Disentangling causation and information: informational parity at issue

[draft]

Abstract

The notion of an informational parity between genes and non-genetic factors appears in two ways in the literature. On the one hand, it is claimed to follow from an information-theoretic approach to account for the notion of information in biology. This consequence, as we shall explain, is considered to be unacceptable for some authors which, therefore, took a different approach in order to save the informational exclusiveness of genes. On the other hand, informational parity is one of the many versions of the causal parity thesis, according to which genes and other developmental factors are causally on a par. According to this view, causal parity is an actual feature of living systems and the concept of information needs to be congruent to this fact. We will argue that in both cases there is a deep conflation between the concepts of information and causation (as concepts undisputedly related) that has not been sufficiently addressed, especially with respect to the quarrel over parity. Such a conflation has a twofold origin: (i) a rough understanding of causation and (ii) a misreading of information theory.

1. Introduction

Lately, an issue has arisen of whether genes and non-genetic biological factors are "on a par". The so-called Developmental Systems Theory (hereafter, DST) has supported parity claims in accordance with its general aim to remove the focus put on the role of genes in development and to emphasize the role of what they refer to as *other developmental resources* of an organism (Griffiths & Gray 1994). Their maxim is that genes are but *one* among the several resources that organisms rely on in order to develop, all of which are "on a par" in this respect, an idea expressed by the so-called "causal parity thesis" (CPT, hereafter). A gene, in short, is not a special resource nor is its role more important than that of other resources.

A closely related idea that features in the literature is that of an informational parity between genes and non-genetic factors, which appears in two ways. One hand, informational parity is a consequence that follows from attempting to account for the

notion of information in biology by appealing to information theory. This consequence is considered to be unacceptable for some authors which, therefore, giving this and other issues, take a different approach in order to save the informational exclusiveness of genes. On the other hand, informational parity is one of the many forms that CPT adopts and one that expresses an actual feature of biological systems. According to this view, then, the idea of information must pick out the fact that genes and other factors are causally on a par.

In this paper, we will scrutinize the sort of reasoning in both cases to argue that it is grounded on a deep conflation between the concepts of information and causation, in the sense that these two concepts seem to be undisputedly related. Such an association has not been sufficiently addressed, especially, with respect to *the quarrel over parity*. We will show this this underlying conflation to have a twofold origin: (i) a rough understanding of causation, and (ii) a misreading of information theory. For the first point, we will argue that if more fine-grained causal analysis can be used to refute CPT, then this drags along also any idea of informational parity that relies on causal considerations. For the second point, we will argue that information theory does not entail any causal notion of information per se. Therefore, informational parity is never correctly inferred -not when endorsed as a version of CPT, neither when rejected as an inacceptable consequence of information theory.

The paper is structured as follows. In Section 2 we introduce and explain CPT within the framework of DST. Here, we show informational parity to be one of the different versions of CPT. The issue with this endorsement of informational parity is put forward in Section 3. Next, in section 4, we track the idea of informational parity in the ongoing debate on the concept of information. There, we show that informational parity is claimed to be a consequence of an information-theoretic approach. The problems with this origin of informational parity are exposed In Section 5. Further questions and issues are suggested as final remarks in Section 6.

2. Developmental Systems Theory, the causal parity thesis and information

CPT originally arose within the theoretical standpoint of DST, a broad approach towards the understanding of development and evolution. Thus, in the next section, we will fist take a look at DTS and its main theoretical proposals, which will allow

understanding the role played by the CPT thesis in their conceptual framework. But advocators of CPT actually sustain a number of different versions of this thesis, so in the following subsections we expose the different shades of CPT, the last of which is the informational version.

2.1. What is Developmental Systems Theory?

As it is commonly presented, DST is a general approach on development, evolution and inheritance, but this is not one unified and fix set of claims (Pradeu 2010; Stegmann 2012b). Indeed, DST consists of several and diverse positions with at least some common ground: a certain consensus on the need to challenge the gene-centered and dichotomic view of development and evolution, received from the Modern Synthesis (Merlin 2010; Griffiths & Gray 1994). Against an oversimplified and sometimes reductionist approach, DST emphasizes the systemic and dynamical nature of biological phenomena. DST theorists also address, from their general view, some already classical topics in theoretical biology and philosophy of biology, as are the nature/nurture debate and the replicators/interactors distinction, among others.

In order to accomplish their particular approach on these matters, DST theorists have redefined their object of study: a *developmental system*. According to Paul Griffiths and Russell Gray, "the developmental system consists of the resources that produce the developmental outcome that are stably replicated in that linage" (1994: 278). Although here we are not concerned with the problem of individuality in biology, the definition shows a feature of DST relevant to the purpose of this paper: *resource* is a key concept in DST approach. In DST, the organism is conceived as a set of different developmental resources. This is not at odds with the received view of development; however, it is the postulation of a strict distinction between two *kinds* of resources -genetic and other- what DST disagrees on (Griffiths & Gray 1994; Merlin 2010; see also Sterelny *et al*. 1996). More specifically, in the received view, the different environmental features constitute the mere background upon which genes "rule" the developmental process (Griffiths & Gray 1994: 281). DST theorists reject this idea categorically: genes are just one among many resources available to the ontogenetic process. In Griffiths and Gray's early words on the matter:

> "There is a fundamental symmetry between the role of the genes and that of the maternal cytoplasm, or of childhood exposure to language. The full range of developmental resources represents a complex system that is replicated in development. There is much to be said about the different roles of particular resources. But there is nothing that divides the resources into two fundamental kinds. The role of genes is no more unique than the role of many other factors". (1994: 277)

This is what the CPT asserts in a rough characterization. However, two things require clarification: what is the precise content of the thesis itself, on the one hand, and how does the parity thesis become an informational parity thesis, on the other. Both will be addressed in the next subsections.

2.2. Different shades of the causal parity thesis

What is CPT exactly about? This has proved to be a complicated matter, for several definitions can be found in the DST literature. It can be tracked back to Susan Oyama's *The Ontogeny of Information* (1985):

> ''What I am arguing for here is a view of *causality* that gives formative weight to all necessary influences, since none alone is sufficient for the phenomenon or for any of its properties''. (1985: 15, emphasis is added).

