

Genetic, epigenetic and exogenetic information in development and evolution

Paul E. Griffiths

*Department of Philosophy and Charles Perkins Centre, University of Sydney, NSW
2006, Australia. paul.griffiths@sydney.edu.au*

*Interface Focus 2017 7 20160152; DOI: 10.1098/rsfs.2016.0152.
Published 18 August 2017*

Abstract

The idea that development is the expression of information accumulated during evolution and that heredity is the transmission of this information is surprisingly hard to cash out in strict, scientific terms. This paper seeks to do so using the sense of information introduced by Francis Crick in his sequence hypothesis and central dogma of molecular biology. It focuses on Crick's idea of precise determination. This is analysed using an information theoretic measure of causal specificity. This allows us to reconstruct some of Crick's claims about information in transcription and translation. Crick's approach to information has natural extensions to non-coding regions of DNA, to epigenetic marks, and to the genetic or environmental upstream causes of those epigenetic marks. Epigenetic information cannot be reduced to genetic information. The existence of biological information in epigenetic and exogenetic factors is relevant to evolution as well as to development.

Keywords: Genetic information, Epigenetics, Specificity

1. Genetic Information

That the development of evolved characteristics is the expression of information accumulated in the genome during evolution and that heredity is the transmission of this information from one generation to the next will strike most biologists as common-sense. But it is surprisingly difficult to cash out this statement in a way that is grounded in the detailed theory and practice of the biosciences¹. Biology today is certainly an ‘information science’, both because it is a science of big data and because many specific models are inspired by the information sciences, but these applications and models do not seem to be unified by a single conception of biological information. If the actual science straightforwardly corresponded to that opening statement, we would expect to find that instructions written in the genetic code are read by gene regulatory networks to make an organism. But the genetic code runs out of steam when it has specified the linear structure of proteins [2]. It is impossible to describe higher levels of biological organisation in the genetic code for the same reason that I cannot write literature using a geodetic coordinate system: the language does not have the expressive power. Nor is it easy to see how the expressive power of the genetic code could be expanded to describe something beyond the order of amino acids in a polypeptide. The ‘histone codes’ [3] and ‘splicing codes’ [4] that have been proposed as supplements to the genetic code are not integrated with the genetic code through a shared measure of coded information. As things stand, histone modification and mRNA splicing are molecular mechanisms that interact with the mechanisms of transcription and translation in the straightforward way that any combination of physical mechanisms can interact. This paper outlines a measure of information that allows us to compare the contributions made by each of these mechanisms to determining a final product in a shared, informational currency.

Turning our attention to gene regulatory networks, these are productively modeled as computing Boolean functions and/or differential equations, but these computational operations are not specified in any of the three ‘codes’ to which we just referred. Instead, these operations are specified by the stereochemical affinities of genomic regions and gene products. The science

¹In his final book the influential evolutionary theorist George C. Williams called for a new, ‘codical’ biology founded on the concept of information precisely because that is not the biology we actually have [1].

34 that connects the ‘codes’ with the ‘computing networks’ is the physics of how
35 stereochemical properties emerge from the linear structure of biomolecules
36 and the cellular contexts in which those biomolecules mature and function.
37 The same is true of the other molecular networks that are at the heart of
38 our understanding of the cell – when we model these networks as performing
39 computations those formal operations do not take as inputs representations
40 written in the genetic code.

41 All this suggests that perhaps ‘biology is an information science’ only in
42 the sense that it uses many models that start with analogies to some aspect
43 of communication or computing, and makes many direct applications of for-
44 malisms from the information sciences. Each of these models or applications
45 stands or falls on its own scientific merits. They do not link together to form
46 a single theory of biological information or a theory of life as an informational
47 phenomenon [5] [6][7][2]. On this sceptical view the ubiquity of information
48 talk in biology is only evidence of the power and generality of theories of in-
49 formation and computation, something we can observe in many other areas
50 of science.

51 This paper defends a more robust view of biological information, however.
52 It argues that there is an important sense of ‘information’ which is related
53 very closely to the older notion of biological ‘specificity’. Biological informa-
54 tion in this sense gives scientific substance to the claim that development is
55 the expression of information accumulated during evolution, and that hered-
56 ity is the transmission of this information from one generation to the next.
57 These claims turns out to be more or less equivalent to the idea that heredity
58 is the ability of one cell to transmit biological specificity to another and that
59 development is the expression of that specificity in a controlled manner.

