**Causing something to be one way rather than another. Genetic information, causal specificity and the relevance of linear order**

**Barbara Osimani**

1. Special Issue on Information: Space, Time, and Identity (DTMD 2013); 43 (6): 865-81.

**Abstract**

Following Crick’s central dogma as well as Schrödinger’s and Monod’s popular disseminating works on the discoveries of molecular biology, the notion of information has enjoyed widespread use in biology. The teleosemantic approach has proposed a specific sense in which the notion of information should be interpreted in relation to genetic phenomena: genetic information should not be understood merely in Shannon’s correlational sense, but in a symbolic sense. This view has been attacked both on substantive and on theoretical grounds. Critics of information talk, especially developmental systems theorists, have claimed that insistence on the notion of information perpetuates gene-centricity, thereby neglecting the complex dynamic of gene-environment interaction in development. Philosophers of science have raised doubts about the symbolic nature of genetic information and have proposed to capture the intuitions related to the teleosemantic projects by drawing on the notion of instruction (Stegmann, 2004) or formal system (Sarkar, 2003). Deflationists such as Boniolo (2003, 2008) and Godfrey Smith (2000) plainly propose to make without the notion of information altogether and to simply substitute it with causal relationships. This has led to a second wave of efforts to justify its use on scientific grounds. Two additional notions of information have emerged: Shea’s infotel semantics (Shea 2013, 2007) and Bergstrom and Rosvall “transmission sense” of information (Bergstrom and Rosvall 2007). The former defends the semantic notion of information by developing the representational approach of teleosemantics in an ontogenetic perspective, while the latter defends Shannon’s information notion as a valid paradigm applied to genetic phenomena to the extent that they are also characterized by the decision-theoretic problem of how to package information for transport. I advance the view that three common minimal denominators are shared by all theories: 1) causal specificity 2) the combinatorial mechanism of the genetic code; 3) code “arbitrariness”; and I propose an analysis of the notion of genetic information based on the conceptual tools developed within philosophical theories of causality on the one hand, and of linguistics as well as philosophy of information on the other. The conclusion is that genetic phenomena are causal in a very special sense: 1) *they cause something to be one way rather than another* (causal specificity) 2) by combining elementary units one way rather than another (linear order). A test for this approach is provided by the notion of genetic error.

1. **The early phase**

The origin of the use of the word “information” and the related concept of “code” is generally ascribed to the famous “Central Dogma” by Francis Crick:

“[O]nce information has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but the transfer from protein to protein, or from protein to nucleic acid is impossible. *Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein* (Crick 1958, 153, my emphasis)”.

In this quote, genetic information is explicitly defined as the determination of the sequence of bases in the nucleic acid or of amino acids in the protein, i.e. by referring to the linear order in which they are disposed. Thus the terminological choice seems to be justified by the language-like combinatorial mechanism of genetic expression. Interestingly enough, this idea is also present in Erwin Schrödinger’s popular book “What is life?” (1945), which was published at a time when the structure of DNA and the related molecular mechanisms was yet to be discovered. Schrödinger identifies chromosomes as the carriers of a ciphered code, the entire plan for the future development of the individual, and speaks of “jump-like” changes (with reference to De Vries experiments) while insisting on the essential discontinuity of genetic phenomena (pp. 32-37). Furthermore the different atom dispositions in the molecule are held responsible for the great variety and at the same time precision of the transmission of hereditary characters, much in the same manner as the different dispositions of the same atoms in a molecule produce isomers of the same substance with possibly different chemical and physical properties:

“From the view we have formed of the mechanism of mutation we conclude that the *dislocation of just a few atoms* within the group of 'governing atoms' of the germ cell suffices to bring about *a well-defined change* in the large-scale hereditary characteristics of the organism” (77, my emphasis).

Schrödinger conjectures that the mechanism at the basis of the transmission of phenotypic characteristics must be based on a combinatorics of a few elementary biological units where any single particle at the micro level may make a difference at the macro level. This distinguishes these phenomena from causal phenomena explained by statistical laws such as those described by thermodynamics, mainly grounded in the collective behavior of particles which, individually taken, do not play but an insignificant role.

Long after the DNA structure has been discovered, another Nobel prize winner, Jacques Monod, develops an anti-metaphysical philosophy out of the phenomena of genetic regulation, where the notion of information has a major role and is used in different senses. His most popular book “Chance and necessity” (1971) generously uses the word information in order to make reference to a set of different biological mechanisms at a molecular level.

1. A first sense relates to *hereditary transmission*. In this sense, information means the structural morphology of a biological entity reproduced from one generation to the next.[[1]](#footnote-1)
2. A second and related sense is that of *specificity*. This is a key term in the history of biology and has been at the center of a vivacious debate between supporters of the “continuity” of nature against advocates of its discontinuity.[[2]](#footnote-2)

A third sense in which Monod uses the word information is the *cybernetic* sense. This refers to genetic regulation and to the “coordination” manifested by catalytic reactions in the presence of different environmental stimuli. For example the catalytic activity of allosteric enzymes depends on the chemical potential of all three “effectors” (one inhibitor and two activators). This mechanism could be simplified as follows: let P be an inhibitor and Q and R activators of the catalytic reactions. Then the activity can be activated or interrupted depending on the aggregate threshold value independently contributed to by the three different effectors. An hypothetical mechanism could then be formalized as follows:

IF “P ≥ p AND Q ≤ q AND R ≤ r”, THEN “suspend catalysis”.

Monod explicitly speaks about Boolean logical properties of allosteric enzymes: these properties consist in measuring the values of the effectors and in triggering the appropriate reaction.

So, whereas Crick’s notion of information strictly refers to the correspondence between sequences of nucleotides and amino acids (and between sequences of amino acids and proteins), and Schrödinger emphasizes the distinction between genetic mechanisms and physical statistical laws, Monod associates to this picture also a series of cybernetic categories. In subsequent decades, the notion of information has undergone severe criticisms (Sarkar, 1996; Mahner and Bunge, 1997; Godfrey-Smith, 1999), and in response to such criticisms philosophers of biology have further refined this concept and investigated its epistemological foundations.