Here, we can see why the thesis is often referred to as a *causal* one. But this version of CPT thesis appears to be quite harmless, for even mainstream biology is aware that genes alone, without all the cellular machinery and several biological agents, could not give rise to their molecular products nor, ultimately, contribute to development. This happens to be the case for any biological agent, being genes, enzymes, organelles. It seems to us that Oyama confuses *plurality* with *parity*. Actually, all we can conclude from the acknowledgement that genes do not act alone is that there is a causal plurality, as in a large number of various causal factors, rather than a causal parity among them. But a complex causal structure, composed by a multiplicity of causal factors, is not necessarily one in which those causal factors are on a par. Plurality-assertions are

quantitative, whereas parity-assertions are qualitative; additional premises are required to move from the former to the latter.

On the other hand, nobody would seriously challenge the claim that genes alone are not sufficient for development. So, for CPT not to be a truism, it must mean something stronger (Stegmann 2012). However, a stronger sense can be conflictive. Ulrich Stegmann (2012b) has done a great job clarifying several versions of CPT spread in the DST literature, showing to what extent do authors stand up for very different claims.

Thus, for example, a "Millean" version of parity rests on the idea that all causes constitute a single kind, and draws on Mill's view that causes and mere background conditions present no ontological differences. This version asserts that genetic and non-genetic factors are on a par insofar as they are both causes. This version of parity can be endorsed with a pragmatic element, as Lisa Gannett (1999) has suggested. She claims that, in the scientific practice, the selection of "the" cause in a given context depends on pragmatic interests rather than on theoretical considerations. But arguing from the perspective that all causes are, as such, ontologically the same, may settle things too early for the problem of causation in biology (or too early for a philosopher anyway). Drawing various distinctions among different causal factors, contributions and relations is actually an important task within the philosophical problem of causation (we will return to this in Section 4).

In any case, the most relevant versions of CPT for the purposes of this paper are those advocated by Griffiths, Gray and Robin Knight (Griffiths & Gray 1994; 2005; Griffiths & Knight 1998), namely, the "no dichotomies-view" and the "distributive" version. The former stems from their insistence on rejecting any (unjustified, according to them) dichotomic view of biological phenomena. In this case, the parity means that genetic and non-genetic factors do not belong to two different *kinds* of factors that play any fundamentally distinct role in development. But more specifically, what this thesis denies is that genes and other factors play roles that differ *metaphysically*. The metaphysical difference alluded comes in a few variants according to Stegmann (2012b): genes as information carriers, replicators and controllers, on the one hand, and non-genetic factors as material, interactors and controlled by genes, on the other. Griffiths and Knight put the challenge they assume in this way:

> "The real developmentalist position is that the empirical differences between the role of DNA and that of cytoplasmic gradients or host-imprinting events do not justify the metaphysical distinctions currently built upon them". (Griffiths & Knight 1998: 254)

> "It has never been part of the developmental systems tradition to deny that nucleic acids and natural language are distinctive elements of developmental systems. The point of the 'parity thesis' is to prevent these empirical differences turning into a kind of scientific metaphysics, as happens when genes are identified with information (or even 'form') and everything else in development with mere matter" (Griffiths 2001: 406).

The distributive version of CPT, in turn, claims that which genes and non-genetic factors are on a par insofar every contribution to development made by a gene is also made by some other factor, and vice versa:

> ''DNA does play a distinctive set of roles in development, but it does not play just one role (partly because DNA elements are themselves so diverse) and the important roles those various DNA elements plays [*sic*] are sometimes played by non-DNA factors in development'' (Griffiths and Gray, 2005: 421).

Karola Stotz (2006) develops this point of view and suggests which contributions genes do share with (at least) another non-genetic factor. For instance, sequence specificity is shared with splicing and editing agents, inheritance is shared with methylation patterns, gene regulation is shared with protein transcription factors and with environmental factors. While the last version rejects major metaphysical distinctions among genes and other resources, this one denies the exclusiveness of any agent in the making of its contribution to development. A particular contribution is not made exclusively by genes nor by any other non-genetic factor.

Other aspects of these versions of CPT can be explored further, but that we are interested in here is how the parity thesis is linked to the notion of information, or how the former "turns" informational.

2.3. The informational version of the causal parity thesis and Developmental Systems Theory on information

DST's attitude towards the informational talk in biology has not been entirely categorical. On the one hand, it seems that they are seriously suspicious about this notion (Pradeu 2010). Oyama, for example, expresses her concern that informational notions might serve in feeding genetic determinism, preformationism or gene-centrism. Thus, she claims that the idea of genetic information is the modern source of the old idea of 'form': everything that an organism requires for its construction is provided by the genome, whereas every other developmental resource provides the mere "material" for the construction. In her words, "[I]nformation is conceived to be a special kind of cause among all the factors that may be necessary for a phenomenon, the cause that imparts order and form to matter". (Oyama, 1985: 3). Relatedly, it has been argued that the idea of genetic information might be at odds with CPT, in the sense that it leads to ascribe a causal primacy to genes that overlooks the fact that other kind of biological factors happen to be essential to the genetic processes and in development in general (Oyama 1985, 2000).

However, the informational talk has not been entirely or straightforwardly rejected either. Rather, DST theorists have hesitated over what the fate of the informational talk should be. And philosophically, they have two options: either giving up the informational talk of genes once and for all, because it is irremediably misleading (in the sense of the objections above), or keeping on the condition of conceiving biological information as something distributed along the whole developmental system in which it gains meaning (Oyama 1985). This second option amounts to broaden the notion of information and claim that containing information is not a purview of genes alone and, thus, leads to endorse the *informational version of CPT*.

