

Appeared in: *Noûs*, 48 (3): 450-465.

PENULTIMATE DRAFT

Causal control and genetic causation

Ulrich Stegmann

University of Aberdeen

School of Divinity, History and Philosophy

1. Introduction

We often make distinctions among causal relations. Sometimes we select one factor as ‘the’ cause of an effect while relegating others to the causal background. Some causal relations seem more fundamental and less accidental to us than others. And some causes generate a remarkable variety of specific effects, whereas others seem less specific. Several studies have argued that these distinctions are not simply ad hoc and pragmatic but trace ontological characteristics of the causal processes in question (Mitchell 2000, Woodward 2003, Weber 2006, 2006, Waters 2007, 2010).

Some studies take genetic causation as a test case, for good reason (Weber 2006, Waters 2007, Bogen & Machamer 2010, Woodward 2010). Genes and DNA are often portrayed, in both scientific and popular contexts, as exceptionally important causes which ‘control’ and ‘determine’ various molecular and developmental processes. Such portrayals have been rejected vociferously by developmental systems theorists, who argue that DNA is just one causal factor among many (e.g., Oyama 1985, Griffiths & Gray 1994, Moss 2003, Stotz 2006). Genetic causation is thus a promising area in which to elucidate characterisations like ‘control’, ‘determining’, and ‘being just one cause among many’, and to ask whether these pick out ontological differences among causal relations. What adds to the appeal of genetic causation as a test case is the fact that this domain of biology is rife with informational metaphors. The legitimacy or otherwise of these metaphors has generated a separate debate in philosophy of biology over the years (e.g., Sarkar 1996, Godfrey-Smith 2000, Maynard Smith 2000, Griffiths 2001, 2003, Stegmann 2005, Shea 2007, Bogen & Machamer 2010, Levy 2010). The relevance of this debate in the present context lies in the prospect of employing new causal concepts (such as ‘actual difference makers’ (Waters 2007) and ‘causal specificity’ (Woodward 2010) in order to reconstruct informational metaphors in purely causal terms (Weber 2006, Šustar 2007).

So far, analyses of causal relations have had little to say about the sense in which we take certain causes to ‘control’ processes.¹ Yet the idea of causes that control is ubiquitous in science and everyday life. In this paper I distinguish between two causal structures at the level of ensembles of cause-effect pairs. I argue that one of the structures captures one distinct sense in which causes can plausibly be construed as ‘controlling’ their effects. I then apply this account of causal control to genetic causation, arguing that claims about the controlling and informational role of DNA can be shown to have precise, empirical content.

2. Causation and causal specificity

The causal structure to be identified in this paper is best characterised in terms of Woodward’s (2003) manipulability account of causation. It also involves causal specificity in Woodward’s sense (2010). I therefore start by sketching, very briefly, the two central concepts, causation and causal specificity. For ease of exposition, I follow Woodward’s recent summary of his notion of causation:

“(M) X causes Y if and only if there are background circumstances B such that if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y or the probability distribution of Y would change.” (Woodward 2010, p. 290)

There are several features of Woodward’s account that need to be mentioned but will not be defended here (for Woodward’s defence see his 2003). For instance, Woodward’s theory is non-reductive insofar as it allows causal notions to enter his analysis of causation. Furthermore, Woodward takes the relata of causation to be variables, rather than events. This allows, among other things, to construe variations in causes as changes in the values of variables. One of the features that is worth highlighting for present purposes is that X’s causing Y does not imply that X is the only cause of Y in the sense that Y’s value depends exclusively on X’s value. For if one or more of the background circumstances were to change (variables other than X taking different values), Y’s value may cease to depend on X’s value. X’s causing Y implies, instead, that under certain background conditions intervening to change X’s value would be sufficient to change Y’s value. For instance, the match’s being lit causes fire because there are background circumstances (e.g. the presence of oxygen) in which intervening to change X’s value (from unlit to lit) changes whether or not a fire occurs. The intervention on X in this case is the striking of the match.

On Woodward's account, not any change in X that would change Y qualifies as an intervention. An intervention in the technical sense is a manipulation that changes Y as a result of only changing X, and in no other way. This is why striking match M is an intervention on X with respect to Y, but pouring petrol over the whole match box (containing M) and igniting it is not. Both are manipulations that set M alight (X) and change whether fire is present (Y). But the ignited petrol changes whether fire is present not only as a consequence of lighting M.

The fact that the lit match causes fire is compatible with there being circumstances in which it does not change the presence of fire, e.g. when there is insufficient oxygen. Roughly, the larger the set of background circumstances in which the X-Y relation obtains, the larger its degree of invariance or stability. Some causal relations are more invariant than others. So the degree of invariance of a given causal relation is one dimension along which it may differ from other causal relations (Woodward 2003, 2010).

Another dimension along which causal relations can diverge is their specificity. One sense of specificity is the degree to which the set of counterfactual dependencies between causes and effects is 'fine-grained'. With reference to Lewis' notion of causal influence, Woodward characterises this kind of specificity as follows:

“(INF) There are a number of different possible states of C ($c_1... c_n$), a number of different possible states of E ($e_1...e_m$) and a mapping F from C to E such that for many states of C each such state has a unique image under F in E (that is, F is a function or close to it, so that the same state of C is not associated with different states of E, either on the same or different occasions), not too many different states of C are mapped onto the same state of E and most states of E are the image under F of some state of C. This mapping F should describe patterns of counterfactual dependency between states of C and states of E that support interventionist counterfactuals. [...]” (Woodward 2010, p. 305).