Proponents of the teleosemantic approach (Sterelny et al. 1996; Maynard Smith, 2000, Jablonka, 2002); consider the notion of information as a central one for contemporary biology: “developmental biology can be seen as the study of how information in the genome is translated into adult structure, and evolutionary biology of how information came to be there in the first place” (Maynard-Smith, 2000: 177). Justification of information talk in teleosemantics follows from four *de facto* interconnected but theoretically independent reasons:

1) the “*symbolic*” nature of the code (its “arbitrariness” as opposed to other sorts of relationships based on indexicality or iconicity). Arbitrariness means that there is no chemical necessity determining which amino acid any nucleotide triplet should code. CAU codes for istidine and CUA for leucine, but there is no chemical reason for which the mapping could not be reversed.

2) The fact that genes *specify form* and functions of the proteins and of the organism as a whole. The “symbolic” nature of molecular biology also means that it “makes possible an indefinitely large number of biological *forms*” (Maynard-Smith, 2000: 185, my emphasis).

3) The “*intentionality*” of the genetic program. The “meaning” of a genetic sequence and its related protein consists in its teleologic properties developed by being confronted with natural selection: i.e. its being functional to the organism’s survival. Thus genetic information is intentional in the sense that it has a goal: the survival of the organism which has inherited it. In analogy to algorithmic programs, DNA contains information that has been programmed by natural selection (Sterelny et al. 1996).

4) Genes are what allows trait inheritance from one generation to the next (genes as “replicators”). The special status of genetic factors is linked to their capacity to replicate themselves and thereby allow inheritance of traits and functions (Dawkins, 1976, 1982; Maynard Smith and Szathmary, 1999).

However, skepticism about informational talk has not spared the teleosemantic account of genetic information either. Opponents of this notion attack it on opposite grounds: whereas developmental systems theorists object that emphasis on this notion hides the importance of the environment in development (since the environment also transmits inheritable “information”), causal reductionists would instead reduce it to the notion of causality altogether.

**2. Objections to informational talk**

Developmental Systems Theorists (DSTs) attack the notion of genetic information mainly in order to defend the role of epigenetic and environmental phenomena in systems development. Their lines of argument mainly impinge upon:

1. So called “parity thesis”;
2. Biological complexity;
3. Scientific use of the word.

According to the *Parity Thesis* (Griffiths and Knight 1998) “any defensible definition of information in developmental biology is equally applicable to genetic and non-genetic causal factors in development” (Griffiths, 2001: 396). Genes are on a par with other environmental factors as causal determinants in the development of the organism because environmental cues in the organism or outside it play an important role in the ontogenetic development of the individual organism (Griffiths and Gray 1994; Oyama 1985). More importantly, information does not pre-exist anywhere and it is the result of genome-environment interaction: the DNA acquires its informational properties exclusively by interacting with the environment (Oyama, 1985). The problem with gene-centricity is to be found in the very notion of genetic information and especially in its supposed “semantic properties” (Griffiths 2001). Ultimately, genetic privilege is the result of an unjustified distinction among kinds of causes, or among causes and background conditions which is based on the idea that genes have a unique and particular way to bring about their effects. Furthermore, genetic determinism is considered to be a direct consequence of this way of thinking about genes (Oyama 2000). Support to such claims is lent by phenomena occurring at diverse levels of development and system-environment interaction: from epigenetic phenomena such as methilation to “host-imprinting” (the acquired tendency in some insects to lay eggs on the same kind of plants where they hatched as offspring). According to DSTs not only do these factors influence organism development in the same way as genes do, but, more importantly from an evolutionary point of view, they are also carriers of information (the plant on which eggs are laid probably optimizes fitness for the species) (Griffiths, 2001). For the same reason, DSTs also refute inheritance as a reason to accept genetic privilege.

*Complexity* makes it difficult to track causal webs in all their multiple paths interacting in different ways. The complex web of mechanisms governing genetic expression, with myriads of pathways and causal interactions both in the cellular environment as well as inputted by the external environment, make genetic causality less robust than expected by “code for” genes (Kendler 2005). One theoretical consequence of this state of affairs is also that the DNA sequence alone is not sufficient to identify the “gene”, hence functional criteria are also needed in order to categorize genes as basic molecular units (Burian, 2004; see also the more general debate on the concept of gene: Mueller-Wille and Rheinberger, 2009; Fox Keller, 2000).

Another objection to teleosemanticists’ informational talk is that it does not correspond to *how biologists use the term information when they explain biological phenomena* (Griffiths, 2001: 410). As Godfrey-Smith puts it: “The solution of the problem of protein synthesis, made possible by the concept of genetic coding, does not require or directly involve any hypothesis of evolutionary history.” (2000: 34). Sarkar (2000: 211) lends further weight to this sort of objection by underlining that neither intentionality nor natural selection are necessary to justify talk of information with reference to genes containing information about proteins.

The only point which DSTs concede to teleosemanticists relates to the distinction between genes and environment on the basis of the unlimited number of possible combinations of the basic constituents which characterize the former with respect to the latter: this allows for an unlimited number of possible heritable states, whereas in the environment these are much more limited. According to DSTs, however, this distinction cannot ground the claim that developmental information resides entirely in the genome (Griffiths, 2001: 404).

Godfrey-Smith further elaborates on the above considerations (Godfrey-Smith 2000) and emphasizes a division of labor for the notion of genetic coding: whereas the code model has been useful to solve the puzzle of protein synthesis, it has instead little explanatory value when used to understand development and evolution, as well as to trace a distinction between what is genetic and what is environmental. All these issues should be instead addressed “just using causal concepts” (2000: 43).