So, as mentioned before (2.1), the second problem regarding the causal parity (and the crucial one in this paper) is the relationship between CPT and the notion of information. The informational parity (Griffiths & Gray 1994; Griffiths & Knight 1998; Sterelny & Griffiths 1999; Griffiths 2001) could be thought to be a case of the no dichotomies-view just exposed in the sense that it rejects a case of dichotomy, namely, between information carriers and not information carriers (Stegmann 2012b). The most explicit claims in this sense are due to Sterelny and Griffiths:

> "Developmental systems theorists argue that in any sense in which genes carry developmental information, nongenetic developmental factors carry information too" (Sterelny and Griffiths 1999, p. 101).

> "Any defensible definition of information in developmental biology is equally applicable to genetic and non-genetic factors in development" (Griffiths 2001, p. 396).

The last claim is in conditional terms as a cautious attitude towards finding an appropriate account of the concept of information (which is the topic of Section 4).

3. Conflation, part I: A rough conception of causation

As explained in the previous section, informational parity is one of the many versions of CPT that expresses an actual feature of biological systems. In this view, the notion of information is only legitimate if it picks out the fact that genes and other factors are on a par. Thus, a satisfactory account of biological information that makes sense of the idea that genes carry information should also be able to pick this fact out, making sense of non-genetic factors as information carriers as well.

The problem with such an endorsement of a thesis of informational parity is not only that it relies on causal considerations in general, but more specifically, that it relies in too rough a notion of causation. To see this point, let us first assume, for the sake of the argument, that the relationship between causation and information is not troublesome. Still, the we face the situation that DST does not appeal to any technical-philosophical account of causation to support CPT, so it is not a surprise that interesting differences among causal factors go unnoticed. Yet, as has been remarked by some philosophers, one of the main tasks regarding causation is to refine causal concepts and causal properties in order to draw more subtle distinctions among causal relations, and not *merely* to elucidate what it is to be a cause in general. Therefore, the possibility of a causal parity among a number of different factors with respect to a certain outcome needs to be examined from the point of view of said fine-grained causal concepts.

Building on Woodward's interventionist theory of causation (2003), Kenneth Waters (2007) has argued against CPT. From an examination of biological practice, he

concludes that a distinction can be drawn between two kinds of causal agents: *potential difference makers* and *actual difference makers*. A potential difference maker is a causal variable that satisfies the counterfactual patterns in the sense of the interventionist theory, regardless whether it actually varies and whether this variation brings about actual differences. An actual difference maker, on the other hand, is such that actually varies and actually brings about a difference. Moreover, *an* actual difference maker can be distinguished from *the* actual difference maker in cases in which there are several difference makers. These notions are relative to a given population. Waters argues that (variation) in DNA and mRNA sequence in prokaryotic cells is the only actual difference maker with respect to protein synthesis (i. e., the population of proteins in a cell), the rest of the elements in the developmental matrix being only potential. But even when he acknowledges that in the case of eukaryotic cells there may be actual difference makers other than DNA (e. g., RNA polymerase in alternative splicing), this does not imply that DNA be on a par with other factors because DNA and mRNA, additionally, causally specific. Specificity holds whenever counterfactual dependencies between variations in an event C and variations in an outcome E are such that there is a (close enough) bijective mapping from the set of C-states into the set of E-states. Thus, specificity singles DNA and mRNA out among the many actual difference makers.

In a paper devoted to biological causation, Woodward (2010) works out three causal notions that are biologically relevant and that can be used to make causal distinctions: *stability* (the insensitiveness of the causal relationship to changes in background conditions), *proportionality* (the level of detail, appropriate or inappropriate, in the causal description); and *specificity* (which refers to the possibility of obtaining fine-grained variations in the effect by means of altering the state of the cause). He refers briefly to CPT:

> "This argument [of causal parity] overlooks the possibility that even if C1 and C2 are both causally relevant to E, the causal relationship between C1 and E may nonetheless differ from the causal relationship between C2 and E in virtue of *other* features (…). For example, (…) the relationship between C1 and E may be more specific, more stable or better satisfy the requirements of proportionality that the relationship between C2 and E, thus introducing an asymmetry between the two factors." (2010: 316)

Thus, Woodward suggests that DNA is more specific than other relevant causes (as splicing agents or the various enzymes that participate in post-translational modification), although the notion of specificity is somewhat different from Water's and refers to the possible range of values for a variable (not to the actual variability in a population, see the full definition in Woodward 2010, p. 305) and admits of degrees in two senses: depending on the number of states that are correlated by the mapping, and depending on the closeness of such mapping to a bijection (Weber forthcoming). Thus, it is the higher degree of specificity which singles DNA out among other (less) specific causes.

In Marcel Weber's view (forthcoming), however, neither Woodward's nor Waters' accounts are enough to refute CPT and single DNA out among causes[1]. (because it entails counterintuitive consequences, because other molecules, such as tRNA, satisfy the specificity criteria, respectively). Instead, he argues, we need to take into account the fact that not all counterfactuals are *equally relevant*. Specifically, those counterfactuals whose antecedents describe normal interventions are more relevant that those which don't. 'Biologically normal' here means that the intervention (i) may also be due to natural processes and (ii) is compatible with the continued persistence of the biological entity under scrutiny. Weber claims that causal selection is often settled by choosing relevant counterfactuals in the above sense. Thus, he states that "among those potential difference-making causes of protein sequence that can be actualized by biologically normal interventions, genes, DNA and mRNA are the causally most specific" (forthcoming, p.?). Conceivable interventions on tRNAs fail to meet the conditions for biological normality.

Genetic causation has also been addressed from the point of view of the idea of *control* by Stegmann (2012a). Given a set of effects <F, G, H>, there are two ways in which it can be brought about, that is, by an external or by an internal ordering. In the first case, F, G and H depend on different variables, say, P, Q and R. In an internal ordering, on the other hand, only one variable of the set depends on an additional variable, P→F, while the rest of them depend on F in chain: F→G→H. The external ordering causal

[1] The arguments include counterintuitive consequences regarding single (token) events of protein synthesis, inability to account for cases where there is no actual difference making (no actual sequence variation), the fact that other molecules, such as tRNA, can be argued to be as specific as DNA (see full paper for details).

structure corresponds to what we usually mean when we say that certain entities control the processes they cause. Stegmann, thus, argues that a single general concept of causation does not capture DNA peculiarities, but the concept of external ordering does: DNA templates are not only specific (again, in a Lewis-Woodward sense) but they also externally order their product molecules, whereas other causal factors in genetic processes are not of this kind.

