequences that are found in many eukaryotic genes) from the primary transcripts (unprocessed RNA copies of certain DNA regions). These splice agents are an important source of actual protein sequence variation and thus also qualify as *a* ADMC in Waters's sense.

In order to claim a special role for eukaryotic DNA in protein synthesis (again, in opposition to the causal parity thesis), Waters uses the concept of *causal specificity*, which he adapts from David Lewis's concept of influence. I will discuss this notion in the following Section. Here, it suffices to note that Waters's concept of *the* ADMC is sufficient to describe the special causal role of nucleic acids in prokaryotic protein synthesis, while a combination of the concept of *a*

ADMC and causal specificity is needed in order to maintain such a special role in the eukaryotic case, as, nucleic acid is only *an* ADMC, not *the* ADMC in cells of higher organisms. Other ADMCs include splice agents, RNA editing, and post-translational processing mechanisms.[7]

I now turn to a critical assessment of Waters's account.

I shall begin with a central feature of ADMCs, namely their always being relative to the choice of a population, which is evident in Waters's definition. It depends on there being *actual* variation in the chosen population (which is usually the case, as cells make thousands of different proteins). Thus, if we consider a single event of protein synthesis *in isolation*, the concept of ADMC is not applicable. This seems counter-intuitive, as they would maintain that whatever special role DNA or RNA in the synthesis of proteins must also be manifested in the single cases.

There is no reason for Waters to be impressed by this objection, as so far it is based purely on intuitions. Indeed, he anticipated it and suggests that his account, in fact, gives the correct analysis of such cases. Imagine that there was no DNA or RNA sequence variation. In this case, Waters argues, we might see the role of these nucleic acids merely in providing a scaffold for the synthesis of (ever the same) protein. Nobody would be inclined to talk about programming or information. Thus, it does seem that the causal role we attribute to different parts of a cell is sensitive to where the actual variation is.

However, the objection can be strengthened. Stegmann (2012a) argues that there were actual *experiments* in biology where there was no sequence variation. For example, the well-known experiment of Matthaei and Nirenberg that paved the way towards the cracking of the genetic code is a case in point. In this experiment, a cell lysate was programmed (if I may use this metaphor once) with poly-U, an artificial RNA that contained only the base uracil. To this, I would like to add that there are *physiological* or *differentiation states* where cells make only a single protein such that the genes and mRNA of this cell are also not ADMCs in Waters's. Reticulocytes (red blood cell precursors) are an example; they make only hemoglobin. Thus, we don't need to resort to thought experiments and intuitions; there are real biological counterexamples.

To these counterexamples, Waters could still reply that if all cells were indeed like reticulocytes, there would be no reason for attributing a special causal role to genes. It is the fact that *not* all cells are like reticulocytes (and that the reticulocytes have precursors that make as many different proteins as any other cell) that defies the causal parity thesis. Nonetheless, it remains true that we seem to have real cases of protein synthesis without actual-difference making where it seems natural to describe the role of nucleic acids in exactly the same terms as in cases where there is actual-difference making going on.

I will make a novel suggestion as to how to handle this problem in Section 5. But first, I shall discuss some other attempts.

3. Woodward's Conception of Causal Specificity

The source idea of causal specificity used by both Waters and Woodward is due to David Lewis (2000), who used the term "influence" rather than "causal specificity". Lewis's goal in introducing this notion was rather different; it was part of an attempt to distinguish conceptually between causal and non-causal relations. By contrast, both Waters and Lewis use this (or a similar) concept in order to distinguish between *different kinds* of causal relations. Here is Lewis' definition of influence:

> Where C and E are distinct actual events, let us say that C influences E iff there is a substantial range C_1, C_2, ... of different not-too-distant alterations of C (including the actual alteration of C) and there is a range E_1, E_2, ... of alterations of E, at least some of which differ, such that if C_1 had occurred, E_1 would have occurred, and if C_2 had occurred, E_2 would have occurred, and so on (Lewis 2000, p. 190).

Thus, Lewis suggested that causal relations differ from non-causal ones by there being some kind of bijective mapping or something close to it between two different kinds C and E of actual event states that holds in counterfactual situations. It should be noted that Lewis does not require a strictly bijective mapping, as in his definition some of the E-states could be the same.

Woodward (2010) modifies this conception in two ways (not counting the different philosophical goal associated with it, which was already mentioned): First, he replaces *events* and their modifications by *variables*. Second, he drops the qualification "not-too-distant alterations of C". Lewis envisioned that for there

to be causal influence, *small* changes in the C-states must lead to changes in the E-states. Woodward, by contrast, places no restrictions on the magnitude of the changes expressed by the C-variables. By doing this, he might lose one of the main potential strengths of the influence relation[8], to which I will be come back in Section 4.

Woodward is also quite vague (deliberately so) on the nature of the mapping from the C-variables into the E-variables. Like Lewis, he does not require a strict bijection. Surjective/non-injective and injective/non-surjective mappings[9] are also permissible, so long as they are *functions* (i.e., if in one or several situations the independent variable(s) take(s) the same value, then so does the dependent variable).

Here is Woodward's definition in its entirety:

> (INF) There are a number of different possible states of C ($C_1 \ldots C_n$), a number of different possible states of E ($E_1 \ldots E_m$) 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 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).