In the same vein, Boniolo (2003, 2008) accords information talk only a metaphorical status and reduces the notion of genetic expression to a (probabilistic) causal chain. He recognizes that two different theoretical questions historically underlay the biophysical vs. the informational approach to molecular genetics, namely the discovery of the concrete molecular mechanisms and the formal question of the correspondence between DNA and amino acids/proteins (2008: 207). But, he claims, even if the notion of information might have had an important heuristic role at the beginning of molecular biology, it has by now lost any useful explanatory function: a purely biophysical approach to gene expression is sufficient in order to provide an exhaustive account of the phenomenon.

John Wilkins (unpublished) recognizes, in line with Sarkar (2000), a heuristic value to the concept of information, however he argues that the notion of information can be substituted by causality because “If information is in the concrete world, it is causality. If it is abstract, it is in the head” (p.1). According to him, the fact that genes are seen as carriers of information is built into their concept; in fact, he goes on, “genes” are an abstract construction defined as that which causes heritable differences. Likewise for teleosemantic approaches: we don’t discover that genes store information selected for by natural selection, instead we define them that way.

Hence, whereas developmental systems theorists would be ready to concede that information is carried by environmental factors as well, and refuse its use only to the extent that it is employed for making a qualitative distinction between genes and the environment, deflationists warn us instead that information discourse hides nothing else than correlational relationships and thus should be replaced by causation. However, the move of reducing the notion of information to that of causation is not less problematic from an ontological point of view, in that it poses the highly problematic question of the ontological status of causality. Furthermore, the concept of cause is itself at the center of strong philosophical controversies; hence reducing the notion of genetic information to that of causality, provides little more information than the notion of information itself. I will elaborate on this point in section 5. In the next section I present counterarguments to the objections raised by DST and deflationist accounts to the notion of genetic information, proposed by a second wave of teleosemantic theories.

**3. Infotel semantics and the transmission sense of information**

As an answer to skeptics about the explanatory role of the notion of information in development, two additional theories regarding the representational content of the genome have been developed: infotel semantics, proposed by Nicholas Shea (2007, 2013), and the transmission sense of information, by Carl Begstrom and Martin Rosvall (2007).

Shea’s proposal intends to oppose Sterelny’s objections that in order for genetic information to be something more than mere causal information (i.e. the basis for an inference drawn because of a supposed correlation between data and source, such as smoke and fire), and thus to bear semantic/representational properties, it must also be *read as such* by the relevant developmental components (Sterelny, 2000). As an answer to this reservation, Shea presents the developing organism as a consumer of genetic information: in the same manner as organisms achieve an adaptive fit to a feature of their environment by detecting information about that feature during their development process, so they achieve the same goal also by reading information about that feature from their genes. So the organism treats the genome as a proxy for facts about the world and behaves consequently (2013: 5). Shea suggests that some aspects of development be understood as the organism’s reading the information carried by the genome and displays a series of cases in which his proposal is shown to have explanatory relevance (2013: 15-27).

Bergstrom and Rosvall (2011) defend the adoption of Shannon’s notion of information against criticisms that take it to merely reflect a “causal sense” of information: they argue that such critiques miss the point in that they neglect the *raison d’être* of the mathematical theory of communication. In fact this is mainly focused on “how far mathematical objects such as sequences and functions can be compressed without losing their identity, and if compressed further, how much their structure will be distorted”. Thus, Bergstrom and Rosvall adopt a “transmission sense” of information and focus on the underlying decision problem of how to optimize the process of information package and delivery through such tools as information compression, exploitation of redundancies, and minimization of distortions.[[3]](#footnote-3)

The teleosemantic approach and the elaboration proposed by Shea and Bergstrom and Rossvall somehow reverse the point of using the notion of information as advanced by the pioneers of molecular genetics. In fact, whereas Crick and contemporaries drew on the notion of information as a primitive concept in order to explain trait inheritance, teleosemanticists feel compelled to justify the notion of information itself: from being the explanans, the notion of information has become the explanandum. More fundamentally, whereas in the early phase the information “contained in the genes” was information about the phenotypic traits of the individuals, for teleosemanticists the information also reveals the traits of the environment which contributed to forge the actual genetic make-up, in an evolutionary perspective. So in one paradigm the information content is protein specificity whereas in another it also refers to the traces left by evolution through natural selection on the gene pools.

However, there is a common ground shared by earlier and more recent defenders of the notion of genetic information. This includes the following characteristics: 1) the combinatorial mechanism of the genetic code, which is accompanied by 2) code arbitrariness, and 3) causal specificity (the systematic relationship of genetic and phenotypic traits). Genetic information is thus seen as an arbitrary but systematic correspondence between *combination of units* in one system and *units* in another system (at multiple levels). Furthermore, such correspondence is endowed with the causal power to reliably generate an enormous diversity of forms. This clarification constitutes a first step to addressing the two kinds of objections raised against the use of informational talk, which can now be reformulated as follows: is there a way to justify the use of the word “information” in relation to these three characteristics, which cannot be equally applied to other biological phenomena (contrary to what is suggested by DST scholars)? Furthermore, can these three characteristics be jointly reduced to mere causation (as would be suggested by deflationists of the notion of information in biology)?

**4. Information and causality**

The debate between advocates of the notion of information and deflationists aiming to reduce it to the notion of causality is bound to remain unfruitful “crosstalk” unless notions of information and causality, which are themselves topics of a strong philosophical debate, are themselves disambiguated.

The main ambiguities concerning the notion of information are related to the confusion between its epistemic vs. its ontological sense (Prigogine and Stengers, 1984). This confusion might well have been nurtured by the connection, fostered by Shannon and Weaver’s paradigm, between entropy (a concept borrowed from physics) and information transmission (see also Wicken, 1987). Such a distinction has been emphasized by Floridi (2002) and Albert Borgmann (1999) who adopted a tripartite articulation of the concept of information (underpinned by a slightly different semantics; I will refer here to Floridi’s analysis): 1) *information as reality* (related to the notion of entropy and structural realism), 2) *information about reality* (semantic information associated with truth-functional values, i.e. “alethically qualifiable”), 3) and *information for reality* (instructions which can be either successfully executed or not). Floridi (2011) also suggests a tentative definition of information as well-formed, meaningful data, where a datum is labeled as any “putative fact regarding some difference or lack of uniformity within some context” or a “fracture in the fabric of being”, thus connecting the ontological dimension with the epistemic one.