Yet another recent distinction meant to refute CPT is that between *causal relevance* and *causal responsibility* and is Christopher Austin's (2015). A factor is causally relevant if its values are functionally correlated to the effect values, but is causally responsible if it is responsible for the very existence of such value correlations. DNA, when considered as realizing dispositional properties directed towards phenotypic outcomes, is causally responsible for those outcomes. And since other factors are not causally responsible for the same outcomes, there is no parity of responsibility, and *causal primacy* can be ascribed to DNA.

Summing up, even if it is true that both genes and non-genetic factors are causally relevant to development, several philosophical treatments of causation challenged CPT. CPT appears to make sense only under a very rough and undefined understanding of causation and can be challenged with a more fine-grained approach. The point we want to make here is that these accounts of genetic causation show that, even if we avoid questioning the very relationship between information and causation, as CPT does not seem to be solid enough, it could hardly ground other sorts of parity claims, particularly, informational parity claims.

However, one could think that the above-mentioned finest causal accounts are unconvincing or indecisive. For the sake of the argument, let us make a concession. Let us admit that even finest causal accounts allow to support CPT, so that CPT can be refined and saved. Still, we are entitled to ask why should we assume a direct relationship between causation and information. Even if a causal parity can be defended, we still lack a justification or explanation of the shift from mere causal parity to informational parity (which requires first a clear concept of information).

Consequently, even if we chose to ignore the issues CPT or the issue regarding the link between causation and information, the result is the same: no informational parity can be inferred or grounded in CPT.

4. Parity in the debate on information in biology

While some DST theorists have explicitly supported the thesis that genes and non-genetic are informationally (and causally) on a par, the idea of an informational parity has also been claimed to arise as a *consequence* of certain way of conceiving information. In order to show this, we will first introduce the problem of the concept of information in biology as one concerned with the elucidation of such a notion. Then, we will take a look at the information-theoretic approach to the problem and the objections against it, in which the idea of parity comes in. Finally, we will give a quick overview of alternative attempts to accounts for the concept of information which have parity as one of their main concerns.

4.1. The main problem around the concept of information

As it has been remarked, "information is everywhere", from our everyday language to several scientific discourses. Biology is no exception: the talk of information is widespread within the life sciences. Multiple areas in biology exhibit this informational language: cell biology, ecology, behavior biology, developmental biology, evolutionary biology. But the concept of information appears somewhat drastically in the context of Molecular Biology. The idea that *genes carry information,* that *this information is inherited*, and that genes contain pretty much all it takes to build up a living organism is incredibly spread. The concept of information seems to be endowed with a differential privilege in the case of the molecular context. It seems to be much constitutive of the field that it is in the rest of the cases. Indeed, the concept of information has been employed since the very beginning of the field, and it appears in the most important theoretical formulations. In his famous 1958 article, "On Protein Synthesis", Francis Crick used the term 'information' to refer to the specificity of gene action in the determination of the lineal order of the components of other macromolecules. Ever since, the concept of information became part of the standard scientific vocabulary and the regular way to understand genetic processes.

Nevertheless, nobody is too sure of *exactly* what 'information' is supposed to mean in this context. This is so mainly because we lack anything like "a theory of biological information". But is also due to the fact that, when we talk of genetic information, we are no referring to a set of data used by external agents in cognitive processes (i.e., a researcher making inferences from data). Instead, it appears to be something "used by the system", that is, the cell. This peculiar aspect of the notion of information in biology has made some people suspicious about it, raising the philosophical debate we will summarize in the rest of this section.

4.2. Information theory, issues and informational parity consequences

If one aims at accounting for the concept of information in biology, it is natural to turn to pre-existing theories of information. The most popular and successful theory of information, Shannon's information theory (1948), has been invoked for this purpose. One of the major insights of the theory is that the channel capacity and the amount of information transferred can be measured, quantified. Shannon modeled a *general communication system* that consists of a relation between a source *S* that produces the messages, a channel *CH* that serves as a medium for the transmission of the signal, and a destination *D* to whom/what the message is directed to.

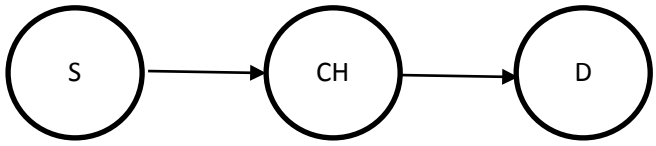

Figure 1: A simplified model of a communication system: a source, a channel and a destination.

Communication, then, requires at least a source, a channel and a destination. S and D must have a range of possible states with a given probability of occurrence. There is communication between S and D when states of S are correlated to states of D and the amount of information can be calculated with the equation $H = -\sum_{i=1}^{n} p_i \log p_i$.

The idea here is that DNA can be seen as a sort of source, while phenotypic or developmental features can be treated as a sort of destination:

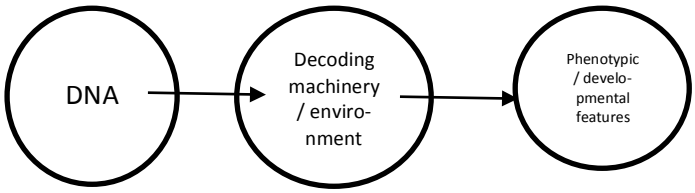

Figure 2: The simplified model of a communication system instantiated biologically.

Typically, DNA is designated as the source of information. Environment is designated as the cannel by which genetic information is transmitted, where 'environment' includes both the extrinsic environment of an organism (e.g. food availability, sunlight hours, etc.) and the cellular environment in which genetic processes take place (the cell's "decoding machinery": cytoplasm, organelles, specific enzymes and other molecules, etc.)[2]. The channel is held constant as a sort of background conditions that allow focusing on the correlations between variations in the source with variations in the destination. The case with the receiver is a bit more ambiguous. To some, *R* must be interpreted as the phenotypic (observable) trait of an organism. To others, the genotype-phenotype relation is too complicated and it is best to interpret it as the most proximate gene products, that is, proteins, RNAs and other DNA molecules[3].