Before I further assess the utility of this concept in analyzing causal relations in molecular biology, I would like to draw attention to the fact that INF admits of *degrees*. Relations that fall under INF may vary quantitatively in at least two dimensions: First, they may differ in the *number of different states* that are mapped by INF. The variables C and E may take discrete values, and their domains may be finite. This is actually the case with DNA and protein sequences. Alternatively, the domains may be infinite but denumerable. The third possibility is that the variables are real-valued and therefore infinite and non-denumerable. Woodward (pers. communication) allows for all these three possibilities, however, I am here particularly interested in cases where the variables are discrete and the domains finite. For it is in these cases that the notion of *degrees* of causal specificity makes sense. A causal relation that satisfies INF is the more causally specific the more different states it maps onto each other.

Second, relations that satisfy INF may differ in their *closeness to a bijective F-mapping*. This closeness also admits of degrees because the elements in the codomain may be mapped onto by different numbers of arguments from the domain (in the surjective and non-injective cases), or different proportions of elements in the codomain may by mapped onto by an argument from the domain (in the injective and non-surjective cases). A combination of both is also conceivable.

Thus equipped with this very specific and subtle causal concept we may now venture a new look at the process of gene expression.[10] It is appropriate to look at the different stages of gene expression separately. In my table below, we assume that there is a more or less stable causal relationship between the entities on the left and those on the right-hand side of the '→', which thus signifies a *causal graph*. Furthermore, we assume that these entities (for instance, RNA or protein), vary in their primary structure (nucleotide or amino acid sequence, respectively), and that their sequence state can thus be expressed by two variables, say *x* and *y*. These variables take *discrete* values. The domain of the variables will depend on the length of the molecules considered, but it is always *finite*. Then, what the table claims is that there exists a range of background conditions under which an intervention on *x* would change the value of *y*, for example, an intervention on mRNA sequence would change protein sequence. The inverse does not hold; an intervention on protein sequence would not change mRNA sequence (under normal conditions). Furthermore, in accordance with Woodward's definition of INF, there exists a mapping, F, that maps different states of *x* onto different states of *y*. The following table states whether or not F is a function (which is necessary for the relation INF to hold) and whether F is bijective, injective or surjective:

Stage of gene expression (x → y)	**F-mapping y=F(x)**
(1) DNA → DNA (=replication)	bijective (INF)
(2) DNA → RNA (in prokaryotes)	bijective (INF)
(3) RNA → DNA (=reverse transcription)	bijective (INF)
(4) DNA → primary transcript (in eukaryotes)	bijective (INF)
(5) primary transcript → mRNA (in eukaryotes)	not a function, therefore not-

	INF
(6) primary transcript (exon parts) → protein domains*	surjective & non-injective (INF)
(7) mRNA → protein*	surjective & non-injective (INF)

*unprocessed/unedited precursors only

This table shows that different stages in gene expression differ considerably with respect to causal specificity. Some stages, namely (1) – (4) are characterized by a bijective F-mapping, which is guaranteed by Watson-Crick base-pairing. A break in causal specificity appears to occur in eukaryotes at stage (5) in the processing of primary transcripts, at least insofar the primary transcript → mRNA link is concerned. This is due to the mechanisms mentioned, namely alternative splicing and RNA editing. In many genes, the same primary transcript can give rise to a variety of different mRNAs. However, this case is different to begin with, because those parts of a processed mRNA that map bijectively to exon (or other) regions in the primary transcript are *token-identical* to the latter (because they result from simply removing the introns by the process known as "splicing").[11] So this relation is not even a candidate for a causal relation such as INF, which requires non-identical relata.

But it should be noted that the INF relation does hold for some of the *exon parts* of a primary transcript in relation to *parts* of the finished proteins (stage 6). In many cases, these protein parts constitute domains. Thus, if Woodward's relation INF is an adequate expression of the causal determination postulated by Francis Crick in 1957, a version of the "Central Dogma of Molecular Biology" (see

Weber 2006 for a conceptual analysis) survives until today, in spite of the zoo of newly discovered RNA processing mechanisms.

Interestingly, the exon – protein domain (6) as well as the mRNA – protein (7) links merely show a *surjective and non-injective* mapping, which is due to the redundancy of the genetic code. Because several base triplets encode the same amino acid, quite a large number of different mRNA molecules can be associated with one and the same protein sequence.

What is not shown in the table above is that there are *other causal variables* in cells that may bear the INF-relation to RNA or protein molecules of a specific sequence. Examples include splicing agents, RNA editing enzymes, or enzymes that introduce post-translational modifications into freshly synthesized polypeptides. However, the *degree* of causal specificity in these cases seems to be smaller than that of the nucleic acid-nucleic acid or nucleid acid-protein links that appear in the table above. For example, even if there are genes that give rise to several hundred different proteins by alternative splicing, this is still far less causally specific than the zillions of different polypeptides that could be made by sequence alterations in DNA or RNA coding regions of some given length.[12] In terms of INF, this means that the domains of the F-mapping are much smaller than those of any of the gene expression stages (1) – (7) according to the table.

It should be noted that this allows for an alternative to both the causal parity and to the dichotomous views: The simple dichotomy of information-bearing and non-information bearing components that was criticized by Oyama (2000) and others

may be replaced by a *continuum* of causal relations that differ in causal specificity.

As these considerations show, Woodward's concept INF is a powerful tool for analyzing causation in molecular biology, in particular the kind of causal determination that Crick couched in terms of information transfer.[13] However, it also is important to understand the limitations of this conception. To expose one important limitation, it should be noted that there are also other cellular components that instantiate INF with respect to gene products to a very high degree. An example is tRNA (transfer RNA), the "adaptor" molecules that carry amino acids to a growing polypeptide chain in accordance with the genetic code and with the mRNA that is being translated.[14] I claim that there is a bijective F-mapping between each set of tRNA molecules and the protein if we hold the mRNA constant. By varying the codon specificity of each tRNA that gets charged with a specific amino acid by aminoacyl tRNA synthetase, we could also make different proteins with one and the same mRNA. Thus, INF alone is too permissive for expressing Crick's "sequence hypothesis" and "Central Dogma".