Woodward's basic idea can be summarised as follows. Causal specificity is a relation between a cause variable and its effect variable (C and E, respectively), where each variable can take a range of values ($c_1... c_n$ and $e_1...e_m$, respectively). Specificity is a matter of degree. It depends on (a) the proportion of cause values that can change values of the effect variable, (b) the proportion of distinct effect values produced by changes to the cause variable, and (c) the proportion of effect values that can be changed as a result of changes to the cause variable. The degree of specificity of C with respect to E increases with these proportions. Woodward's

example of a causal relation with a high degree of specificity is choosing a radio station by turning the tuner. Many positions of the tuner (C) map to the receiver frequencies of different radio stations (E) and not too many different tuner positions map to the same radio station. Turning the tuner even slightly tends to change the station. Note also that INF suggests a maximal degree of specificity. Cause variables are maximally specific (in the sense of INF) with respect to E if all C-states map to E-states and vice versa (mapping is onto) and all C-states map to only one E-state and vice versa (1-1).²

3. External and internal causal ordering

With these notions in place, let us now consider an ordered set of effects, for example the triple $\langle F, G, H \rangle$. There are two ways in which the triple may arise. One possibility is that one effect becomes the cause of the next, i.e. F causes G and G causes H (we may assume that F is caused by another variable P). For example, pressing the push-button of an electric bell (F) closes an electric circuit (G) so that a current flows through the conductor (H). This sequence of events can be described as an ordered set of variables taking on certain values depending on the value of the previous variable. The effects G and H are caused by variables that are part of this set, i.e. they are caused by F and G, respectively. We can call this causal structure *internal ordering* or I-ordering, for short (fig. 1 a).

Another mode in which a triple of effects may arise is through cause variables that are not part of the set. In addition to effect variables F, G, and H as well as the external cause P, there are causes Q and R, such that P causes F, Q causes G, and R causes H. For example, barometer readings on three consecutive days (F, G, and H) depend on atmospheric pressure at the time (P, Q, and R) rather than the previous reading. Here the effects G and H are caused by variables that are not part of the triple (i.e. they are caused by Q and R, respectively). Let us call this causal structure *external ordering* or E-ordering (fig. 1 b).

In external ordering, the variables composing the product, here $\langle F, G, H \rangle$, are only effect variables with respect to one another, and the cause variables are not part of the product. No effect variable in an E-ordered process becomes the cause with respect to the variable next in the ordered set, unlike in I-ordering. There is consequently a strict separation between an entity that acts as a series of causes and an entity that is caused by it. E-ordered processes, here $\langle F, G, H \rangle$ taking certain values, are therefore processes in which every 'step' of the process is an effect of an external cause. It is natural to think of such external entities as 'controlling' the process they

cause because every step in such a process depends on the external entity. This sense of control is manifest in various technological artefacts.

Consider the devices categorised as “sequence controlled machines” (Brennecke 2000, p. 57): music boxes, automatic looms, and other machines (Ceruzzi 1983, Randell 1994). In a broad sense, any machine that passes through a sequence of steps in an ordered, repeatable, and automatic manner is sequence controlled. Even I-ordered devices like electric bells are sequence controlled in this sense. But when historians of technology single out music boxes, looms and some other devices as being sequence controlled, I suggest their classification traces sequence control in the more specific sense of E-ordering.³

Take a music box with a pegged, rotating cylinder. First note that the pegs cause individual tones in Woodward’s sense: there are background circumstances such that if we changed whether a peg is present or absent (its ‘value’), this would change whether or not a tone is played and, if it is played, which tone is played.⁴ Crucially, which tone is played at a particular point in time depends on which peg strikes the comb at that time, not on the tone played before. Since a peg is not part of the melody in the sense in which a tone is part of the melody, the series of tones is caused by external factors. In other words, the pegs externally order tone sequences. Another example are Jacquard looms. These looms used series of punchcards to raise and lower so-called Bolus hooks, which in turned raised and lowered a harness. Whether a given Bolus hook is raised or lowered at a particular point in time does not depend on which Bolus hooks were raised before. It rather depends on whether or not there is a hole at a certain position in the punchcard. And the hole in the punchcard is not part of the triple of Bolus hook states <F, G, H>. The sequence control exemplified by these machines is not simply the property of stepping through a sequence of states in a regular, repeatable manner. It is the more specific property of the machine’s operations being externally ordered.

Interestingly, music boxes and looms are regarded as precursors of modern day computers exactly because they are sequence controlled (Ceruzzi 1983, Randell 1994, Brennecke 2000). So it is not surprising to find the same kind of sequence control in some early computers. In the IBM ASCC (Harvard Mark I), for example, operations performed at a given point in time did not depend on the previous operation but rather on the pattern of holes in the linearly moving paper tape (called “control tape”, Harvard Computational Laboratory 1946, p. 11). Early computers like the ASCC, whose operations were specified by continually feeding a punched tape through the machine and which lacked branching orders, have been labelled “tape-controlled” computers

(Randell 1973, p. 352).⁵ Again, E-ordering captures the sense in which these tapes ‘control’ machine operations.⁶

Before applying this notion of control to the biological case, a few comments about the E-/I-distinction itself are in order. First, E- and I-ordering are the extreme ends of a spectrum. Within an otherwise E-ordered set there may be some effect variables which depend on variables in other positions. In addition, a given effect variable may itself cause other effects that do not belong to the ordered set (e.g. raised Bolus hook raise the harness). And some of these effects may in turn influence an effect variable within the set, breaking simple E-ordering (e.g. conditional branching orders in stored-programme computers, see footnote 12). On the other hand, within an otherwise I-ordered set some of the variables may not depend on variables in other positions.

Second, whether or not a given variable belongs to an ordered set is critical to E- and I-ordering. So far I have simply stipulated that, say, variable G is part of the effect triple $\langle F, G, H \rangle$, whereas R is not. Groups of variables in real-world systems may be causally and functionally integrated to a degree that they constitute the parts of an individual. What it takes for something to be part of an individual is a complex metaphysical issue. But, assuming that a collection of entities can reasonably be construed as forming the parts of an individual, it is a separate question whether the parts are E- or I-ordered. It is the latter issue that is at stake here. The question of external or internal ordering becomes an interesting empirical issue only on the assumption that the existence of a multi-component individual can be established on independent grounds.