As some authors note, attempts to interpret the concept of biological information in information-theoretic terms face serious issues. Mainly, because (*if* righteously applicable) it cannot be applied in any way that is *interesting* from a biological viewpoint. Some of the arguments against appealing to information theory are the following:

- Information in Shannon's sense has nothing to do with the specificity alluded by Crick in the *central dogma* and fails to capture such a critical property of genetic processes (Sarkar 2004).
- The claim that genes carry information in Shannon's sense only means that there is some correlations between genes and their products, and we certainly did not need Shannon's theory to make this point. Relatedly, information in this sense is "everywhere" and "very little is being claimed" when we say that genes carry information for developmental outcomes in this sense (Godfrey-Smith 2004).

[2] This makes no difference to the point we are about to develop, as long as anything falling under this label is not DNA or genes.

[3] This makes no difference to the point, either.

– Shannon's theory is a quantitative approach of information and, hence, insensitive to *meaning* or *function*, which are relevant biologically speaking (Jablonka, 2002). (To some people, the case of genetic information requires an approach with at least some semantic character).

Yet there is a fourth popular argument that is particularly relevant for our purposes, because it brings a form of parity into the debate. This argument states the following: Information theory requires co-variation between source and destination. This requirement is met by genes, but is also met by many other non-genetic or environmental factors because they have reliable effects in development as well: "any of the causal factors that are necessary for the production of proteins can be treated as 'channels' against which variation in some particular causal factor can causally co-vary with protein production" (Austin 2015, p.?). So, one can always treat DNA as a source (just as in Fig. 2), but one might just as well hold the genetic factors constant, treat the environment (which exhibits variation as well) as the source, and find correlations between environmental factors and phenotypic or developmental outcomes (Fig. 3). Thus, Maynard Smith claims 'we can equally well say that a baby's environment carries information about its growth; if it is malnourished, it will be underweight'' (2000, p. 189). The important point is that this situation is claimed to show that the theory cannot be used to single out genes, on the one hand, and to rule out non-genetic factors, on the other, as sources of information. If environment is designated as the source (which nothing prevent us from doing), then it follows that environment carries information just as much as genes do (Maynard Smith 2000, Sarkar 2004; Godfrey-Smith 2004; 2008; Godfrey-Smith & Sterelny 2016; Austin 2015). In other words, information theory implies a parity of information between genes and non-genetic factors.

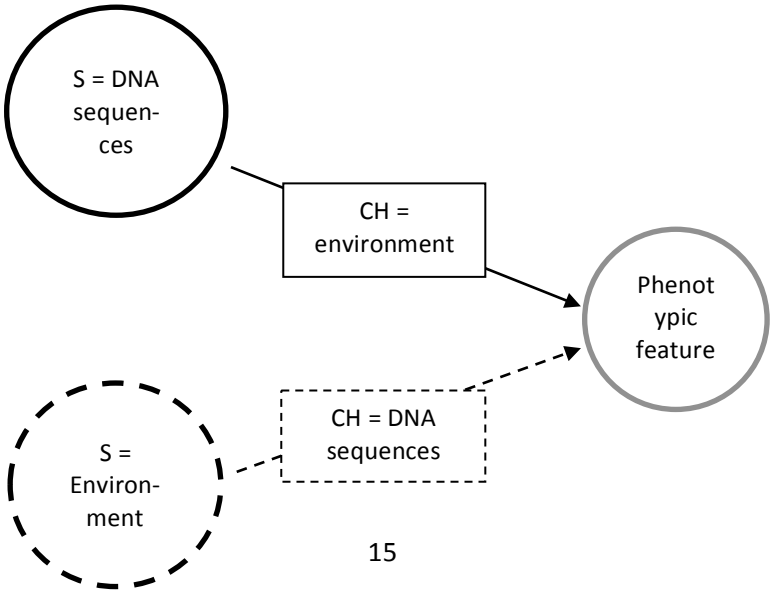

Figure 3: The punctuated lines represent a possible but less common instantiation of the simplified schema, which is an inversion of the roles of DNA and environmental or non-genetic features in the common view (solid lines).

But if it is a well-known fact that, indeed, there are many non-genetic factors-phenotype correlations (what no biologist would ever deny), then what is the problem with such a theoretical permissiveness? Doesn't it make justice to the actual sort of correlations found in living systems? If this is the case, this feature of information theory would be a virtue rather than a problem when applied to biological contexts. However, many authors claim the informational descriptions of genes and genetic processes have been traditionally used by biologists in order to highlight certain peculiarities of their roles in the organism. If one aims at elucidating or philosophically sharpening the notion of information *while favoring this orthodoxy*, then information theory will not do.

4.3. Alternative concepts of information and informational parity

For the different reasons mentioned above, several philosophers and biologists advocate for a more qualitative approach instead, and have put forward a number of different accounts that are (allegedly) better prepared to make sense of the informational talk in biology. In these alternative accounts, the notion of information is related to various other concepts and theoretical frameworks, such as teleosemantics and selected functions (Maynard Smith 2000a, 2000b, Jablonka 2002; Shea 2007, 2011a, 2011b), inheritance (Jablonka 2002, 2005, Bergstrom & Rosvall 2011), coding and causation (Godfrey-Smith 2000, 2004; Sarkar 2000, 2004; Kjosavik 2007), the new mechanistic philosophy (Darden 2006; Bogen & Machamer 2011; Kjosavik 2013), naturalized representations (Shea 2007, 2011a, 2011b), or naturalized instructions (Stegmann 2005), signaling games (Planer 2014, 2016; Calcott 2014), among others.