At this stage, Waters's concept of actual-difference making cause could be called to the rescue. In contrast to mRNA, tRNA does not *actually* vary in living cells. In fact, the genetic code that is laid down by a cell's complement of tRNAs is almost invariant across all the five kingdoms of life (known exceptions include mitochondria and trypanosomes, the causative agents of sleeping sickness). Talk about a causal factor that doesn't actually vary! Thus, INF and ADMC could be combined in order to express the very subtle kind of causal determination that

some prefer to couch in terms of information or programming. INF alone doesn't do the job, it needs to be at least supplemented with something like ADMC.

The main problem with this approach is that we still cannot account for those cases where no actual-difference making occurs. Note also that this approach inherits the counterintuitive consequences that were mentioned in Section 2: that a case considered in isolation will not exhibit the special role that at least some people want to attribute to genes and mRNA. The source of both of these problems is the relativity of ADMCs to the choice of a relevant population or reference class. As a result, the resulting distinctions may be viewed as not being *ontological* distinctions, or at least not ontological enough.

This is why we will need to dig into richer conceptual resources if we really want to understand the principles that guide causal selection in certain areas of biology. I will first check if Lewis's original notion of influence might be such a resource. Furthermore, I want to examine the potential of Franklin-Hall's (this volume) account of scientific explanation for solving the problem at hand. Then, I will show that all this will still not do, that we need yet another conceptual resource.

4. Lewis's Conception of Influence and Franklin-Hall's Account of Explanation

As we have seen in the previous section, Lewis's notion of influence, unlike Woodward's, contains the idea that causal relations are characterized by the counterfactual dependence of modifications of the effect event on *not-too-distant* modifications of the cause event. Lewis introduced this idea in response to

difficulties of his counterfactual theory of causation in dealing especially with cases of causal redundancy and pre-emption.

Consider the case of Suzie and Billy throwing rocks at a bottle. They both hit, but Suzie's rock arrives first and shatters the bottle into thousands of pieces. Now, it is not true that the bottle would not have shattered had she not thrown her rock, because Billy's rock was also on its way and would have smashed the bottle had Suzie's rock not already done so. This is clearly a difficulty for the counterfactual theory. In order to deal with this difficulty, Lewis wanted to make use of the fact that there is a short time span during which Suzie has more control over the exact time as well as the specific way in which the bottle bursts into pieces than Billy. She could not have delayed the time of shattering beyond the time when Billy's rock arrives, but before this happens she is in the driver's seat so far a timing is concerned, because she threw her rock first. This being in the driver's seat of the actual cause events is the hallmark of causal influence according to Lewis's 2000 paper. For this approach, it is instrumental that the modifications of the cause events are not too distant, for otherwise the difference to the redundant cause vanishes. What precisely "not too distant" means is difficult to specify, however.

Using this idea of Lewis, it could be argued that there is the following difference between, say, an mRNA (which is said to "code" for protein) and a tRNA (which is not): By tampering with the mRNA and holding everything else constant, we could make any possible protein. As we have seen, the same is true for tRNA, and also for other molecules involved in protein synthesis such as aminoacyl-tRNA synthase. However, there is a difference: In the case of mRNA and genes we could make the full range of possible proteins by just *altering a single molecule,*

(mRNA) or a part of a single molecule (genes). By contrast, in the case of tRNA we would have to replace many tRNA molecules. Also, we might have to replace (some of) them after each ribosome cycle, while the mRNA or genes have to be changed only once for all of the ribosome cycles needed to complete the protein, and even to make more of the same kind. So one could argue that in the case of mRNA, the changes needed to realize the full causal specificity of the causal link are smaller than in the case of other biomolecules.

Why should this matter? In other words, what makes this difference in the causal link *relevant* for causal selection and scientific explanation?[15] Note that the case is different from the cases of causal redundancy discussed by Lewis (2000). There, the modifications of the cause events have to be small because otherwise the counterfactual differences between the actual cause and the pre-empting cause vanish. But here, there is no pre-empting cause and no causal redundancy; all of the cell components that we are talking about are causally necessary for gene expression to be operational. So, again, why does it matter that the changes needed to realize the causal specificity of the nucleic acid-protein link appear smaller in the case of so-called information-bearing molecules than in other cases?[16]

Is it possible that this question could be answered with the help of Laura Franklin-Hall's causal economy account of explanation (in this volume)? Franklin-Hall thinks that events are explanatory to the extent in which they provide the most bang for the buck, where the "buck" is the amount of detail needed to describe the event and the "bang" is the modal stability boost that it provides. A stability boost means that the cause event makes the effect event significantly more modally

robust, which means that the event would still have occurred (in the same way) even if the world had been significantly different in the past. In other words, the cause event stabilizes the event in a range of nearby possible worlds.