On the other side, the philosophical debate on causation addresses: 1) *semantic* questions concerning the *ti estì* of causality (“what is a cause”, in technical terms: the intensional meaning of the term “cause” and its extensional domain, i.e. “what relationships can be said to be causal”); 2) the debate inaugurated by Hume on the *metaphysical status* of causal relationships (objective reality vs. mind-dependent category), 3) *epistemic* questions concerning the validity of inductive leaps – the inference from correlation to causation and 4) *methodological* efforts devoted to the distinction of causal from non-causal correlations (see Schaffer, 2008). While these domains are obviously interrelated, sometimes they have been unduly confused; however, for our purposes, the first point is mostly relevant. In fact, among the various characterizations of causality offered by the philosophical literature (causes as INUS conditions - Insufficient but Non-redundant parts of a condition which is itself Unnecessary but Sufficient for the occurrence of the effect, see Mackie 1988; probabilistic causality, the regularity view, the invariance view, causes as mechanisms, causes as processes, causal powers), the notion of causation as influence (Lewis, 2000), which is itself based on a counterfactual analysis of causation (Lewis, 1973), seems to provide the bridge between the intuitions which underlie information talk related to genetic phenomena through the notion of “causal specificity” and formalizing them into a testable theory. As a matter of fact, the notion of causal specificity comes very close to the notion of information (as reality) and is akin to the Aristotelian notion of “formal cause” or “eidos”: the principle which causes something to be as it is (Aristotle Physics. II.3; and Metaphysics A.3 ff; Δ.2), in contrast to efficient causality, which causes something to be, period.[[4]](#footnote-4)

1. **Causal specificity**

Philosophers of science addressing the issue of genetic causality have been mainly interested in drawing a line between genetic and other kinds of causal phenomena in biology (rather than addressing the issue of genetic information *per se*). The first contribution in this sense can be found in Šustar (2007): Šustar also intends to reduce the notion of genetic information (particularly, as found in Francis Crick’s formulation of so-called “Central Dogma of molecular biology”) to causality, however he insists on the importance of the distinction between different kinds of causal notions. According to him, unlike other non-genetic phenomena, in the case of protein synthesis from DNA molecules, *precise and punctual changes* of the DNA sequence determine as many and as much precise changes in the protein product. By drawing on the notion of “template”, coined by Crick himself, and on the notion of colinearity applied by Yanofsky and Brenner to the correspondence between gene structure and protein structure (1964), Šustar proposes a definition of genetic causality as a “template correlative determination”, thereby blending in one formula both the regularity notion of causality as correlation, as well as that of information entailed by the concept of template.

Independently from Šustar, Waters (2007) and Woodward (2010) base their analysis of genetic phenomena on a more sophisticated counterfactual notion of causality (Woodward, 2003) and adopt Lewis notion of causation as influence (Lewis, 2000) to justify their distinction between genetic and nongenetic phenomena related to organism development. In Lewis’ terms, causation should be understood as a systematic counterfactual relationship between a range of alterations of a kind of event and a range of corresponding alterations on another kind of event (Lewis, 2000: 190), hence they develop the notion of “influence” as a way to flesh out the idea of multiple dependence relationships from one kind of event to another.

Waters and Woodward’s distinction between genetic causality and other kinds of biological mechanisms at the molecular level are based on these patterns of dependence and on Woodward’s (2003) manipulative account of causation.[[5]](#footnote-5) Woodward’s (2003) approach to causality combines the notion of intervention (which in some senses substitutes the Lewisian concept of “miracle”) with the notion of invariance. Invarianceis a systematic relationship holding between the values of a variable and the values of another variable whenever they are connected by a causal relation.

While drawing on both Lewis’ (2000) and Woodward’s (2003) analyses of causation, Kenneth Waters is committed to the ontological distinction between causes and necessary conditions. This is precisely the distinction which is contested by opponents of genetic privilege such as for instance Gannett (1999).[[6]](#footnote-6) Given that both Lewis’ counterfactual and Woodward’s counterfactual-manipulationist account are also not sensitive to this distinction, and because Waters thinks this is precisely the distinction which is needed in order to explain the peculiarity of genetic causality with respect to other kinds of developmental mechanisms, he uses the notion of “actual difference maker” for this purpose. The actual difference maker is, relative to a population p, the factor which differs among individuals (for instance a genetic mutation being present in some but not all individuals) and which fully accounts for observable differences among them (for instance a phenotypic trait). Waters’ central point is that the actual effect relates to a difference in a population (p. 567). In order to account for genetic causality in eukaryotic cells, Waters also develops the concept of *an* actual difference maker (or, better, an actual difference changer). In fact, in this case, also the different kinds of RNA polymerase are difference makers in RNA synthesis in that intervening on them could for instance stop the synthesis of mRNA and thus affect the structure of RNA. However, Waters insists, DNA is a specific difference maker in the sense that:

“different changes in the sequence of nucleotides in DNA would change the linear sequence in RNA molecules in many different and very specific ways. RNA polymerase does not have this specificity. Intervening on RNA polymerase might slow down or stop synthesis of a broad class of RNA molecules, but it is not the case that many different kinds of interventions on RNA polymerase would change the linear sequence in RNA molecules in many different and specific ways” (574-575).

Among difference makers, DNA has a peculiar character with respect to other enzymatic mechanisms contributing to genetic expression in that *many specific differences in the DNA account for corresponding differences in the resulting protein* (“specific difference maker”). Woodward’s contribution to this specific issue (Woodward, 2010) further elaborates on this point. Woodward’s (2010) explicit concern is to specify what it means to claim that the causal relationship between a DNA sequence and the “proteins” for which it codes is “specific”. Like Waters he intends to contrast the idea of causal parity between genes and cellular machinery as advocated by development system theorists. Furthermore he aims to provide a way to distinguish genetic causality manifested in mendelian vs. non-mendelian traits.