In many of these various accounts, parity is a common and explicit concern, and in some of these the aim is to elaborate an account that succeeds in making sense of genes as information carriers while singling them out as such. Maynard Smith notes the need for a concept of information that succeeds in capturing the traditional idea that only

genes carry information, because in fact biologists have use the idea of information to distinguish the role of genes in development from the role of non-genetic factors. He argues that only an intentional sense of information (and not a sense related to information theory) is able to rule out non-genetic factors as information carriers (although he is forced to admit, later, that his intentional criteria are also met by regulatory proteins and, therefore, these too can be said to be informational). In a similar vein, Stegmann defends the need to preserve the "paradigmatic cases" of the informational talk in biology, and he puts forward a notion of 'instructional content' that applies only to genes. Sarkar, in turn, argues that even when non-genetic factors, say, proteins, are relevant to explain a certain trait, this by itself does not mean that they carry information about the trait, because the relationship between proteins and traits fails to meet his criteria for information ('semiotic information', as he labels his notion), namely, arbitrariness and specificity. Similarly, for Godfrey-Smith, the only causal relationship of a coding nature (featuring arbitrariness, specificity, and combinatoriality) is that of nucleic acids and codons, whereas environmental factors are ruled out. Kjosavik, drawing on the mechanistic approach, argues that genes and non-genetic factors are on a par when considered in their regulatory role (because entities other than genes take part in regulation), genes keep a primacy over other factors when considered in their sequence-specifying or "structuring power" role (that is, their role in synthesis of other molecules).

But not all of these accounts aim at preventing informational parity consequences from following. Shea has a rather moderate stance: while he rejects a too-extreme idea of parity (according to which every causal factor in development can be argued to carry information) he does not restrict his account to genes. Other (but not every) non-genetic factors may convey representations as well (most likely, epigenetic factors), although this possibility is constrained by the requirement that the candidate factor be part of an inheritance system. Furthermore, in Jablonka's account of information, a form of parity between genetic and non-genetic factor is explicitly (though somewhat obscurely) supported. Specifically, she argues that carrying information is not a property of genes alone and that we need an unbiased concept of information that allows the acknowledgement that there are different types of information. In her view, transmission of information is involved in inheritance whatever its vehicle is, so for example, methylation patterns and heritable behaviors are also information bearers.

5. Conflation, part II: A mistaken view of information theory

Informational parity as a consequence derives from choosing a particular concept of information, specifically, it is typically brought up when discussing a correlational concept of information, such as Shannon's. In Section 4.2 we explained that in some philosophers' opinions, from the framework of information theory we can compute either the environment or the genome as a source of biological information, because they are both correlated to phenotypic outcomes. This consequence is unacceptable for some authors and is therefore a reason to reject such an approach to the problem of information in philosophy of biology.

What is important to stress here is that such correlations are actually and very often conceived in explicit *causal* terms (Godfrey-Smith 2008). So, the reasoning can be put in this way: Information (according to information theory) requires causal relations. Both genetic and non-genetic factors can be proved to be causally linked to developmental/phenotypic outcomes. Therefore, both genetic and non-genetic factors carry Shannon information about such outcomes. Opponents of informational parity, then, conclude that if one is unwilling to accept that genetic and non-genetic factors are *informationally* on a par, one should turn to a different approach elucidate the notion of information in biology. This conceptual or non-empirical origin of the informational parity (as a consequence of endorsing a specific theoretical framework) has barely been noted (though Stegmann 2012 does). What we now want to argue is that, *even if rejected*, informational parity is not correctly inferred.

The most important feature of Shannon's information theory is often overlooked: it is a formal, quantitative theory that offers a purely formal, abstract definition of central terms (for instance, 'channel' and, more importantly, 'information'!). Thus, it has been argued that the theory requires an interpretation in order to be applicable to certain domains. And in fact, it has been given several quite different interpretations (epistemic, physical, deflationary, see Lombardi 2004, Lombardi et al. 2015). Despite this fact, in the debate on genetic information, Shannon's information is simply *assumed* to provide a *causal* information concept (Griffiths 2001; Sterelny & Griffiths 1999; Downes 2006; Godfrey-Smith & Sterelny 2016). For instance,

> "There are essentially two concepts of information, which we can label "causal" and "intentional". Causal notions of information derive from the mathematical theory of communication, the discipline originally invented to design efficient telephone systems in the 1940s (Shannon and Weaver 1949)." (Sterelny & Griffiths 1999, p. 101).

But does the theory actually entail a *causal* concept of information? Although a causal view of Shannon's theory could be plausible[4], it cannot be something taken for granted. For example, the theory can be presented syntactically, without referring to anything like signals, sources or receivers, but only to random variables, probability distributions over their possible values, and the correlations between them (Cover &Thomas 1991, Lombardi 2004, Lombardi et al. 2015, 2016).

For these reasons, it strikes us as a very strange thing to say that Shannon's is "the most unproblematic technical use of information in biology", that "It is common to begin the analysis of information in biology with an uncontroversial but minimal notion: a causal or correlational conception", or that such a notion of information "does not require much philosophical attention" (Godfrey-Smith & Sterelny 2016, see also Godfrey-Smith 2004, p. 276). First, the very expression "causal or correlational" passes off as synonymic what should be two disjunctive terms. As to the claim that information theory is unproblematic, actually the opposite is the case. Shannon's theory still raises important philosophical questions and debates (particularly in the philosophy of physics). It is in this sense that we believe that extracting an informational parity consequence from information theory is incorrect because one cannot take a causal understanding of Shannon's theory for granted – not at least without properly addressing such an interpretation.

[4] For example, Griffiths et al. (2015) use information theory as a tool to measure the specificity of causal contributions to an effect, making use of Woodward's interventionist theory of causal explanation. Their idea is that the more specific the relationship between a cause variable and an effect variable, the more information we have about the effect after we perform an intervention on the cause. However, theirs is an application of the theory on what *happens to be* a causal relation, not a causal interpretation of the theory. It is not very different from ecologists' use of information theory to measure species diversity. The ontological status of information within information theory (whether the transmission of information *is fundamentally* a causal relation) is a different matter.