The last comment concerns the relation between the E-/I-distinction and Woodward’s causal specificity. The degree of fine-grained counterfactual dependence (INF) and E-/I-ordering are distinct features of causal relations. Suppose that variables F, G, H, P, Q, and R can take two values each (f_1, f_2, g_1, g_2 , and so on) and that all cause-effect pairs exhibit maximal causal specificity (INF). That is, each value of a cause variable C maps to exactly one value of an effect variable E and vice versa (the relation is 1-1 and onto, or bijective): c_1 maps to e_1 and c_2 maps to e_2 . Under these conditions, intervening to change C’s value from c_1 to c_2 would change E’s value from e_1 to e_2 . This would hold for all cause-effect pairs irrespective of whether they are part of an I- or an E-ordered process. If the triple of effects $\langle F, G, H \rangle$ were I-ordered, then F’s taking on a given value would affect all values of the other members of the triple. If F took on value f_1 , then G would take value g_1 and H would take value h_1 . On the other hand, if F took value f_2 , then G would take value g_2 and H would take value h_2 . By contrast, if the triple $\langle F, G, H \rangle$ were E-

ordered, then F 's taking on f_i would have no consequences on the values of G and H . This would be the case despite all cause-effect pairs in both triples sharing the same degree of causal specificity.

In this section I have argued that we should understand a familiar and ubiquitous sense of 'control' in terms of a certain causal structure. The rest of the paper applies this exposition of control to a particular class of biological processes. This will shed light on two contested issues in philosophy of biology: the causal role of DNA and its status as information carrier. I start with the causal role of DNA.

4. DNA: from causal specificity to external ordering

Recent work on the causal role of DNA has paid careful attention to *kinds* of causal relations (Weber 2006, Waters 2007, Woodward 2010). In order to see what is involved in claims about the 'special' causal role of DNA, it is useful to begin by considering its role in a particular molecular process, DNA replication.

The textbook account of DNA replication goes something like this. A new strand of DNA (daughter strand) is synthesised by successively adding new components (nucleotides) to a linear chain. Several causal factors are involved, chief among them a pre-existing DNA strand (parent strand) and the DNA polymerase, an enzyme. Yet, of all causes involved, only the parent strand determines or specifies the base sequence of the daughter strand. It acts as the template, i.e. as the entity that determines which of the four possible nucleotides is being added to the daughter strand. By contrast, the polymerase merely assists in adding the new components by catalysing the chemical reaction that binds them to the growing strand.

On closer inspection it is unclear what 'determining' or 'specifying' consist in. It cannot mean counterfactual dependence simpliciter, because the daughter strand depends on both the template and polymerase. Adding an adenine to the growing daughter strand counterfactually depends on there being a thymine in the parent strand (if there had not been a thymine, a base other than adenine would have been added). But crucially, it also depends counterfactually on the presence of DNA polymerase, because adenine would not have been incorporated without it. Adding adenine depends, furthermore, on a host of other factors and conditions, such as the facts underlying Watson-Crick base-pairing (complementarity, a set of four bases, and so on). Suppose the thymine-adenine rule was to change to thymine-thymine (assuming this would be chemically possible), then the thymine in the template would yield another thymine, not an adenine. Hence, the fact that adenine is being added also depends counterfactually on the base

pairing rule being thymine-adenine. In short, the addition of adenine depends not only on the DNA template. But once all this is granted, what justification remains for viewing the template as ‘determining’ the daughter strand?

Woodward’s (2010) answer to the puzzle is to analyse ‘determining’ and ‘specifying’ in terms of causal specificity (INF). A variable C_1 is a specific cause of an effect to the extent that the effect depends in a fine-grained way on C_1 . And C_1 is a more specific cause than C_2 just in case the degree to which C_1 causes the effect in a fine-grained way is greater than that of C_2 . This, Woodward continues, holds in the case of processes like transcription (RNA synthesis). In transcription, many variations in the DNA template will cause corresponding variations in the RNA sequence, whereas we cannot intervene on the RNA polymerase, of which there is just one kind in prokaryotes, and cause the same degree of variation in RNA sequences. One may vary the concentration of polymerase, but this will only result in modifying the rate of synthesis, not in altering the sequences produced. In transcription, therefore, templates are the factor that is more causally specific than polymerase (a non-genetic factor). Note that Woodward takes the variables to be entire templates and their products (or sections thereof), not individual bases at certain positions within a template. The values of the cause variable ‘template’ (C_1) are token DNA templates with distinct base sequences, and the values of the effect variable ‘RNA sequence’ are token RNA product strands with different base sequences. ‘Polymerase’ is the other, less specific cause variable (C_2).

For present purposes it is worth pointing out that the difference in causal specificity between templates and polymerases also obtains at the level of individual bases within sequences. Let C and Γ be the two cause variables parent base and polymerase, respectively, and E the effect variable daughter base. C and E can take on five values, the bases adenine (subscript a), thymine (t), guanine (g), and cytosine (c) as well as s for absent (no base). The mapping between variables C and E is fairly specific (INF) insofar as most values of the cause variable map to exactly one value of E , and vice versa. That is, c_a maps to e_t , c_t maps to e_a , c_g maps to e_c , c_c maps to e_g (these mappings represent the Watson-Crick base pairings), and c_s maps to e_t , e_a , e_c , e_g , and e_s (in the absence of a template base, any or none of the four bases may be added to the new strand). So, if an intervention changed C ’s value from, say, c_a to c_g , then there would be a corresponding change in E ’s value (from e_t to e_c). The identity of the new base in the daughter strand thus depends in a fine-grained way on the base in the parent strand. By contrast, the dependence on the polymerase is not fine-grained. The polymerase Γ can take on two values (in prokaryotes): present (p) or absent (s).⁷ The mapping between Γ and E is such that γ_p maps to e_t ,

e_a , e_c , and e_g , and γ_s maps to e_s . Introducing a polymerase may lead to the addition of any of the four bases. The two kinds of causes therefore differ in their degree of causal specificity (INF) with respect to their effect, the newly added nucleotide.

However, templates and polymerases differ not only in their degree of causal specificity.⁸ As I will argue in the following paragraphs, they also differ insofar as templates externally order daughter sequences, whereas polymerases do not.