Analogously to Waters (2007), Woodward develops the notion of “*specificity*” in order to formalize the distinction between genetic causality and other kinds of cellular mechanisms. However Woodward provides a different formalization of this property and he suggests to use it not as a discriminator between causes and background conditions, or causes and non-causes, but as a criterion for distinguishing *different kinds of causes*. Specificity is fleshed out by elaborating on Lewis’ notion of “influence”. Woodward’s proposal is that, other things being equal, we are more or less inclined to think of C as having more rather than less influence on E (and as a more rather than less specific cause of E) to the extent that: there are a number of possible states of C (c1, c2, … cn), a number of possible states of E (e1, e2, … en); a mapping from C to E that describes patterns of counterfactual dependency between states of C and states of E that support interventionist counterfactuals (the mapping need not be bijective, an onto function suffices): “C will influence E to the extent that by varying the state of C and its time and place of occurrence, we can vary the state of E *in a fine-grained way*” (p. 305, my emphasis).

Thus, the transcription of a DNA sequence into an amino acid sequence – and the related translation into proteins – differs from other catalytic processes which contribute to this process in that “interventions that change this [the DNA] sequence in many different and specific ways, will also change the RNA molecule in different and specific ways” (pp. 303-304). DNA transcription and translation can be described as cases of causal influence because there are many possible states of the DNA sequence, and many variations in this sequence are systematically associated with different possible corresponding states of the linear sequence of the mRNA molecules and of the proteins synthesized. By contrast, DNA polymerase, or other catalytic processes involved in gene expression are responsible for protein production in an undifferentiated fashion: the cellular machinery is involved in the production of any protein. In order to illustrate this point Woodward draws on the metaphor of the radio switch and dial and compares the cellular machinery to the switch, whereas genetic causality is paralleled to the dial: the switch can take only two values (on/off), and its state is only causally relevant to whether *any* station at all is received; by contrast, the position of the dial “influences” *which* station is received.

Both in Waters’ as well as in Woodward’s account of genetic privilege, the difference between cellular machinery and the contributing roles of gene to protein specificity is framed in terms of distinct causal stories: whereas enzymes and other proteins help to determine any amino acid and protein, the role of the basic DNA sequence is to provide the structure for the production of *specific* amino acids and proteins. These two causal stories can be seen on one side as causes which cause something to be, while on the other they can be seen as causes causing something to be *one way rather than another.[[7]](#footnote-7)* I will use this definition of causal specificity as the first component of my definition of genetic information. This formulation only partly responds to the objections raised by developmental system theorists – by clarifying why genetic causality is a particular kind of causality with respect to other developmental phenomena – but it idoes not straightforwardly legitimize the use of the notion of information, and hence does not answer the objection raised by deflationists either, who want the notion of biological information (whether genetic or not) to be reduced to that of causality. Full justification of the notion of information will come by taking into account also the other two characteristics of genetic causality: combinatorial mechanism and code arbitrariness.

1. **Combinatorial mechanism and code arbitrariness: genetic causality and language**

As mentioned above, another fundamental dimension emphasized by teleosemanticists and other defenders of the notion of genetic information relates to the combinatorial mechanism at the basis of this generation of diversity. Indeed, what mostly explains the success of information terminology such as the adoption of such terms as “transcription”, “translation”, “expression” and the like is that in this mechanism it is not the single nucleotide which is relevant, but rather *its place* in the linear sequence. The linear order in which the nucleotides are distributed constitutes the centre of the linguistic/information metaphor. If we consider standard linguistic phenomena, we also observe that an “atomic” morphological difference makes a macroscopic difference at the level of meaning. In order to illustrate this point, I will refer to the phenomenon of “double articulation”, first described by the French linguist André Martinet (1960), which refers to the two lowest levels of linguistic meaning: the highest between the two (n-1) refers to morphological signs, for instance the letter *s* which, when added to nouns, makes them plural (hence *s* “means” plural):

n-1) “**f**riend + **s” = plural of friend 🡪 s = plural**

The lowest of all levels of meaning concerns opposing pairs (or “minimal pairs”, i.e. pairs of phonemes which can be distinguished by being included in otherwise identical words having different meanings) like the role of the letters *s* and *t* in producing the semantic difference between “saw” and “taw”:

n) “**s**aw” ≠ “**t**aw” **🡪 s = ?; t = ?**

At this lowest level, neither the letter *s* nor the letter *t* mean anything per se, but they contribute substantially to the meaning of the word in which they appear as a function of where exactly they appear. Order relevance is also evident in the following pairs: “take” vs. ”Kate”; “sit” vs. “its”; or “rat” vs. “art”, where the words consist of exactly the same letters, but with changing order. This parallels the role played by linear order in the nucleotide sequences. The following table shows a series of different permutations resulting from changing the linear order of triplets of nucleotides. Changing the order of the nucleotides result in different amino acids even if the combination of nucleotides is the same (U taken two times and A taken one time; or A, C and U taken one time each):

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **X** | **Y** |
| **Combinations**  |  | **Nucleotide sequence** | **Amino acid** |
| (U, U, A) | Permutations | UUA | Leucine |
| AUU | Isoleucine |
| UAU | Tyrosine |
| … | … |
| (A, C, U) | CAU | Istidine |
| AUC | Isoleucine |
| CUA | Leicine |
| … | … |