60 The paper builds on Paul Griffiths and Karola Stotz’s ‘bottom-up’ ap-
61 proach to biological information, starting with a simple concept of informa-
62 tion that plays a straightforward role at the heart of molecular biology and
63 seeing how many other aspects of biology can be clarified by applying this
64 sense of information. That starting point is what they termed ‘Crick infor-
65 mation’, the sense of information introduced by Francis Crick (1958) in his
66 ‘sequence hypothesis’ and ‘central dogma of molecular biology’ [8][9]²

²Griffiths and Stotz used the phrase ‘Crick information’ to refer to what, in this article, will be called ‘sequence specificity’. In more recent work I and my collaborators have reserved the term ‘Crick information’ for a measure of the intrinsic information content of a sequence, rather than for the measure of the relationship between a sequence and its

67 Given the central role of Crick’s ideas in molecular biology it is surprising
68 that previous efforts to explicate the idea of biological information have not
69 adopted Crick’s straightforward approach. Instead, they have mostly focused
70 on the richer connotations of the term ‘information’: ideas like meaning,
71 representation, and semiosis.³ Some authors have even attributed this rich
72 sense of information to Crick: “The sense of information relevant to the
73 central dogma is of course the sort which requires ‘intentionality’, ‘aboutness’,
74 ‘content’, the representation of other states of affairs...” [13][pp. 550-1].
75 As we will see in the next section, nothing could be further from Crick’s
76 intentions. The problem with rich approaches to biological information is
77 that we do not have developed, technical theories of information in this sense.
78 The various terms used in the passage just cited are, as the author admits,
79 merely “one or another facet of a philosophically vexed concept” [13][p. 151].
80 So the approach amounts to taking this vexed concept, for which we have no
81 developed theory, and placing it at the foundations of an account of living
82 systems. In this paper, in contrast, we will use only the standard formalism
83 of information theory and the idea of biological specificity.

84 2. Crick’s conception of information

85 The key move made by Crick in his work on protein synthesis was to
86 supplement the existing idea of stereochemical specificity, embodied in the
87 three-dimensional structure of biomolecules and underlying the well-known
88 lock-and-key model of interaction between enzymes and their substrates, with
89 the idea of informational specificity, embodied in the linear structure of nu-
90 cleic acids that determine the linear structure of a gene product [14][5]. This
91 idea is present in Crick’s statements of both the sequence hypothesis, and
92 the central dogma (Figure 1):

93 *The Sequence Hypothesis* ... In its simplest form it assumes that
94 the specificity of a piece of nucleic acid is expressed solely by the
95 sequence of its bases, and that this sequence is a (simple) code
96 for the amino acid sequence of a particular protein.

causes that is the subject of this article.

³Sahotra Sarkar [5] gives a brief history of efforts by molecular biologists to construct a theory of biological information. Key papers in philosophical literature are [10][11]. For ‘biosemiotics’ see [12]

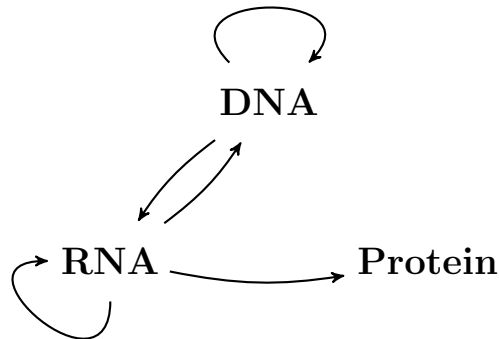


Figure 1: The Central Dogma, as it is held today. After [16], with modifications. In particular, an arrow from DNA to protein has been removed.

97 *The Central Dogma* This states that once ‘information’ has passed
 98 into a protein *it cannot get out again*. In more detail, the transfer
 99 of information from nucleic acid to protein may be possible, but
 100 transfer from protein to protein, or from protein to nucleic acid
 101 is impossible. Information means here the precise determination
 102 of sequence, either of bases in the nucleic acid or of amino-acid
 103 residues in the protein. [15][pp. 152-153, italics in original]

104 According to Crick the process of protein synthesis involves “the flow of
 105 energy, the flow of matter, and the flow of information.” While noting the
 106 importance of the “exact chemical steps”, he separated this transfer of mat-
 107 ter and energy from what he regarded as “the crux of the problem”, namely
 108 how to join the amino acids in the right order – “the crucial act of sequen-
 109 tialization.” His solution to this problem would “particularly emphasise the
 110 flow of information” where “By information I mean the specification of the
 111 amino acid sequence of the protein” [15][144].