We saw that in the circumstances in which DNA replication occurs in organisms there is only one kind of causal factor on which daughter bases depend in a fine-grained way, and that factor is another base, not the polymerase.⁹ Furthermore, that base belongs to the parent strand. By contrast, the previously added base does not influence in any way the identity of the base added next. These facts make DNA replication an E-ordered process. In more detail, consider a three-unit sub-section of a newly synthesised DNA sequence, F-G-H, where F, G, and H denote the linearly arranged bases of the daughter strand. There are three bases (P, Q, and R) which are not part of the daughter strand, such that P causes F, Q causes G, R causes H (E-ordering). Since all variables stand for bases in either the parent or daughter strand, each variable can take on five values, i.e. the bases adenine (subscript a), thymine (t), guanine (g), or cytosine (c) as well as s for absent (no base). The mapping between variables is fairly specific (INF) insofar as most values of cause variables P, Q, and R map to exactly one value of a given effect variable F, G, and H, and vice versa. That is, the following mappings hold for the ordered set $\langle\langle P, F \rangle, \langle Q, G \rangle, \langle R, H \rangle\rangle$, where C is a placeholder for cause variables P, Q, and R, and E stands for the effect variables F, G, and H: c_a maps to e_t , c_t maps to e_a , c_g maps to e_c , c_c maps to e_g , and c_s maps to e_t , e_a , e_c , e_g , and e_s (see above). As we saw, if an intervention changed C's value from, say, c_a to c_g , then there would be a corresponding change in E's value (from e_t to e_c). Suppose the template took values $p_c-q_t-r_g$, then the newly synthesised section would take on the values $f_g-g_a-h_c$. If we intervened on the third cause variable R, changing its value from r_g to r_c , then the new section would be $f_g-g_a-h_g$.¹⁰

For questions of symmetry the crucial point is that the E-ordering entity in DNA replication is the DNA template. The polymerase is also an external cause of synthesis insofar as it is not part of the new daughter strand. But it is not the E-ordering entity, nor a part of it. So, the template differs from the polymerase (and other causally relevant factors) in the degree of causal specificity (INF), but it also differs with respect to E-ordering.¹¹

This finding generalises to other instances of template-directed syntheses: the E-ordering entities in these processes are the (DNA or RNA) templates, not the other causally relevant

factors. Consider protein synthesis, which is an E-ordered process because the addition of any given amino acid to the nascent polypeptide does not depend on the nature of any previously added amino acid. Protein synthesis proceeds in circumstances in which the mapping between mRNA codons and amino acids (the ‘genetic code’) persists. In these circumstances, the kind of amino acid integrated at a particular position depends in a fine-grained way only on the type of codon in the corresponding position in the mRNA. The amino acid sequence is therefore E-ordered by the mRNA base sequence.¹²

It is time to address a worry. It may seem as if the E-/I-distinction is contingent on how a process is described. Take DNA replication again. Successive rounds of replication yield a series of single DNA strands, in which every single strand templates the next.¹³ The nature of the new strand depends on the old, i.e. on the previous (terminal) element of this time sequence of DNA strands. So described, DNA replication may seem to be an I-ordered process. And it is. But in so describing replication, we simply consider a different product (effect variable). The ‘polymer’ we are now considering is a time series of single strands, not a spatial series of individual bases, and there is no inconsistency in the former being I-ordered but the latter E-ordered. The observation only shows that one must be clear about which effect variable is being considered.

DNA is a ‘special’ causal factor both in terms of its degree of causal specificity and in its ability to externally order causal processes. Perhaps DNA is also special in the sense that hardly any causes other than nucleic acid templates E-order molecular processes. This is an open empirical issue which cannot be fully settled here. But it is worth noting that even exemplars of non-genetic factors that may be said to ‘control’ development do not E-order the respective developmental processes. Take temperature-dependent sex determination. In some reptiles, temperature modulates the expression of a gene regulatory network implicated in gonad differentiation. One possible modulation mechanism is that colder temperatures trigger the expression of the tumor-repressor gene *Wt1*, which in turn affects the downstream gene regulatory network (Valenzuela 2008). If something like this is correct, then temperature does not E-order the series of steps in developing an organism’s sexual organs. Instead, the process involves partial dependence on previous steps: if the temperature had not been below a certain threshold, *Wt1* would not have been expressed; and without the expression of *Wt1* the regulatory network would not have been affected in the way it was, and so on. Examples like this suggest that, as a minimum, E-ordering is not spread *evenly* among genetic and non-genetic factors of development. Whatever the empirical findings, the notion of E-ordering provides a clear means of assessing and comparing claims about developmental control.¹⁴

In short, nucleic acids are a special causal factor because of their role as E-ordering entities.¹⁵ The next section explores the implications of this fact for the nature and legitimacy of a range of metaphors taken from information technology.

5. Varieties of causal relations: specificity, control, and information

Earlier it was noted that the DNA daughter strand depends not only on the template, but also on the polymerase because replication would not proceed without it. This generates the apparent symmetry in causal status between templates and polymerase. At this point molecular biologists would have been quick to emphasise that DNA polymerase is insensitive to the chemical nature of the nucleotides it conjoins; it does not influence which base is added. In other words, there is an asymmetry insofar as the *contrast* of adding (say) adenine, rather than a different base, depends exclusively on the parent base (whereas the *addition* of one base or another depends on both). This, I suggest, is the difference captured by causal specificity (INF): when applied at the level of *individual bases* in a template, causal specificity accounts for the sense in which bases are said to ‘determine’ or ‘specify’ exactly which new bases are added.

However, when *templates* are said to ‘control’, ‘specify’, or ‘direct’ product sequences, more is at work than a claim about the causal specificity of their components (INF). Genes and DNA have often and controversially been regarded as loci of control, a notion that has proven hard to explicate (cf. Weber 2005). To the extent that DNA (and RNA) templates E-order the synthesis of macromolecules, it is natural to regard them as ‘controlling’ each step in that process, in just the sense in which we take music boxes and looms to be sequence controlled. Template-directed synthesis is as externally ordered a process as operating a music box or a loom.

The similarity in causal structure can be pursued further. Externally ordered processes have E-ordering entities. In template-directed syntheses the E-ordering entities are templates, in music boxes, looms and punch-card machines they are pegged cylinders, perforated tape or stacks of punched cards. Whatever the many differences between these entities, they all share a causal role. It is then not surprising to find DNA templates being compared to punched cards and tapes¹⁶, and template-directed synthesis being conceptualised as a “tape-reading” processes¹⁷. Such metaphors have been popular since the 1950s, especially in the context of research into the basic mechanism of protein synthesis. They can be understood as highlighting a causal structure that neither professional nor lay audiences had encountered in the biological realm before the

1960s, but one with which they were familiar at the time from early computers and punch-card machines in research laboratories and business offices.