While this is a rich and interesting conception of explanatory power, I don't see how it could solve our current problem, which is to say why it *matters* that it takes smaller changes to realize any arbitrary protein if we intervene on mRNA compared to tRNA (for example). For in the present case, we are not holding the effect event (=the synthesis of a specific protein) fixed and examine its presence or absence in nearby worlds present or absent some cause. Rather, we are asking what *would* happen had the mRNA been *different*, or had the tRNA been different. It might be true that we could realize the full causal specificity of this link with smaller changes in the case of mRNA (or other molecules that are said to transmit information, such as DNA). But this is not about the modal stability of the effect event, which according to Franklin-Hall is what we want to spend our buck on. It seems to me that the modal stability boost afforded by the presence of a particular species of mRNA, say, is about the same as that provided by a specific set of tRNAs. In both cases, no protein would be made if the molecule in question was absent, and a different protein would (possibly) be made would there be different RNAs present. But no one of these causes makes the synthesis of a specific protein any more modally robust than the other (i.e., insensitive to changes in conditions in the environment). So the bang is the same. Also, I don't see why describing one set of entities compared to the other should be more expensive in terms of the amount of detail needed, so we don't get more bang for our buck. The only difference that I can see here is that for some molecules it takes smaller (if there is such a thing) or less interventions to realize the full

causal specificity of the causal link. Franklin-Hall's account does not tell us why this makes a difference with respect to causal selection.

In the following section, I will present an account that answers this question. On this account, it is not the size of the interventions that would be needed to change protein sequence in a maximally causally specific way that matters, but the question of whether these interventions are *biologically normal*.

5. Relevant Counterfactuals and Biologically Normal Interventions

Let us take stock first. I have shown that Woodward's relation INF does not single out DNA and mRNA as a uniquely specific cause of protein sequence. tRNA and a few other molecules fit that bill as well, as there is a bijective mapping from tRNA to protein sequence. In other words, if we hold the mRNA constant we can still make any arbitrary protein sequence by changing the amino acid-codon specificity (which amounts to manipulating the genetic code). This may be hypothetical, but that's all an interventionist causal theorist needs. Could we perhaps argue that, by manipulating the mRNA and holding the tRNAs fixed, we could make any possible protein sequence of any length, while the converse does not hold: By manipulating the tRNAs and holding the mRNA fixed we cannot make any arbitrary protein? Well, there is one difference: In the first case, we can make any protein of any arbitrary length, while in the second case the length of the protein is given by the length of the mRNA which we hold fixed. However, maybe this could be overcome somehow.

If this is granted, then the following problem arises: We can *still* make any arbitrary protein sequence by tinkering with the tRNAs simply by exchanging the tRNAs during the process of translation.[17] Maybe we would have to do this after each ribosome cycle. But since we are talking about *hypothetical* interventions and counterfactuals, nothing can prevent this. So we are still left with the problem that tRNA exhibits a bijective INF relation to protein sequence just as much as mRNA does. Must we therefore finally give into the CPT?

I suggest that we don't. The key to differentiating the role of genes and mRNA from that of other parts of the gene expression machinery lies in the recognition that *all counterfactuals are not equal*. Some counterfactuals are more relevant than others.

Here is an example (Hitchcock and Knobe 2009, 590f.). A student has got an F in a test. She reasons how this undesirable state of affairs might have been prevented. Clearly, she would not have gotten the F had the teacher been eaten by a lion, or had the Earth's gravitational pull suddenly ceased. She would also not have gotten the F if she hadn't stayed up drinking until late the night before. Now, the first two of these counterfactuals seem awfully irrelevant, while the third doesn't. Unlike the others, this last counterfactual is worth entertaining, because doing so will allow the student to improve her performance on the long run. The other counterfactuals simply aren't worth even to consider. Thus, the student will select her drinking as the cause of her F, not the Earth's gravitational field that held the teacher on the chair while she was grading the test or the lion's not being hungry.

Hitchcock and Knobe (2009) argue that the relevance of counterfactuals does matter in our causal reasoning. Namely, it matters in *causal selection*. We often select those causes from a causal field that would have been potent targets for interventions to change some outcome in some *realistic* way. In everyday life, the realistic ways may be the ones that are within the grasp of our own actions.

What is more, we often select causal factors that are targets for *normatively acceptable interventions*. An example for the latter case is also provided by Hitchcock and Knobe: The departmental administrators keep pens the are reserved for them. A professor takes a pen, which she was not supposed to do. One morning, there are no more pens. Who is the cause of this? The professor who took the pen, because it would haven been normatively acceptable to intervene in her taking the pen. Thus, causal selection is sometimes guided by norms.

According to Hitchcock and Knobe, the function of causal selection is (sometimes) to direct us to those parts of a causal structure where we could intervene in a way that is normatively acceptable. While, for a causal interventionist, the set of all possible interventions defines a causal structure, norms sometimes pick out a subset of possible interventions that are in the range of actions for which we might have good reasons.

I suggest that something similar is going on in certain causal explanations in biology. Of course, I don't want to suggest that causal selection in biology is also guided by the normative acceptability of certain interventions. However, there is a biological analogue: There is a class of interventions on biological systems that is

distinguished by being *biologically normal*. In what follows, I will try to spell out this notion a little.

The idea is not to introduce some value-laden notion of normality, nor a statistical notion. By biologically normal interventions, I mean simply interventions in the sense of Woodward (2003) that satisfy the following two additional conditions:

> (1) the intervention *may also be due to natural processes* such as spontaneous mutation, replication error, transposition, etc. (the *cetera* include all known natural causes of genetic variation)
>
> (2) the intervention is *compatible with the continued persistence of the biological entity that is being considered*

As in Woodward (2003), such interventions may figure in counterfactual claims of the sort "if an intervention would occur that changed the nucleotide sequence at position X of the *Drosophila* genome, the resulting fly would make protein Y_1 instead of Y_2." Woodward's (2003) interventionist theory of causation does not systematically distinguish between different kinds of interventions, e.g., ones that require human agency and ones that could occur by natural processes. They are equally suitable for expressing causal claims. I want to claim here that, *given biology's explanatory goals*, this makes a difference. More precisely, I want to argue that such counterfactuals that state biologically normal interventions in their antecedents, which I shall for the sake of brevity simply refer to as "biological counterfactuals", are what's sometimes behind the selection of certain causal factors in biology.