Barton Moffatt argues that much of the literature on information, in philosophy of biology, is "fundamentally confused" regarding what it is often called the "causal information"[5]: "It neglects the core idea of mathematical information and conflates three distinct senses of information—the sense found in information theory and two related notions developed in Dretske (1981)" (2011, p. 287). The core idea of information theory is not that of systematic causal dependence of D-states on S-states, but that the amount of transmitted information can be measured. As to Dretske, he performs a modification of Shannon's theory so that individual messages could also be dealt with, and another concept focused on content (non-quantitative). The literature confuses Shannon´s theory with the former, inasmuch as it lacks references to averages; from the latter, the literature takes the explicit causal element. Importantly for our purposes here, causality is not sufficient in this version of Dretske's either, since there are other requirements (so note that even if it was Dretske´s second notion the one underlying some authors' claims, this one has been misunderstood as well).

Be as it may, the question remains why should we appeal to Shannon's theory (or interpretations of it) to elucidate the notion of information biology, in the first place. If the objections raised against the fruitfulness of Shannon's theory to elucidate the informational talk in biology (cf. Section 2) are on the right track, it should be concluded that the philosophical task of elucidating the notion of biological information requires a more qualitative (rather than quantitative) approach. This is the motivation of many alternative accounts. While it is most certainly true that not all of these accounts are equally "defensible" (to use Griffiths' own expression) because of their different

[5] Moffatt's remarks on the information-theoretic *mélange* is indeed insightful, although he sees the whole picture in way very different than ours. First, he understands CPT to be "the position that information cannot be used to differentiate between the role of genes and the environment in development" (2011, p. 284). We believe this characterization is mistaken. When Oyama first suggested parity between genes and other factors, her main concern was the illegitimate (in her view) attribution of a unique, exclusive character of the causal role of genes. The fact that this exclusiveness has been expressed via informational descriptions is not accidental or absurd, of course, but it is a different thing. This way of characterizing CPT derives from not distinguishing among: (1) causal parity as the thesis that genes and non-genetic factors are causally on a par, (2) informational parity as a version of the parity thesis, and (3) informational parity as a corollary of (a particular way of understanding) information theory. This is the reason why his analysis focuses on the misuses of information concepts and how they should be used instead. Our concern here is a different one: the conflations that make both rejections of informational parity-*consequences* and endorsement of an informational parity *thesis* incorrectly reasoned.

shortcomings[6], it appears to be currently agreed on, however, that developing a concept of information more grounded on the uses of this notion in specific (biological) context is a more promising enterprise than that of appealing to information theory (cf. Section 2.2). So not only is the inference of informational parity mistaken, but the mistake is contingent on the choice for a discouraged approach.

In sum, the inference of an informational parity-consequence from a causal understanding of information theory is incorrect. Therefore, the (undesired, for many) consequence of informational parity can be easily dissolved by dropping the information-theoretic approach to the problem of biological information (which we have independent and more serious reasons to do anyway). It wasn't a true problem after all.

6. Final remarks

In this article, we showed that on the one hand, informational parity is a consequence that follows from attempting to account for the notion of information in biology by appealing to information theory. This consequence, we explained, is considered to be inacceptable for some authors which, therefore, took a different approach in order to save the informational exclusiveness of genes (Section 4). On the other hand, informational parity is one of the many versions of CPT that expresses an actual feature of biological systems. According to this view, the idea of information must pick out the fact that genes and other factors are on a par (Section 2). In this paper, we exposed the flaws of both kinds of reasoning. The general problem is that there is a deep conflation between the concepts of information and causation (as concepts undisputedly related) that has not been sufficiently addressed, especially with respect to *the quarrel over parity*.

The current state of affairs is rather confusing and messy: we count with explicit pleas for causal parity, a number of attempts to refute CPT in strict causal grounds, a number of anti-parity accounts of information, and a couple of pro-parity accounts of information as well. It is not surprising that this issue had gone pretty much unnoticed.

It seems to us that the quarrel over causal parity can do well without the concept of information, as the "finest" causal analyses (Section 3) indicate. But what about

[6] Counter-intuitive corollaries, counter-examples, theoretical issues, conceptual issues, limitations, etc.

informational parity? A proper defense of informational parity based on the endorsement of CPT should, first, determine a causal relationship that is shared by DNA and other factors (where causation is understood in more fine-grained terms), and then pin down *in what sense* does this causal relation have anything to do with information (and what is it meant by this polysemantic term), although we don't see any obvious way to achieve this. One could also believe that, since the two concepts (information and causation) are oddly conflated (as we have argued here), then perhaps informational parity could be argued for in non-causal grounds. However, this too is tricky, because many accounts of information in biology do rely on causal considerations and because the most promising ones are those that actually reject parity claims (we lack space to develop this idea here).

The final conclusion we should draw from this paper is this: no talk of informational parity dependent on causal considerations (being CPT or a mistaken causal view of information theory) seems to be motivated so far. Informational parity does not actually follow from any such considerations.

References

Austin, C. (2015). "The dispositional genome: primus inter pares". *Biology & Philosophy*, 30, 2, pp. 227–246.

Bergstrom, C. & Rosvall, M. (2011). "The Transmission Sense of Information". *Biology and Philosophy*, 26, pp. 159-176.

Bogen, J., & Machamer, P. (2011). "Mechanistic Information and Causal Continuity", in P. Illari, F. Russo, & J. Williamson (eds.), *Causality in the Sciences*, Oxford University Press, pp. 845-864.

Calcott, B. (2014). "The Creation and Reuse of Information in Gene Regulatory Networks". *Philosophy of Science*, 81(5), pp. 789-890.

Cover, T. & Thomas, J. (1991). *Elements of Information Theory*. New York: John Wiley & Sons.

Crick, F. (1958). "On Protein Synthesis". *Symposium of the Society of Experimental Biology*, 12, pp. 138–163.

Darden, L. (2006). “Flow of Information in Biological Mechanisms”. *Biological Theory* (3), pp 280—287.

Downes, S. (2006). “Biological Information.” In Pfeiffer and S. Sarkar (Eds.) *Philosophy of Science: An Encyclopedia*. New York: Routledge, pp. 64–68.

Dretske, R. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Gannet, L. (1999). "What's in a Cause? The Pragmatic Dimensions of Genetic Explanations", *Biology & Philosophy, 14:* pp. 349–374.