Philosophers of biology have focused their attention on another group of metaphors. These are metaphors describing DNA as an information carrier and a repository of instructions (see Kay 2000, for many examples). An explanation of these metaphors falls out of the considerations above: the machine-level instructions of many computers in the 1940s and 1950s contained, in common parlance, the information about what the computer will do, and they tended to E-order each step of a computer's operations. One example is the "control tape" of the ASCC (operational in 1946), which contained the machine-level instructions in the form of hole patterns (Harvard Computational Laboratory 1946). As the tape was fed into the machine, the holes externally ordered the series of operations (see Section 3).

Early stored-programme computers provide another source of examples. In computers like the EDVAC (1945-52) and IAS (1946-52), series of bit patterns were loaded from memory into the instruction register and executed one at a time (serially). The instructions for successive operations were by default located next to each other in memory. The computers therefore operated serially in the (additional) sense of stepping from one instruction to the spatially next. The kind of operation performed at any one point normally depended on the bit string presently in the instruction register, not on the previous operation or its computational result (Ceruzzi 1983).¹⁸ This suggests that in the IAS and similar early stored-programme computers the machine-level instructions (bit strings) E-ordered the sequence of operations to a significant extent.¹⁹ Note also that for several decades the programme of stored-programme computers was loaded into memory by encoding it on tape and then feeding it to the computer (Ceruzzi 1983, Williams 1997). Thus, DNA templates share the causal role of E-ordering with some of the early machine-level instructions and their storage media. Conceptualising DNA as a carrier of information and instruction can therefore be understood as a means of emphasising that shared role.

This understanding of informational metaphors in molecular biology allows their reconstruction in causal terms.²⁰ Suppose a bacterial polypeptide is said to have amino acid sequence *S* *because* of the information contained in sequence *T* of its mRNA template. On the present account, this explanation would amount to the following claims: *S* was produced by E-ordering, *T* was the E-ordering entity with respect to *S*, and the polypeptide has *S* because the template had *T*. The content of the informational explanation for the polypeptide's sequence being *S* is, in short, that it has *S* because it was E-ordered by an mRNA with *T*. In this

explanation information is supposed to explain an individual product sequence. This idea poses no difficulty if ‘T carries information for the production of S’ is analysed as ‘T externally orders the production of S’. The wider point here is that informational descriptions and explanations can play a specific and useful theoretical role. And this positive role stems from highlighting certain causal features, not from attributing semantic information or content.

The notion of E-ordering can also be employed to reconstruct Crick’s (1958) ‘central dogma’. The central dogma can be recast as a set of claims about the sort of molecules acting as E-ordering entities in the synthesis of nucleic acid strands and polypeptides: nucleic acids E-order the synthesis of other nucleic acids and of amino acid sequences, but amino acid sequences do not.²¹ Once the informational claims are rendered as causal claims about E-ordering, their truth or falsity becomes an empirical matter. Furthermore, since E-ordering comes in degrees, templates may also be said to carry information *to some extent*. As Sarkar (1996), Stotz (2006) and others emphasised, nucleic acid templates often have very little impact on product sequences. For example, DNA templates have little influence on amino acid sequences in eukaryotes due to RNA splicing and other mechanism. It therefore seems false to say that DNA carries information about those sequences. But to the extent that nucleic acids play an E-ordering role, informational metaphors are apt and their use justifiable²².

Two further features of the proposal are worth pointing out. First, the causal reconstruction in terms of E-ordering not only shows that a specific causal role happens to coincide with informational language; it also explains why *that* role in particular motivates informational descriptions. E-ordering motivates informational descriptions because the entities that tended to E-order the operations of early computers (certain machine-level instructions) were and are considered to be information carriers. Comparing a molecular process with the new information technology was an effective way of gesturing towards a causal structure for which there was no biological precedent at the time. It is this similarity in causal roles that can justify the use of information concepts.²³ Second, there are interesting differences between kinds of template-directed syntheses. In translation, for example, template and product elements belong to different biochemical classes and, arguably, are related arbitrarily to one another. In replication and transcription, by contrast, template and product elements belong to the same biochemical class (nucleic acids) and are not arbitrarily related. But according to the central dogma, these differences are irrelevant to the question of whether they carry information; they all do. This is what one would expect if information transfer is tied to E-ordering, as I have argued.²⁴

6. Conclusion

The idea of causes that ‘control’ processes is familiar, not least from machines in which every step appears to depend on entities that are not part of the outcome. The ‘sequence controlled’ operations of music boxes, automatic looms and punchcard machines are examples. I have argued that ‘control’ in this sense can be identified with a certain causal structure, external ordering. Since series of effects may vary in the degree to which they are externally ordered, this causal structure is one dimension along which causal relations can differ. It is a dimension distinct from variations in other causal features, such as causal specificity or invariance.

External ordering is particularly useful in analysing genetic causation.²⁵ Questions about the causal status of DNA (‘in what sense, if any, does it determine effects?’) and its status as information carrier (‘in what sense, if any, does it carry information?’) have sometimes spawned disconnected lines of inquiry in philosophy of biology. I suggest that the causal structure identified here provides an answer to both questions. Information-technological metaphors are based on a shared causal structure between certain artefacts and molecular templates, and this structure is unlikely to be present in many other biological processes.²⁶