The idea is that genes are the most specific *potential*-difference making causes if the potential differences in question are limited to differences the could be brought about by *biologically normal interventions*.

In order to make this claim plausible, I shall consider the case of protein synthesis again. As we have seen, there exist highly specific patterns of counterfactual dependence between the amino acid sequences of proteins made by a cell and the nucleotide sequence of genes as well as mRNA. However, there are equally specific patterns of counterfactual dependence between the amino acid sequences and the primary structures of tRNA and aminoacyl tRNA- synthetase molecules, and possibly others. There are no grounds for selecting one class of cause as "determining the amino acid sequence of proteins" (as Francis Crick might say). So far, there is a perfect causal parity of determination between the various components involved in the process. However, things change when we ask what kinds of interventions would change the outcome of the process. Some interventions are consistent with continuing biological functioning, whereas others are not. Substituting one mRNA for one with a different sequence is normally so consistent, so is substituting a gene. Substituting tRNAs and aminoacyl-tRNA synthetases for molecules with different specificities, by contrast, is not consistent with continuing biological functioning of the process of protein synthesis. For if the cell suddenly contains different tRNAs and aa-tRNA synthetases, this will affect the sequence of other protein molecules made by the cell, which makes it impossible for it to survive.

Against this suggestion, it could be argued that the relevant interventions must be conceived as "surgical". This means that we imagine changing just those token tRNA or aa-tRNA synthetase molecules that will bring about the desired change in the structure of the final product, while leaving all the other molecules in the cell unaffected. We might even imagine that the modified molecules are removed from the cell right after they have done their job, such that the cell will not be poisoned by them. Construed in this manner, these interventions would also be compatible with biological functioning. However, they would not *naturally occur* under normal biological functioning.

Thus, it seems that we have two kinds of possible interventions: Such interventions that could naturally occur as part of the normal biological functioning of an organism and interventions that do not normally occur in this manner. It is part of the biological function of tRNAs and aa-tRNA synthetases that their biological specificity remain constant. They define a cell's genetic code, and a cell can only survive if that genetic code stays the same. By contrast, it is not part of the biological function of genes and mRNAs to maintain their biological specificity.

It is a somewhat controversial issue if genes have any biological function. Dawkins (1975) has argued that they lack a function altogether; as they are the ultimate beneficiaries of all biological functioning. However, we do not have to commit ourselves on this issue. All I need for my account is that the functional significance of tRNA & Co. differs from that of genes and mRNA such that only the former, but not the latter have the function of providing for a stable genetic code (i.e., codon-amino acid association in protein synthesis).

We also need not commit ourselves to a specific view about biological functions. Such views are many: some think that functions are evolutionary adaptations (Sober 1993), others that they are dispositions to survive (Bigelow and Pargetter 1987), yet others that they are contributions to some systemic capacities (Cummins 1975, Weber 2017b), and so on. No choice needs to be made here: All I need is that all of these accounts recognize that it is a biological function of tRNAs & Co. to provide a constant catalytic environment for protein synthesis, such that the association of nucleic acid codons and amino acids in protein synthesis stays the same. It seems to me that life as we know it would be impossible without such a stable association. At the same time, no such function must be attributable to genes.

For genes and mRNA it is biological normal to exhibit sequence variation, both within the same genome and in whole populations of organisms. Again, life as we know it would be impossible without it.

To come back to the relevant counterfactuals: I suggest that the relevant counterfactuals that guide causal selection in biology are counterfactuals that describe interventions that are *biologically normal*. These counterfactuals pick out a set of causes, some of which may be actual-difference makers (in some relevant populations), others merely potential-difference makers. From this set, I want to claim, biologists select DNA, genes and mRNA for the following reason:

Among those actual- and potential-difference making causes of protein sequence that can be actualized by biologically normal interventions, genes, DNA and mRNA are the causally most specific.

This claim obviously requires some elaboration. I shall elaborate it in three steps.

First. By "causally most specific" I mean that genes bear Woodward's relation INF to proteins in the highest degree. By "the highest degree" I mean that the number of values that the variables on both sides of the INF relation can take is vastly higher (i.e., many orders of magnitude) than that of any other causal variables that bear the relation INF to protein sequences (e.g., splicing agents).[18]

Second. By "potential-difference making cause" I mean that there exists a counterfactual dependence of the protein sequence on the nucleotide sequence of the gene such that some interventions on the gene sequence would change the amino acid sequence provided that all other difference-making causes w.r.t. to protein sequence remain constant. This is already implied by the INF relation.

Third. According to my definition given above, biologically normal interventions are such that they (1) could be brought about by natural biological processes and (2) don't kill the organism considered.[19] Normally, clause (2) will amount to the cell's protein synthesis machinery as well as other biochemical mechanisms continuing to perform their biological function. Also note that clause (2) excludes evolutionary change, because it normally requires the death of individuals. The two clauses (1) and (2) are meant to together exclude such interventions that tamper with the protein synthesis machinery (tRNAs etc.) in such a way that only

the synthesis of a single protein (type or token) is affected. For this kind of "surgical" intervention would be required to change the sequence of protein with a causal specificity that matches that of genes and mRNAs. I see no way of how this could occur in the course of normal biological processes that allow a cell's protein synthesis machinery to continue to perform its full function (including the synthesis of other proteins).