**Tab. 1: Different amino acids resulting from the different permutations of the same nucleotide triplet.**

For instance the combination (A,C,U) corresponds to different possible permutations: CAU, AUC, CUA. These triplets are then “translated” into different amino acids (istidine, isoleucine, leucine respectively). So what makes a difference is not only the “value” of the variable, but also its place in the sequence.[[8]](#footnote-8) Hence, the intuition underpinning the use of the notion of information relates to the relevance of linear order for the genotype-phenotype correspondence. Furthermore, it is clear that neither C nor A cause their effect in virtue of their chemical composition, but rather in virtue of their position in the combination with the other two nucleotides. Evidence of this is given by the fact that AUA and AUC translate to the same amino acid (isoleucine). The relevance of linear order in genetic mechanisms characterizes the role played by each nucleotide in determining the final protein as fundamentally relational, in that it is not a function of the nucleotide chemical structure, but (mainly) of the place it occupies in the sequence. This distinguishes genetic causality from other sorts of biological phenomena mainly involving exchange of energy or matter. Informational talk is legitimized by the relevance of linear order and by its predominance over other biological factors such as matter or energy exchange. However, the sort of information involved here is not “information about reality”, as in verbal language, but rather “information as reality”. More precisely, “information as reality” at a lower level (nucleotide sequences) produces “information as reality” at a higher level (amino acids, proteins and phenotypic traits). Hence my proposal is to combine the notion of causal specificity presented above with an operational definition of the coding process and then define genetic causality as: a special kind of cause which causes something to be one way rather than another, by combining elementary units one way rather than another (linear order). This characterization allows to reconcile all common denominators adduced by different advocates of the notion of genetic information (causal specificity as well as combinatorial and arbitrary mechanism) in one definition, and at the same time to clarify the sense in which the notion of information is understood when applied to genetic phenomena (“information as reality”). In these terms the notion of genetic causality and genetic information are interchangeable, to the extent that they are intended in these very special senses.

**7. Testing the Theory: the notion of genetic error**

A possible test to see how the proposed account fares in explaining the association of genetic phenomena with informational facts is represented by the notion of genetic “error”.

Both Lewis and Woodward’s as well as Waters’ notion of influence are based on the notion of a pattern of counterfactual dependence. Whereas the standard counterfactual account of causality – and, all the more, other unsophisticated accounts – is silent about the consequences of a counterfactual whose antecedent is false (or, more generally, about the consequences of causal failure), Lewis’ pattern of counterfactuals C1 🞎🡪 A1, C2 🞎🡪 A2, C3 🞎🡪 A3, … Cn 🞎🡪 An (where Ci stands for a specific codon and Ai for a specific amino acid) or Woodward’s influence pattern, can tell you not only what happens when a given nucleotide triplet is expressed but also what would happen if the nucleotide sequence would be replaced by another one because of “misspelling” errors (i.e. errors in duplication, transcription, or translation). This recalls one of the distinctions drawn by Šustar with reference to DNA translation and enzymatic activity: whereas changes in the DNA sequence produce predictable changes in the protein product, by contrast, when the enzyme is changed, the enzyme might function no longer or produce “unexpected” effects. This unpredictability of the effect is precisely linked to a lack of systematic associations between possible alternative changes of the enzyme and possible alternative effects of these changes. This also accounts for Stegmann’s claim that “customary notions of genetic information involve a strong form of error which is evaluative and non-statistical” (2009: 8) and that “genetic information can be expressed correctly or incorrectly in a sense not adequately captured by uncommon effects” (2005: 429). In fact whenever you have a given triplet, then the amino acid is *necessarily determined*. For instance, if the triplet is CUA, then the resulting amino acid is invariably leucine. If a different amino acid is produced, then at least the linear order of the bases must have changed. A duplication error which would substitute the first nucleotide C (cytosine) in the triplet CUA with the nucleotide A (adenine) would result in the triplet AUA and thus invariably produce isoleucine instead of leucine. The systematic relationship between nucleotide sequences and amino-acids allows for the production of a well-defined macroscopic “change in the large-scale hereditary characteristics of the organism” (Schrödinger, 1945: 77) by means of a point difference in the DNA. This characterizes genetic causality with respect to other kinds of causal phenomena in the biological realm or to other entropy related phenomena such as thermodynamic laws, in that point changes in “information as reality” at one level produce systematic changes in “information as reality” at a higher level.

1. **Summary and conclusion**

Criticisms against informational talk with reference to genetic phenomena are mainly raised on two grounds. Developmental systems theorists object to a simplistic nature/nurture distinction based on the idea that only genes have semantic properties: “genetic privilege” is refuted on account of the semantic properties of epigenetic phenomena and other phenomena related to the inheritance of environmental features (Griffiths, 2001; Griffiths and Gray 1994; Oyama 1985; Griffiths and Knight 1998). Other philosophers of biology argue for the lack of explanatory value of informational talk with respect to main topics of interest in biology (Godfrey-Smith, 2000, Boniolo 2003, 2008). So, whereas developmental systems theorists would be ready to concede that information is carried by environmental factors as well, and refuse its use only to the extent that it is employed for making a qualitative distinction between genes and the environment, deflationists argue instead that information discourse masks nothing else than correlational relationships and thus should be replaced by causation.

This paper addressed both challenges in three steps. First of all, the common denominators shared by earlier and more recent defenders of the notion of genetic information have been identified. These include: 1) the *combinatorial mechanism* of the genetic code, 2) code *arbitrariness*, and 3) *causal specificity* (the systematic relationship of genetic and phenotypic traits). Genetic information is thus seen as an arbitrary but systematic correspondence between *combination of units* in one system and *units* in another system. In a second step, genetic causality has been distinguished from other kinds of developmental phenomena by drawing on the analysis of genetic causation developed by Šustar (2007), Waters (2007) and Woodward (2010), following which I defined genetic causation as a kind of cause which causes something to be one way rather than another (causal specificity). In a third step, the combinatorial mechanism and code arbitrariness have been compared to the linguistic phenomenon of double articulation which points to the relevance of linear order for semantic effects produced at different linguistic levels. This characterizes nucleotide sequences as “information as reality” in that structure and relations among them predominate over other features such as energy and matter, as causal determinants. Furthermore, such “information as reality” is able to produces “information as reality” at higher levels. Hence, genetic causality has been defined as a special kind of cause which causes something to be one way rather than another, by combining elementary units one way rather than another. Finally, this definition has been tested against the notion of “genetic error” in comparison with other kinds of causal failures in non-genetic phenomena.