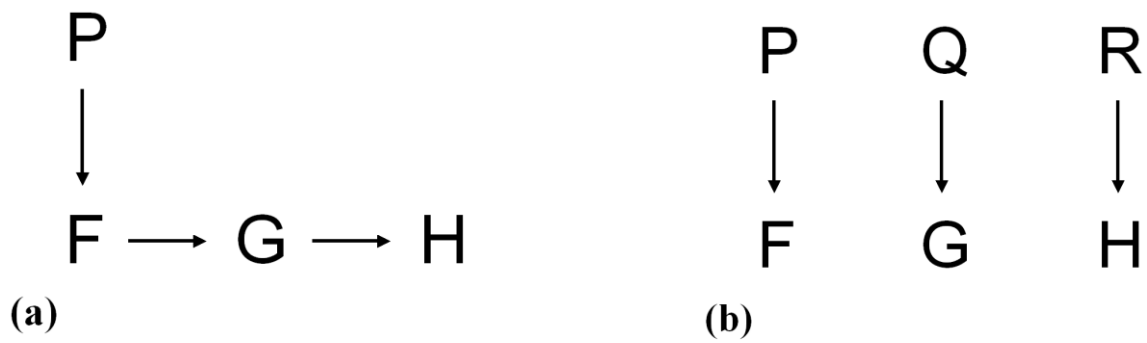


Figure 1. Causal graph of (a) internal ordering and (b) external ordering. Orderings are modes of generating an ordered set of variables, here the triple $\langle F, G, H \rangle$. Internal and external orderings are the two extremes of a spectrum. Arrows denote a causal relation in the sense of Woodward's (2010) notion of **M** (see text). Roughly, there are circumstances in which intervening on (say) P to change its value would change the value of F. In I-ordering, this would also change the values of G and H. But in E-ordering changing P's value would not change G's or H's values, because their causes (Q and R) do not depend in any way on P or F.

References

- Aspray, William. (1990). *John von Neumann and the Origin of Modern Computing*. Cambridge, MA: MIT Press.
- Bogen, Jim, & Machamer, Peter. (2010) "Mechanistic information and causal continuity". In *Causality in the Sciences*, eds. Phyllis McKay Illari, Federica Russo and Jon Williamson. Oxford: Oxford University Press, pp. 845-864.
- Bonner, James F. (1965). *The Molecular Biology of Development*. Oxford: Open University Press.
- Brennecke, Andreas. (2000) "A classification scheme for program controlled calculators". In *The First Computers: History and Architectures*, eds. Raul Rojas and Ulf Hashagen. Cambridge, MA: MIT Press, pp. 53-68.
- Ceruzzi, Paul E. (1983) *Reckoners: The Prehistory of the Digital Computer, from Relays to the Stored Program Concept, 1935-1945*. Westport, Connecticut: Greenwood Press.
- Ceruzzi, Paul E. (2003) *A History of Modern Computing*, 2nd edn. Cambridge, MA: MIT Press.
- Copeland, B. Jack. (Ed.) (2006) *Colossus: The Secrets of Bletchley Park's Codebreaking Computers*. Oxford: Oxford University Press.
- Crick, Francis H. (1958) "On protein synthesis". *Symposia of the Society for Experimental Biology*, **12**, 138-167.
- Darden, Lindley. (2006) "Flow of information in molecular biological mechanisms". *Biological Theory*, **1**(3), 280-287.
- Godfrey-Smith, Peter. (2000) "On the theoretical role of "genetic coding"". *Philosophy of Science*, **67**, 26-44.
- Griffiths, Paul E. (2001) "Genetic information: A metaphor in search of a theory". *Philosophy of Science*, **68**, 394-412.

- Griffiths, Paul E., & Gray, Russell. (1994) "Developmental systems and evolutionary explanation". *Journal of Philosophy*, **91**, 277-304.
- Harvard Computational Laboratory. (1946) "A manual of operation for the automatic sequence controlled calculator". *The Annals of the Computation Laboratory of Harvard University*, **1**, 1-561.
- IBM. (1953) "IBM 407 Accounting Machine, Manual of Operation". New York: IBM, pp. 1-210.
- Jacob, François. (1974) *The Logic of Living Systems - A History of Heredity*. London: Allen Lane.
- Kay, Lily E. (2000) *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford: Stanford University Press.
- Levy, Arnon. (2010) "Information in biology: A fictionalist account". *Noûs*, **45**(4), 640-657.
- Maynard Smith, John. (2000) "The concept of information in biology". *Philosophy of Science*, **67**, 177-194.
- Mitchell, Sandra D. (2000) "Dimensions of scientific law". *Philosophy of Science*, **67**, 242-265.
- Moss, Lenny. (2003) *What Genes Can't Do*. Cambridge, MA: MIT Press.
- Neumann-Held, Eva. (2006) "Genes - causes - codes: Deciphering DNA's ontological privilege". In *Genes in Development*, eds. Eva Neumann-Held and Christian Rehmann-Sutter. Durham: Duke University Press, pp. 238-271.
- Northcott, Robert. (2009) "Is actual difference making actually different?". *Journal of Philosophy*, **106**(11), 629-634.
- Oyama, Susan. (1985). *The Ontogeny of Information*. Cambridge: Cambridge University Press.
- Piccinini, Gualtiero. (2008) "Computers". *Pacific Philosophical Quarterly*, **89**, 32-73.

- Randell, Brian. (Ed.). (1973) *The Origins of Digital Computers: Selected Papers* (1st ed.). Heidelberg: Springer.
- Randell, Brian. (1994) "The origins of computer programming". *IEEE Annals of the History of Computing*, **16**(4), 6-14.
- Sarkar, Sahotra. (1996) "Biological information: A skeptical look at some central dogmas of molecular biology". In *The Philosophy and History of Molecular Biology: New Perspectives*, ed. Sahotra Sarkar. Dordrecht: Kluwer, pp. 187-231.
- Sarkar, Sahotra. (2003) "Genes encode information for phenotypic traits". In *Contemporary Debates in Philosophy of Science*, ed. Christopher Hitchcock. London: Blackwell, pp. 259-272.
- Shea, Nicholas. (2007) "Representation in the genome and in other inheritance systems". *Biology and Philosophy*, **22**, 313-331.
- Stegmann, Ulrich E. (2004) "The arbitrariness of the genetic code". *Biology and Philosophy*, **19**, 205-222.
- Stegmann, Ulrich E. (2005) "Genetic information as instructional content". *Philosophy of Science*, **72**, 425-443.
- Stotz, Karola. (2006) "With 'genes' like that, who needs an environment? Postgenomics's argument for the 'ontogeny of information'". *Philosophy of Science*, **73**, 905-917.
- Šustar, Predrag. (2007) "Crick's notion of genetic information and the 'central dogma' of molecular biology". *The British Journal of Philosophy of Science*, **58**(1), 13-24.
- Valenzuela, Nicole. (2008) "Evolution of the gene network underlying gonadogenesis in turtles with temperature dependent and genotypic sex determination". *Integrative and Comparative Biology*, **48**(4), 476-485.
- Waters, C. Kenneth. (2007) "Causes that make a difference". *Journal of Philosophy*, **104**, 551-579.