112 Crick maintained the same, straightforward view of information through-
 113 out his career. In his well-known paper clarifying the central dogma he
 114 reiterated that his key achievement in 1958 was to reduce the problem of
 115 protein synthesis to “the formulation of the general rules for information
 116 transfer from one polymer with a defined alphabet to another.” [16][561]
 117 Information is a causal concept, referring simply to precise determination.
 118 Crick reiterated this forty years later: “...‘Information’ in the DNA, RNA,

119 protein sense is merely a convenient shorthand for the underlying causal ef-
120 fect.” (Crick to Morgan, March 20 1998). “As to ‘information,’ I imagine
121 one could avoid the word if one didn’t like it and say ‘detailed residue-by-
122 residue determination’ ” (Crick to Morgan, April 3 1998). Moreover, “As to
123 ‘meaning’ . . . I would keep away from the term.” (Crick to Morgan, April 3
124 1998) ⁴

125 So if we take Crick at his word, then information is about (1) precise
126 determination and (2) transfer of biological specificity from one biomolecule
127 to another (in both development and in heredity).

128 These two aspects of Crick’s ideas about information can be made precise
129 using Shannon information measures and algorithmic information measures
130 respectively. This paper concentrates on the first aspect of information and
131 on Shannon measures of information.⁵

132 **3. Information as precise determination**

133 When Crick said that he would emphasise information in his account of
134 protein synthesis, rather than matter and energy, he meant that he would
135 focus on the precise determination of the structure of one biomolecule by
136 another. There are variables through which the cell exercises this precise
137 determination, notably coding sequences of nucleic acid, and other variables
138 through which it does not, such as the presence or absence of an RNA poly-
139 merase in the transcription process. Variables of this second kind are ab-
140 solutely required to construct the downstream biomolecule: without them
141 nothing will happen. But they do not precisely determine the structure of
142 that biomolecule: their role will remain the same no matter what particular
143 structure is produced. Crick’s distinction between ‘matter and energy’ on the
144 one hand and ‘information’ on the other thus corresponds to the standard
145 distinction between the *efficiency* and *specificity* of a molecular process. The
146 efficiency of a molecular process is a matter of how much product is obtained

⁴Philosopher Gregory Morgan received two letters from Crick in response to questions about how and why Crick came to use the concept of information in his work. These were kindly made available to us by Morgan. Crick also states that the inspiration for his use of ‘code’ in the sequence hypothesis was the Morse Code’s purely syntactic mapping between two alphabets (Crick to Morgan, April 3 1998)

⁵A treatment of the second aspect of Crick’s ideas about information using algorithmic information measures is in preparation

147 for a given quantity of inputs. The specificity of the process is the extent to
148 which the process produces just one output, rather than other energetically
149 equivalent outputs. A well-designed polymerase chain reaction, for example,
150 will produce just one DNA product (specificity) but many copies of that
151 product (efficiency).

152 Biological specificity is explained by locating the variables through which
153 cells exercise precise determination of outcomes. In philosophy these vari-
154 ables are known (coincidentally as far as the author can discover) as ‘specific
155 causes’[17][18]. In earlier work the present author and collaborators have
156 developed an information-theoretic approach to measuring the specificity of
157 causal relationships [19][20].

158 This work was a contribution to the so-called ‘interventionist’ approach
159 to causation[21][22], which is based on the insight that “causal relationships
160 are relationships that are potentially exploitable for purposes of manipula-
161 tion and control”[17][p. 314]. Interventionists treat causation as relationships
162 between the variables that characterise an organised system. These rela-
163 tionships can be represented by a directed acyclic graph. In such a graph,
164 variable C is a cause of variable E when a suitably isolated manipulation
165 of C would change the value of E . With suitable restrictions on the idea
166 of ‘manipulation’ this test provides a criterion of causation, distinguishing
167 causal relationships between variables from merely correlational relationships
168 [21][pp. 94-107].