 **References**

Barbieri, M. (2003) Biology with Information and Meaning. History and Philosophy of Life Sciences, 25: pp. 243-254.

Bergstrom C.T., M. Rosvall (2011) The Transmission Sense od Information (2011) Biology and Philosophy (26): pp. 159-176.

Boniolo, G. (2008) L’informazione biologica. In G. Boniolo and S. Giaimo (eds) Filosofia e scienze della vita. Milano: Mondadori: pp 198-219.

Boniolo G. (2003) Biology without information. History and Philosophy of Life Sciences, 25: pp. 255-273.

Borgmann A. (1999) Holding On to Reality. The University of Chicago Press.

Crick F. (1958) On Protein Synthesis. Symposia of the Society for Experimental Biology. 12: pp. 138-163.

Danchin, A. (2002) The Delphic Boat. What genomes tell us. Cambridge, Mass: Harvard University Press

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

Dawkins R. (1982) The extended phenotype: the long reach of the gene. Oxford: Oxford University Press.

Fantini B. (1990) Introduction to: Per un’etica della conoscenza (Italian edition of Monod, J. (1988) Pour une éthique de la connaissance). Turin : Bollati & Boringhieri.

Floridi L. (2002) What Is the Philosophy of Information? Metaphilosophy, 33: pp. 123-145.

Floridi L. (2011) Semantic Conceptions of Information. The Stanford Encyclopedia of Philosophy (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/information-semantic/>.

Gannett L. (1999) What’s in a cause? The Pragmatic Dimensions of Genetic Explanations. Biology and 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. (2007) [Information in Biology.](http://www.people.fas.harvard.edu/~pgs/InfoBio-PGS-CUP07.pdf%22%20%5Ct%20%22_blank) In David L. Hull & Michael Ruse (eds.) The Cambridge Companion to the Philosophy of Biology. Cambridge: Cambridge University Press.

Godfrey-Smith P. and Sterelny K. (2008) Biological Information. The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/information-biological/>.

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

Griffiths P. E. (2006) The Fearless Vampire Conservator: Philip Kitcher, Genetic Determinism and the Informational Gene. In: E. M. Neumann-Held and C. Rehmann-Sutter, C (Eds). Genes in development: Rereading the Molecular Paradigm: Duke University Press, PP.175-198.

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

Griffiths P. E., R. D. Knight (1998) What is the developmentalist Challenge? Philosophy of Science, 65 (2): pp. 253-258.

Itzkovitz S. U. Alon (2007) The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res (17): pp. 405-12.

Kendler K. (2005) A gene for …: the nature of gene action in psychiatric disorders. Am J Psychiatry, 162: pp. 1243-1252.

Kitcher, P. (1984) 1953 and All That. A Tale of Two Sciences. The philosophical Review, 93 (83): pp. 335-373.

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

Lewis D. (1973a), Counterfactuals. Oxford: Blackwell Publishers and Cambridge: Harvard University Press, 1973, Reprinted with revisions, 1986.

Lewis D. (1973b) Causation. The Journal of Philosophy, 70 (17): pp. 556–567.

Lewis D. (2000) Causation as influence. The Journal of Philosophy, 97 (4): pp. 182-197.

Lewontin R. (1974) The analysis of variance and the analysis of causes. American Journal of Human Genetics, 26: pp. 400-411.

Mackie, J. L. (1988) The Cement of the Universe: A study in Causation. Clarendon Press, Oxford, England.

Mahner. M. and M. Bunge (1997), Foundations of Biophilosophy. Berlin. Heidelberg, New York: Springer-Verlag

Martinet, A. (1960) Éléments de linguistique générale. Paris: Armand Colin.

Maynard Smith, J. (2000) The Concept of Information in Biology. Philosophy of Science, 67: pp. 177-194.

Maynard Smith J., E. Szathmáhry (1995) The Major Transitions in Evolution. Oxford: Oxford University Press.

[Maynard Smith](http://cscs.umich.edu/~crshalizi/reviews/authors.html#john-maynard-smith) J., [E. Szathmáry](http://cscs.umich.edu/~crshalizi/reviews/authors.html#eors-szathmary) (1999) The Origins of Life. From the Birth of Life to the Origin of Language. Oxford: Oxford University Press.

Mayr E. (1961) Cause and Effect in Biology. Science, 134 (3489): pp. 1501-1506

Mill, J.S. (1843) A System of Logic. Longmans Green, London (1949).

Mitarai N., K. Sneppen, S. Pedersen (2008) Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. Journal of Molecular Biology (382): pp. 236-245.

Monod J. (1971) Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology. New York: Alfred A. Knopf

Moss L. (1992) A Kernel of Truth: On the Reality of the Genetic Programme. In: D. Hull, M. Forbes, K. Okruhlik (eds.) PSA 1992, vol 1. East Lansing: Philosophy of Science Association: pp. 335-348.

Mueller-Wille S., H.J. Rheinberger (2009) Das Gen im Zeitalter der Postgenomik, Suhrkamp.

Oyama S. (1985) The Ontogeny of Information. Cambridge: Cambridge University Press.

Oyama S. (2000) Evolution’s Eye: A Systems View of the Biology-Culture Devide. Durham, NC: Duke University Press.

Perlmann G.E., J.M. Manning, P. Schimmel, (2006) Protein. In AccessScience@McGraw‐Hill,

http://www.accessscience.com.libproxy.mit.edu, DOI 10.1036/1097‐8542.550200, last modified: January 11, 2006. Accessed: 7 July 2013.

Prigogine I., I. Stengers (1984) Order out of Chaos: Man’s New Dialogue with Nature. New York: Bantam Books.

Sarkar S. (1996) Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology, In: S. Sarkar (ed.) The Philosophy and History of Molecular Biology. Dordrecht: Kluwer Academic Publishers: pp. 187-231;

Sarkar S. (2000) Information in Genetics and Developmental Biology: Comments on Maynard Smith. Philosophy of Science, 67: pp. 208-213.