Weber, Marcel. (2005) *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.

Weber, Marcel. (2006) "The central dogma as a thesis of causal specificity". *History and Philosophy of the Life Sciences*, **28**, 595-609.

Williams, Michael R. (1997) *A History of Computing Technology*. Los Alamitos: IEEE Computer Society Press.

Woese, Carl R. (1967) *The Genetic Code: The Molecular Basis for Genetic Expression*. New York: Harper and Row.

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

Woodward, James. (2006) "Sensitive and insensitive causation". *Philosophical Review*, **115**, 1-50.

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

Endnotes

¹ I have in mind the sense of control exerted by causes like the pegged cylinders in music boxes. Woodward's (2010) notion of causal specificity (INF) explicates another sense of control.

² It is not entirely clear what should count as maximal causal specificity. In his main text, Woodward (2010, p. 305) appears to rely mostly on bijectivity. But the corresponding footnote emphasises "the "many different states" requirement in INF" (his footnote 17), potentially excluding two-valued variables from having any causal specificity. Yet these two features are (partially) independent of one another; they could be separate dimensions of specificity. In this paper I will count cause-effect pairs as maximally specific as long as they are bijective, even if the variables have very few distinct values.

³ Randell (1994) appears to suggest that sequence control is a matter of selecting one sequence of operations out of a range of possible sequences: "A most important art in this regard was that of means for [sic] *specifying a sequence of choices, among a set of possible actions*, in such a way that the machine could carry out the sequence completely automatically" (p. 6, my italics). The idea seems to be that a Jacquard loom, for instance, weaves different patterns depending on the kinds of punched cards it is fed, whereas other machines always perform the same sequence of operations. When pressing the button of an electric bell we could not make it perform different operations, short of modifying the bell's design. I think this characterisation of sequence control is unsatisfactory. First, finite state machines can undergo different sequences of operations depending on the kind of input they receive (e.g. a vending machine performs different operations depending on which coins are inserted and which command buttons are pressed). But there still seems to be a distinction between merely triggering one automatic machine sequence out of several possible sequences and the sort of sequence control achieved with pegged cylinders and punched cards. Second, the characterisation excludes machines with non-exchangeable cylinders or cards (early music boxes and tape-fed looms), because they could not perform different sequences of operations. Randell's characterisation glosses over the difference between sequence control and programmability.

⁴ The causal relation between pegs and tones only holds across a certain range of values of other cause variables. For instance, if the box were played in a vacuum, adding a peg would not add a tone. Furthermore, pegs are not the only causes of tones: the presence or absence of a vacuum is another, as is the thickness and length of the comb's teeth, and so on. Changing the thickness of

teeth would also change which tone is played. Nonetheless, changing the pegs would change tones across a range of values of other cause variables.

⁵ Other early computers, like the ENIAC, operated differently. In these computers, the programme the machine executed was defined by the configurations of patch cables and switch settings (e.g. Ceruzzi 1983, Copeland 2006).

⁶ The E-ordering entities in early stored-programme computers and the ASCC were machine-level instructions (Harvard Computational Laboratory 1946). But series of data can also E-order operations. For example, the IBM Accounting Machine 407 was fed one card at a time from a stack of cards whose hole pattern encoded commercially relevant data. The same kind of operation was performed on all cards in a stack, e.g. adding numbers. But the precise action performed by the machine depended on the holes in the card currently being read, not on the nature of the previous operations (IBM 1953). The stack of punch-cards or, more precisely, the hole patterns in the cards, therefore acted as the E-ordering entity with respect to the temporal sequence of operations. By re-wiring a plugboard, the IBM 407 could perform a different kind of task on a given stack of cards, e.g. printing data. This made it unnecessary to carry the stack of cards to another special-purpose machine (e.g. collators, tabulators, reproducers), as had been the case with previous punched card machines (Ceruzzi 2003).

⁷ Alternatively we might choose the concentration of polymerase as the variable, in which case it can take on a range of different values. But the outcome with respect to DNA templates would be the same.

⁸ Causal specificity alone is unlikely to separate genetic from non-genetic factors of development. In protein synthesis, for example, DNA is less causally specific than some of the enzymes involved. Aminoacyl-tRNA-synthetases (aaTS) are the enzymes catalysing the ‘charging’ of tRNAs with their corresponding amino acids. The first step in this process is binding the amino acid to the corresponding aaTS. Most cells make twenty aaRSs, one for each amino acid. The mapping of aaTSs to the twenty amino acids is (nearly) bijective. If we intervened on the amino acid-binding site to make it complementary to a new kind of amino acid, then a given tRNA would be loaded with the new amino acid. By contrast, DNA triplets map to amino acids many-to-one (degeneracy of the genetic code). DNA is therefore less causally specific than aaTSs with respect to amino acids. (See Weber 2006 for a similar argument regarding RNAs).

⁹ In a distant possible world in which the Watson-Crick base pairing rules keep changing from one base to the next during replication, the bases in the daughter strand also (or perhaps exclusively) depend in a fine-grained way on the value of the base-to-base mapping. However, there being such a possible world is irrelevant to the question of which causal factor E-orders the daughter strand in the *actual* world.

¹⁰ Compare template-directed synthesis with the assembly of protofilaments from individual monomers (actin molecules), a clear example of I-ordering. The terminal monomer specifies the identity of the next, because the next monomer needs to be complementary to the previous one. Consider a three-unit sub-section of a newly synthesised polymer, F-G-H, where F, G, and H denote the linearly arranged monomers. Protofilament assembly is I-ordered, so F causes G and G causes H. Let f_a stand for an actin molecule at first position, g_a an actin molecule at second position and h_a and actin molecule at third position. Then if F took the value f_a , G would take the value g_a and H would take the value h_a . The new polymer would be $f_a-g_a-h_a$. If we intervened on the second cause variable G and changed its value from g_a to g_b , where b denotes an actin molecule with altered surface properties, then protofilament assembly might come to a halt because b would lack the complementarity required to attaching another actin molecule. The new polymer would be f_a-g_a .