Fourth. Why does the combination of biological normality and causal specificity matter? The reason is that large parts of biology are concerned with explaining how organisms manage to stay alive and how they can maintain the properties characteristic of life such as metabolism, growth, reproduction, and so on. Biological knowledge is not always "maker's knowledge" (Craver, in this volume), which is not to deny that "makers knowledge" exists also in biology. Biological knowledge is usually knowledge about the living state, its maintenance and evolvability. It is this explanatory focus which selects biologically normal interventions as the targets of relevant "what if things had been different"-questions (Woodward 2003). For biologically normal interventions are such changes that do actually occur in the living state and are ultimately responsible for its maintenance and its evolution. This is why they matter biologically.

Thus, the following picture of how biologists select explanatory causes in such cases emerges: First, list *all the causally relevant factors*. These include DNA, mRNA, splice agents, post-transcriptional and -translational editing machinery, RNA polymerases, ribosomal RNAs and proteins, tRNAs as well as aminoacyl-tRNA synthases. Second, select from this list those causes that can be actualized by *biologically normal interventions*. This removes tRNAs and aminoacyl-tRNA

synthases, but retains splicing agents and their like. Finally, from the remaining causes, pick those with the *highest degree of INF*. This leaves only DNA and mRNA, i.e., precisely those molecules that are said to be information-bearers. Typically, this selection procedure will pick out actual-difference makers in Waters's sense (because biologically normal variation is often at least partly actualized), but not exclusively so.

How does the present account differ from Waters's and Woodward's? While it is obviously indebted to both accounts, it also differs in important respects. The main difference to Waters's account is that this account allows us to say what is special about the causal role of genes even in cases where there is no actual variation. I see the causal uniqueness of genes not in their accounting for *actual* variation. This clearly is an important aspect of genes, but it is not *specific* to them. On my account, the special role of genes only becomes visible in the modal realm. This should be metaphysically acceptable to anyone who buys into a counterfactualist theory of causation.

My account differs from Woodward's in that it does not rely on INF alone for explicating the causal role of genes. As I have shown in Section 3, there are other cellular components--namely parts of the protein synthesis machinery--that also bear the relation INF to proteins. For INF is a *counterfactual* relation (by definition), so nothing prevents us from letting things like tRNAs or aa-tRNA synthetases vary in such a way as to making any arbitrary change in protein sequence possible (thus satisfying INF the same extent as genes or mRNA). My strategy is to exclude such variation by declaring the necessary counterfactuals *irrelevant* for biological explanation. This is justified. What biologists want to

explain is *life*. Therefore, they don't care about what might happen in some highly contrived possible worlds; they only care about what might happen under *biologically normal circumstances*, and the kinds of variation necessary to manipulate protein sequences in a highly specific way by tampering with anything else but genes or mRNA are not biologically normal circumstances.[20]

In the final section, I shall attempt to draw some more general conclusions from these considerations.

6. Conclusions: Causal Explanation and Biological Normality

The causal parity thesis derives its initial plausibility from the contemplation of the enormous causal complexity of the processes that constitute life at the molecular level. Somewhat ironically, we would know close to nothing about this complexity if it wasn't for gene-centered research, but this is besides the point here. I am concerned here with the attempt to reject the CPT on the grounds of causal considerations alone, neglecting the methodological and information-theoretic attempts. Like Waters and Woodward, I think that our causal concepts need to be refined in order to express the special role of genes and DNA in ontogeny in an abstract way, lest not all causes of ontogeny look the same.

Waters tried to do this with the help of a distinction between actual- and potential-difference making cause, Woodward with the help of the concept of causal specificity (INF). I have shown that neither approach is satisfactory. As Waters himself has noticed, his concept of (the) ADMC is not sufficient for the more complex cases (eukaryotic gene expression) and needs to be strengthened by

causal specificity. The problem with this turned out to be twofold: First, the account has counterexamples such as cells that are differentiated to make only one protein. Second, the concept of ADMC is sensitive to the choice of a population or reference class, which some find undesirable because the resulting distinctions don't seem to be ontological enough. On my account, we can have more if we make our causal concepts sufficiently rich.

The key to solving the causal selection problem lies not in actual difference making (although this is an important causal notion) but in the causal specificity of the *potential*-difference makers, i.e., in the counterfactual realm. This is not a radical move at all, given the broad consensus that the causal relation itself supervenes on counterfactual scenarios anyhow.

The trouble with this move--which is basically Woodward's--is that it allows other potential-difference makers to move into sight that are a perfect match to genes as far as causal specificity is concerned (tRNAs, aa-tRNA synthetases, and possibly others). My approach here is to block these potential-difference makers as being *irrelevant to biological explanation*. All causal explanation must make a choice of relevant and irrelevant counterfactual states of affair (we want to know why Willy robbed the bank rather than getting a job, not why he robbed the bank rather than the dairy). The counterfactuals required for claims about the specificity of other potential-difference making causes (other than genes) are irrelevant[21] because to make the antecedents of the corresponding counterfactuals true would require events that do not occur in biologically normal circumstances. But such normal circumstances typically provide the backdrop against which requests for causal explanation emerge in biology.