Sarkar S. (2003) Genes Encode Information for Phenotypic Traits. In C. Hitchcock (ed.) Contemporary Debates in Philosophy of Science. London, Blakwell: pp. 259-272.

Schaffer, J. (2008) The Metaphysics of Causation. In: E. N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), URL = <http://plato.stanford.edu/archives/fall2008/entries/causation-metaphysics/>.

Schrödinger, E. (1945) What is Life? The Physical Aspect of the Living Cell. Cambridge: Cambridge University Press.

Shannon, C. E., W. Weaver (1949) The Mathematical Theory of Communication. Urbana: University of Illinois Press.

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

Shea N. (2013) Inhereted Representations are Read in Development. British Journal for The Philosophy of Science, (64): pp. 1-31.

Silverstein A. M. (1989) A History of Immunology. San Diego, CA: Academic Press.

Sørensen M., S. Pedersen (1991) Absolute in vivo translation rates of individual codons in *Escherichia Coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. Journal of Molecular Biology (222): pp. 265-280.

Stegmann U. (2004) The Arbitrariness of the Genetic Code. Biology and Philosophy, 19: pp. 205-222.

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

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

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

Sterelny K. (2000) The “Genetic Program” Program: A commentary on Maynard Smith on Information in Biology. Philosophy of Science, 67: pp. 195-2001.

Sterelny, K., P.E. Griffiths (1999) Sex and Death: An Introduction to the Philosophy of Biology. Chicago: University of Chicago Press

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

Šustar P. (2007) Crick’s Notion of Genetic Information and the ‘Central Dogma’ of Molecular Biology. British Journal for the Philosophy of Science, 58 (1): pp. 13-24. doi: 10.1093/bjps/axl018

Weber M. (2005) Philosophy of Experimental Biology. Cambridge. Cambridge University Press

Weber M. (2006) The Central Dogma as a Thesis of Causal Specificity. History and Philosophy of the Life Sciences, 28: pp. 595-610.

Weismann, A. (1892) Das Keimplasma: Eine Theorie der Vererbung. Jena: Fischer. Traduzione inglese: J.A. e M.R. Thomson (1904) The evolution Theory. London: Edward Arnold

Wicken J. S. (1987) Entropy and Information: Suggestions for a Common Language. Philosophy of Science, 54 (2): pp. 176-193.

Wilkins, J. (forthcoming) A deflationary account of information in biology: philsci-archive.pitt.edu/4834/01/genetic-info.doc

Williams, G.C. (1992) Natural Selection: Levels, Domains, and Challenges. Oxford: Oxford University Press.

Woodward J. (2003) Making Things Happen. A Theory of Causal explanation. Oxford: Oxford University Press.

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

1. According to Monod, it is not information *per se* which distinguishes organic from inorganic life – as advanced by Schrödinger – in that crystals also manifest structural information. Instead it is its overwhelming quantitative disparity in favour of living beings which makes the difference. [↑](#footnote-ref-1)
2. The former saw nature as a seamless continuum (*Natura non facit saltus*) where gradual and quantitative transitions resulted in apparent discrete species; the latter insisted instead in complete discontinuities between types. This dispute has involved scientists such as Buffon, von Nägeli, Karl Landsteiner, Paul Ehrlich, among many others, and was nurtured by metaphysical underpinnings (Gilson, 1971; Silverstein 1989: 92-93, Fantini 1990 p. XXIII – XXIV). Monod’s move is that of recognizing specificity while at the same time de-metaphysicizing it by devoiding it of any teleological interpretation. [↑](#footnote-ref-2)
3. Central to their comparison with the genetic code is the notion of degrees of freedom. It is this, together with code arbitrariness, which singles out the transmission of genetic traits across generations as a an informational phenomenon in the Shannon sense. Bergstrom and Rosvall also adduce empirical evidence showing that natural selection has taken into account such degrees of freedom in the of choice the actual code (Itkovitz and Alon, 2007), and benefits from it in various ways, e.g. by selecting nucleotide sequences so as to minimize “traffic congestions” during translation (by exploiting the difference in translation rates between synonymous codons: Sørensen and Pedersen, 1991, Mitarai et al. 2008). [↑](#footnote-ref-3)
4. Weber (2006: 599) reminds us that both the word “species” and “form” are common translations of Aristotles’ “eidos”. [↑](#footnote-ref-4)
5. In Woodward’s (2003) account, a sufficient condition for causation is that there is a possible intervention on X such that carrying out this intervention will change the value of Y or its probability distribution. Woodward’s claim is that “Causal relationships between variables […] carry a hypothetical or counterfactual commitment: they describe what the response of Y *would be if* a certain sort of change in the value of X were to occur” (Woodward, 2003: 40). Woodward’s definition of cause takes into account that causal relationships often occur in a web o causal links, so that the intervention might be masked by counteractive [↑](#footnote-ref-5)
6. Opposition to this distinction has a long tradition in the philosophy of science, starting with J.S. Mill (1843), who qualified it as non-scientific. However, Waters does not mean that the selection of causes is always based on identifying actual difference makers, but he objects to the idea that all causes are on an ontological par. Some causes are actual difference makers while others are not, regardless of the epistemic context. So the choice of a cause as the cause is not always and only contextual. [↑](#footnote-ref-6)
7. There is a continuum between the two, where epigenetic phenomena occupy an intermediate position. [↑](#footnote-ref-7)
8. This is also true with regard to the higher level of correspondence between amino acid sequences and proteins. The number of possible permutations of amino acid sequences in even so small a protein as insulin, with 51 amino acid residues, is enormous: 1051 permutations. Nevertheless, it has been established that the pancreatic cell of a given species has only one of these possible sequences (Perlmann, et al. 2006). In turn, the three-dimensional form of the protein, the secondary structure, is determined by the folding of the sequence of amino acids of which it is composed. [↑](#footnote-ref-8)