¹¹ Waters' (2007) account of the relation between DNA and polymerases is different. He argues that the alleged symmetry presupposes the widespread but false assumption that there is no ontological difference between causes and conditions. But they are distinct, he insists. On Waters' account of the difference it turns out that what biologists identify as 'the' cause of the daughter strand sequence, the template, corresponds to a special kind of cause ('the actual difference maker'), whereas factors like polymerase do not. However, as Waters points out himself, his account is only applicable to a *population* of effects that manifest a difference. There are no difference making causes with respect to a set of uniform effects (or with respect to a single effect). He regards this implication as unproblematic. But note that molecular biologists regard templates as the sequence-determining factor even when considering the synthesis of a single polymer. What they regard as the determining cause in template-directed synthesis does not appear to track actual difference making causes. So, even if Waters account of the ontological distinction is correct (see Northcott 2009 for an objection), it is doubtful whether it can shed light on the sense in which molecular biologists causally privilege DNA in the context of template-directed synthesis. I will pursue this concern in detail elsewhere.

¹² The code is the result of specificities of hydrogen bonding between codon and anticodon and of the specificities of various types of bonds between aminoacyl-tRNA-synthetase, on the one hand, and tRNAs and amino acids, on the other hand. Changing some of these specificities may change which amino acid is incorporated in response to a given codon. Molecules like tRNA and their specificities are therefore also cause variables with respect to the addition of amino acids, and at least some changes in some of these variables will have a fine-grained effect on the effect variable. However, they are not the E-ordering factors in the circumstances in which protein synthesis occurs in organisms, because these are circumstances in which the genetic code (and the underlying specificities) persist during synthesis. In addition, it is hard to see how the ordered set of (say) tRNAs involved in protein synthesis could constitute an individual in the way in which the ordered set of codons constitutes an individual (i.e., an mRNA). Maintaining that in actual organisms the E-ordering factors are templates is compatible with there being possible worlds in which other molecules satisfy this role, or none do. Neumann-Held (2006) provides an interesting thought experiment that can be interpreted along these lines.

¹³ Of course, replication does not yield ‘single strands’ in the sense that daughter strands remain isolated from their parent strands.

¹⁴ For instance, DNA is unlikely to control gene regulation and downstream developmental processes in the sense of E-ordering. Successive rounds of switching on and off genes seem closer to I-ordering (although morphogen gradients may be interesting exceptions). Again, these are open empirical issues that require further investigation.

¹⁵ I take E-ordering to be an explication of what I referred to earlier as “advance specification” (Stegmann, 2005).

¹⁶ DNA has been compared to a series of “punch-cards” (Lederberg 1955, quoted in Kay 2000, p. 114), a “punched tape” (Bonner 1965), a “magnetic tape” (Jacob 1974), and so on.

¹⁷ Woese (1967, p. 5) offers an extensive description of template-directed syntheses in terms of “tape-reading”, in which “output tapes” (new nucleic acid and amino acid sequences) are generated by feeding “input tapes” (DNA and RNA templates) linearly through “tape readers”.

¹⁸ Writing about the IAS computer, Ceruzzi (1983, p. 140) observes: “The address of the next instruction was likewise not given; it was assumed to be in the memory location right next to the instruction just executed. Thus although the IAS machine was a stored-program computer, it normally executed instructions in a steady linear stream coming from the memory, just as if they were coming off a tape”.

¹⁹ Most modern desktop computers are von Neumann machines, whose design principle was established by the EDVAC and the IAS (Randell 1973, Ceruzzi 1983, Aspray 1990). It is also clear, however, that modern computers are not normally E-ordered due to conditional branching and other features in which the decision about which instruction to execute next depends on the outcome of the previous step. Indeed, much of the power and versatility of SPCs stems from such features (Aspray 1990), which break strict E-ordering. The computation of functions arguably *requires* this dependence on previous operations and their outcomes (Piccinini 2008). Note, however, that the E-/I-distinction remains valuable for characterising causal structures that are intermediates between E- and I-ordered processes, including transfer orders, conditional branching, and so on.

²⁰ Obviously not all uses of informational metaphors in molecular or developmental biology should be construed in terms of E-ordering. But I suggest that E-ordering makes sense of how some of these metaphors were used in research on the mechanism of protein synthesis and, more generally, in the context of template-directed synthesis (e.g. in Crick's 'central dogma').

²¹ Weber (2006) has offered an alternative causal reconstruction of Crick's 'central dogma'. He draws on Waters (2007) notion of actual difference makers to argue that Crick's dogma should be construed as the claim that "protein is not the actual difference-making cause in protein synthesis with respect to a cell's population of proteins or in nucleic acid synthesis with respect to a cell's nucleic acid population" (Weber 2006, p. 605). This reconstruction inherits the limitation of Water's difference makers: it is difficult to reconstruct informational claims about *individual* templates and their products in this way (see footnote 8).

²² Similar causal roles do not imply that nucleic acids carry information or instructions in a demanding, semantic sense.

²³ Other causal reconstructions (Weber 2006, Šustar 2007) leave it open why only some causal relations invited and sustained information talk.

²⁴ Accounts in which features like arbitrariness are important (e.g. Godfrey-Smith 2000) have a difficulty explaining this characteristic of information talk (Stegmann 2004, 2005, Darden 2006).

²⁵ External ordering may also be useful in elucidating the notion of programmability as well as the relation between algorithms and the kinds of artifacts often used to illustrate them (e.g. cooking recipes).

²⁶ This paper has had a long gestation period. Early versions were presented in Birmingham, Bristol, Groningen, Leeds, Oslo, and at the 2007 meeting of the ISHPSSB, and I am grateful for

the feedback from the audiences. I owe special thanks to Paul Griffiths, Tim Lewens, Ulrich Krohs, David Papineau, and Stuart Presnell for their queries on earlier versions. Marcel Weber, Gerry Hough and an anonymous referee provided helpful comments on the last draft.