This result has interesting implications for the topic of causal explanation in biology in general. This topic is often treated as if causation in biology did not differ from causation in other areas, except, perhaps, for the fact that causal generalities in biology tend to be more fragile (i.e., less stable in the sense of Woodward 2010) than others, in particular if compared to the physical sciences. Specifically, there seems to be a widespread confidence that the topic can be dealt with independently of considerations concerning normal functioning or teleology (see, e.g., Strevens, in this volume). Of course, there are good reasons for this; not least the desire for ultimately reducing functions and teleology to strictly causal explanations (see Weber 2017b). However, the arguments presented here show that causal explanations at least in some significant cases are, as it were, infected by teleology. What this means is that the causal relations that biologists care for often concern states of affairs -- including counterfactual ones -- that have some sort of *functional significance* for the organism. The mechanisms of gene expression are clearly an example for this.

Acknowledgements

This paper went through many stages and was presented and discussed at several places, including the British Society for the Philosophy of Science 2010 meeting in Dublin, the IHPST Paris, the Department of Ecology and Evolution at the University of Lausanne, the DFG research group "What If?" at the University of Konstanz, and last but not least at the workshop "Causal Reasoning in Biology II" at the Minnesota Center for Philosophy of Science. I am grateful for lots of helpful suggestions, in particular to Ulrich Stegmann, Jim Woodward, Ken

Waters, Paul Griffiths, Frédérique Théry, Brian Leahy, Daniel Dohrn, Wolfgang Spohn, Daniel Sirtes, Alan Love, Chris Hitchcock, Laura Franklin-Hall, and David Danks.

Notes

[1] The problem of causal selection arises because most explanations in science and elsewhere are highly selective about the kinds of causes they consider to be explanatorily relevant. While the issue has been discussed in philosophy of law (e.g., Hart and Honoré 1959), philosophers of science have become interested in the subject only very recently (e.g., Waters 2007, Ross 2017).

[2] Frédérique Théry (pers. communication) complains at this stage that a change of subject has occurred *vis à vis* the causal parity thesis. The latter makes a claim about causal reasoning concerning the ontogeny of the whole phenotype, while we are now suddenly talking about protein synthesis. However, it seems to me that to play a special role in protein synthesis is sufficient (even though perhaps not necessary) for playing a special role in ontogeny, because protein synthesis is a constitutive part of ontogeny.

[3] Note that Waters's claim about causal specificity concerns the *potential* variation of the actual-difference makers, not the *actual* variation of the actual-difference makers.

[4] Recently, Griffiths et al. (2015) have used mathematical information theory in order to provide a quantitative measure of causal specificity. I do not count this among the information-theoretic approaches because, even though these authors use information theory, they are not trying to define a sense of "information" that applies exclusively to DNA.

[5] It is particularly important not to confuse actual or token causation as it is discussed in the general causation literature (e.g., Hitchcock and Knobe 2009) with Waters's notion of actual-difference making. The latter is not applicable to singular causes.

[6] I lack the space to present Woodward's theory in any detail here; the reader is referred to the original publications, in particular (Woodward 2003). Very briefly, Woodward analyzes causation in terms of patterns of counterfactual dependence between possible interventions and some effect variable. In other words, a statement such as "X causes Y" is taken to mean that, given appropriate conditions (which are spelled out explicitly in Woodward 2003), an intervention that would change the value of X would change the value of Y. What is important to understand about this account is that it is non-reductive, i.e., it does not aspire to define causal relations in non-causal terms. This is evident in the above formulation in the use of the term "change", which is still a causal notion. Woodward contends that it is not possible to state the truth conditions for the relevant counterfactuals without already using causal notions; this contention is a subject of much current debate in the philosophy of causation.

[7] Waters (2007) also mentions different RNA polymerases as ADMCs in eukaryotic gene expression. This case strikes me as being qualitatively different from the ADMCs mentioned in the text above, because what species of RNA polymerase is active in gene expression only affects the rate of transcription, not the sequence or identity of the resulting gene products. Thus, RNA polymerase could be ruled out as an ADMC by specifying that we are talking only about causes of primary structure, not of the rate of synthesis. However, the same job

can be done in one fell swoop by using the notion of causal specificity (see below).

[8] I am indebted to Chris Hitchcock for calling this difference and its potential importance to my attention.

[9] In an *injective* mapping (also: one-one), the elements of the codomain (= values of the dependent variable) are mapped to by *at most* one argument of the domain (= values of the independent variable). In a *surjective* mapping (also: onto), all the elements in the codomain are mapped to by *at least* one argument from the domain. Bijective mappings are both injective and surjective.

[10] In line with standard terminology in molecular biology, this term stands for all the processes that result in the synthesis of RNA (=transcription) and protein (= translation).

[11] I owe this point to Ulrich Stegmann.

[12] Griffiths et al. (2015) point out that the causal specificity of the actual variation in splice variants, by the lights of their information measure, may exceed that of actual DNA sequence variation in biologically relevant cases. In a discussion note to Griffiths et al., I argue that this is not the relevant kind of causal specificity that guides causal selection in this case (nor is it according to Waters). The relevant variation, i.e., the biologically normal potential variation, is far more causally specific in the case of DNA mutations than it is for alternative splicing (Weber 2017a).

[13] I think that Woodward's INF is better suited to analyzing biological theories than Sarkar's (2003) concept of "formal information system". Although the latter also has its merits and contains a similar idea about mappings of earlier to later states, it is also ridden with difficulties, the most serious one being that it fails to

capture the causal relevance of the input to the output states (see Stegmann 2009 for a full critique).

[14] tRNA is not traditionally assumed to be an information-bearing molecule; it is considered to be part of the protein synthesis machinery. Therefore, tRNA is a real challenge for the CPT opponent. The same is true for the enzymes that charge the tRNAs with amino acids, the aminoacyl-tRNA synthetases. They are thus also responsible for the specificity of the tRNA-codon relation.