Godfrey-Smith, P. (2000). “On the theoretical role of ‘genetic coding’”. *Philosophy of Science*, 67, pp. 26–44.

Godfrey-Smith, P. (2004). “Genes do not encode information for phenotypic traits”, in: Hitchcock (ed.), *Contemporary debates in philosophy of science,* Oxford: Blackwell Publishing Ltd., pp. 259-274.

Godfrey-Smith, P. (2008). “Information in Biology”, in: Hull, D., y Ruse, M. (Eds.), *The Cambridge Companion to the Philosophy of Biology*. Cambridge: Cambridge University Press, pp. 103-119.

Godfrey-Smith, P. y Sterelny, K. (2016). "Biological Information", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2016/entries/information-biological/>.

Griffiths, P. & Gray, R. (1994). “Developmental Systems and Evolutionary Explanation”. *Journal of Philosophy*, 91, pp. 277-304.

Griffiths, P. & Gray, R. (2005). “Three ways to misunderstand developmental systems theory”. *Biology and Philosophy*, 20: 417-425.

Griffiths, P. & Knight, R. D. (1998). “What is the developmentalist challenge?” *Philosophy of Science*, 65, pp. 276–288.

Griffiths, P. (2001). “Genetic Information: A Metaphor in Search of a Theory”. *Philosophy of Science,* 68, 3, pp. 394-412.

Griffiths, P. (2005). “The fearless vampire conservator: Philip Kitcher, genetic determinism and the informational gene”. In: E. Neumann-Held and C.

Rehmann-Sutter (eds.), *Genes in Development: Re-reading the Molecular Paradigm*. Durham: Duke University Press.

Jablonka, E. (2002). "Information: Its Interpretation, Its Inheritance and Its Sharing." *Philosophy of Science,* 69, pp. 578-605.

Jablonka, E. (2002). "Information: Its Interpretation, Its Inheritance and Its Sharing." *Philosophy of Science,* 69, pp. 578-605.

Kjosavik, F. (2007). "From symbolism to information? –Decoding the gene code". *Biology and Philosophy*, 22, pp. 333-349.

Kjosavik, F. (2014). "Genes, Structuring powers and the Flow of Information in Living Systems." *Biology and Philosophy*, 29 (3), pp. 379-394.

Lewis, D. (2000). "Causation as influence." *The Journal of Philosophy,* 97, pp. 182-197.

Lombardi, O. (2004). "What is information?". *Foundations of Science*, 9, pp. 105–134.

Lombardi, O., Fortin, F., and Vanni, L. (2015). "A pluralist view about information." *Philosophy of Science*, 82, pp. 1248-1259.

Lombardi, O., Holik, F., and Vanni, L. (2016). "What is Shannon information?" *Synthese*, 193, pp. 1983-2012.

Maynard Smith, J. (2000a). "The concept of information in biology". *Philosophy of Science*. 67, pp. 177–194.

Maynard Smith, J. (2000b). "Reply to commentaries". *Philosophy of Science*, 67, 2, pp. 214-218.

Merlin, F. (2010). "On Griffiths and Gray's concept of expanded and diffused inheritance". *Biological Theory*, 5(3), pp. 206-215.

Moffatt, B. (2011) "Conflations in the Causal Account of Information Undermine the Parity Thesis". *Philosophy of Science,* 78, pp. 284-302.

Oyama, S. (1985). *The Ontogeny of Information: Developmental Systems and Evolution*. Cambridge: Cambridge University Press.

Oyama, S. (2000). "Causal democracy and causal contributions in Developmental Systems Theory". *Philosophy of Science,* 67, S332-S347.

Oyama, S.; Griffiths, P. & Gray, R. (2001). *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge: The MIT Press.

Planer, R. (2014). "Replacement of the "Genetic Program" Program." *Biology and Philosophy*, 29, pp. 33-53.

Planer, R. (2016). "Are Genetic Representations Read in Development?" *The British Journal for the Philosophy of Science*, 67, pp. 997-1023.

Pradeu, T. (2010). "The Organism in Developmental Systems Theory". *Biological Theory*, 5(3), pp. 216-222.

Sarkar, S. (2004), "Genes Encode Information for Phenotypic Traits", in Hitchcock, C. (Ed.), *Contemporary Debates in Philosophy of Science*, Oxford: Blackwell Publishing Ltd., pp. 259-274.

Shannon, C. (1948). "The Mathematical Theory of Communication". *Bell System Technical Journal* 27, pp. 379–423.

Shea, N. (2007). "Representation in the Genome and in other Inheritance Systems". *Biology and Philosophy*, 22, pp. 313–331.

Shea, N. (2011a). "Developmental Systems Theory Formulated as a Claim about Inherited Representations. *Philosophy of Science,* 78(1), pp. 60-82.

Shea, N. (2011b). "What's Transmitted? Inherited Information". *Biology and Philosophy, 26*, 2, pp. 183-189.

Stegmann, U. (2005). Genetic information as instructional content. *Philosophy of Science*, 72, 3, pp. 425–443.

Stegmann, U. (2012a). "Causal Control and Genetic Causation. *Noûs*, doi: 10.1111/j.1468-0068.2012. 00867.x

Stegmann, U. (2012b). "Varieties of parity". *Biology and Philosophy,* 27, 6, pp. 903-918.

Sterelny, K. & Griffiths, P. (1999). *Sex and death. An introduction to philosophy of biology*. Chicago: The University of Chicago Press.

Stotz, K. (2006). "With 'genes' like that, who needs an environment? Postgenomics's argument for the 'ontogeny of information'". *Philosophy of Science*, 73, pp. 905–917.

Waters, C. K. (2007). “Causes that make a difference”. *Journal of Philosophy*, 104, pp. 551–579.

Weber, M. (forthcoming). “Causal Selection versus Causal Parity in Biology: Relevant Counterfactuals and Biologically Normal Interventions. In: K. Waters, M. Travisano and J. Woodward (eds.), *Philosophical Perspectives on Causal Reasoning in Biology*. Minneapolis: University of Minnesota Press.

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, pp. 287–318.