[15] Here, I am indebted to David Danks.

[16] Note also that it might be difficult to define a measure for the magnitude of interventions. Furthermore, it is not entirely clear how we should count interventions.

[17] As Ulrich Stegmann pointed out to me.

[18] To my knowledge, so-called "epigenetic" modifications of the chromatin such as methylation or acetylation do not affect the sequence of the gene products, just the rates of their expression.

[19] I must obviously allow for a certain class of exceptions here, namely such cases where an intervention occurs on the DNA that introduces a lethal mutation in the protein coded. In such cases, we must construe biological normality itself as a counterfactual scenario and assume that the function wiped out by the mutation is somehow provided (or neutralized if its a gain of function mutation). This case is much closer to normal biological functioning than those scenarios where we change the genetic code after each ribosome cycle, which would completely incapacitate the cell's gene expression machinery.

[20] My intention here is not to introduce an Aristotelian dichotomy such as *physis* versus *techne* or things that "exist by nature" and things that "exist from other

causes" (as Aristotle's *Physics*, Book II famously begins). A technical intervention (where it is possible) is not necessarily biologically abnormal, so long as the same result could also have been produced by a normal biological process. This happens all the time in genetic engineering. Nonetheless, perhaps the biological normality constraint on causal selection does introduce a whiff of Aristotelianism into the theory of biological explanation.

[21] Note that I am not claiming that all counterfactuals about these other factors are irrelevant; of course, some counterfactuals are needed to express their causal relevance (which I am not denying). My point is that those counterfactuals that pertain to the *causal specificity* (INF) of these other causes are biologically irrelevant.

References

Bigelow, J. and R. Pargetter (1987), Functions. *Journal of Philosophy* 84: 181-196.

Craver, C. (forthcoming), Intervention in Biological Experiments: The Case of Optogenetics. In this volume.

Cummins, R. (1975), Functional Analysis. *Journal of Philosophy* 72: 741-765.

Dawkins, R. (1976), *The Selfish Gene*. Oxford: Oxford University Press.

Franklin-Hall, L. (forthcoming), The Causal Economy Account of Scientific Explanation. In this volume.

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

Griffiths, P.E. and R. D. Gray (2005), Three Ways to Misunderstand Developmental Systems Theory. *Biology and Philosophy* 20 (2): 417-425.

———, A. Pocheville, B. Calcott, K. Stotz, H. Kim and R. Knight. 2015. "Measuring Causal Specificity", *Philosophy of Science* 82: 529-555.

Hart, H.L.A., and A.M. Honoré (1959), *Causation in the Law*. Clarendon, Oxford.

Hitchcock, C. and J. Knobe (2009), Cause and Norm. *The Journal of Philosophy* CVI: 587-612.

Jantzen, B. and D. Danks (2008), Biological Codes and Topological Causation. *Philosophy of Science* 75: 259-277.

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

Mackie, J.L. (1980), *The cement of the universe. A study of causation*. Oxford University Press, Oxford

Maynard Smith, J. (2000), The concept of information in biology. *Philosophy of Science* 67:177-194

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

Oyama, S., P. E. Griffiths and R. D. Gray (eds.) 2001. *Cycles of Contigency: Developmental Systems and Evolution*. Cambridge, MA: MIT Press.

Ross, L. (forthcoming), Causal Control: A Rationale for Causal Selection. In this volume.

Sarkar, S. (2003), Genes encode information for phenotypic traits. In: C. Hitchcock (ed.), *Contemporary debates in philosophy of science*. Blackwell, London, pp 259–272

Schaffner, K. F. (1998), Genes, Behavior, and Developmental Emergentism: One Process, Indivisible? *Philosophy of Science* 65: 209-252.

Shea, N. (2007), "Representation in the Genome, and in Other Inheritance Systems", *Biology and Philosophy* 22:313-331.

Sober, E. (1993), *Philosophy of Biology*. Oxford: Oxford University Press.

Stegmann, U.E. (2005), Genetic Information as Instructional Content. *Philosophy of Science* 72: 425-443

Stegmann, U.E. (2009), DNA, Inference, and Information. *British Journal for the Philosophy of Science* 60: 1-17

Stegmann, U.E. (2012a), Varieties of Parity. *Biology and Philosophy* 27: 903-918

Stegmann, U.E. (2012b), Causal Control and Genetic Causation. *Noûs* 48: 450-465

Sterelny, K., K. Smith and M. Dickison (1996), The Extended Replicator. *Biology and Philosophy* 11: 377-403.

Strevens, M. (forthcoming), The Causal Role of Aggregate Properties. In this volume.

Waters, C.K. (2006), A Pluralist Interpretation of Gene-Centered Biology. In S. Kellert, H. E. Longino and C. K. Waters (eds.), *Scientific Pluralism. Minnesota Studies in Philosophy of Science, Volume XIX*. Minneapolis: University of Minnesota Press, p. 190-214.

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

Weber, M. (2005), Genes, Causation and Intentionality. *Studies in History and Philosophy of the Life Science* 27: 399-411.

Weber, M. (2006), The "Central dogma of molecular biology" as a thesis of causal specificity. *History and Philosophy of the Life Science*s 28:565-580

Weber, M. (2017a), Which Kind of Causal Specificity Matters Biologically? *Philosophy of Science* 84: 574-585.

Weber, M. (2017b), How Objective Are Biological Functions? *Synthese,* doi:10.1007/s11229-017-1483-z

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

Woodward, J. (2010) Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy* 25:287-318