169 Using this definition most events have many, many causes. But only some
170 of these causal relationships are highly specific. The presence of oxygen in
171 the atmosphere was one cause of the bushfire, but the arsonist was a more
172 specific cause. The intuitive idea of specificity is that interventions on C
173 can be used to produce any one of a large number of values of E , so that
174 the cause variable has what Woodward terms “fine-grained influence” over
175 the effect variable [17][p. 302]. This idea can be quantified using Shannon
176 information theory with the addition of an intervention operator that allows
177 us to isolate the causal component of the correlation between variables:

178 SPEC: the specificity of a causal variable is obtained by measur-
179 ing how much mutual information interventions on that variable
180 carry about the effect variable.⁶

⁶[19][20]. This measure has been independently proposed in neuroscience [23]and in
the computational sciences [24]. For other related measures see [25][26].

181 Formally, the specificity (I) of C for E against a background of other
 182 variables B is:

$$I(\widehat{C}; E|\widehat{B}) = \sum_b p(\widehat{b}) \sum_c p(\widehat{c}|\widehat{b}) \sum_e p(e|\widehat{c}, \widehat{b}) \log_2 \frac{p(e|\widehat{c}, \widehat{b})}{p(e|\widehat{b})} \quad (1)$$

183 Equation 1 is a variant on the equation for Shannon’s mutual information,
 184 which measures the overlap, or redundancy, in the probability distributions of
 185 two variables. The $\widehat{}$ (‘hat’) on a variable denotes Judea Pearl’s intervention
 186 operator [22] and indicates that the value of that variable is determined
 187 by intervention rather than observation. These interventions transform the
 188 symmetrical mutual information measure into an asymmetric measure of
 189 causal influence, since it now represents not the observed correlation between
 190 the variables, but the effect on E of experimentally intervening on C whilst
 191 controlling for background variables B . If two variables are not causally
 192 connected, then however strongly they are correlated, $I(\widehat{C}; E|\widehat{B}) = 0$.

193 A more intuitive way to think about the specificity measure is that it
 194 measures the extent to which an agent can reduce their uncertainty about
 195 the value of the effect variable if they can change the value of the cause, that
 196 is, the extent to which the agent can *precisely determine* the value of E by
 197 intervening on C .

198 SPEC can be used to measure either how specifically two variables are
 199 connected (potential causal influence) or how much of the actual variation
 200 in E in some data is causally explained by variation in C (actual causal
 201 influence) [19][20]. Whilst the use of Shannon information theory means
 202 that the measure is restricted to discrete variables, equivalent measures of
 203 metric variables are possible. None of these additional complexities need
 204 concern us in the present discussion, however. Instead, we will briefly see how
 205 SPEC can be used to elucidate the difference between sources of specificity,
 206 such as coding sequences of DNA, on the one hand and sources of efficiency,
 207 such as RNA polymerase, on the other. We will then turn our attention to
 208 generalising this approach to sequence specificity.

209 4. Genetic and epigenetic information

210 If biological information is precise determination, as measured by SPEC,
 211 then it is easy to see that DNA is a rich source of information in the produc-
 212 tion of biomolecules in a way that distinguishes it from many other causes

213 of those biomolecules. Varying the sequence of DNA exerts fine-grained con-
214 trol over the structure of the molecules produced. Griffiths and collaborators
215 [19][pp. 539-40] constructed a toy causal model of transcription with three
216 variables: RNA Polymerase (POL), which is either Present or Absent, DNA,
217 whose values are alternative DNA sequences, and RNA, whose values are
218 alternative RNA sequences. The value of RNA depends on both POL and
219 DNA. Nothing is transcribed if POL = absent and when POL = present,
220 each value of DNA determines a unique value of RNA. This is roughly how
221 Crick imagined transcription, although, of course, the chemical nature of the
222 transcription machinery was unknown. Assuming for simplicity a maximum
223 entropy distribution over both POL and DNA, the specificity of POL for
224 RNA can never exceed 1 bit, since POL has an entropy of 1 bit and the mu-
225 tual information between two variables cannot exceed the lowest maximum
226 entropy of either variable. However, once the number of possible values of
227 DNA each determining a unique RNA product exceeds 4, then DNA will
228 always have > 1 bit of specificity for RNA. ⁷

229 Calculations on a toy model are of limited interest. However, the approach
230 that lies behind them has some immediate exciting consequences. The first is
231 that this measure can be applied to *both* coding and non-coding regions in the
232 genome to allow a quantitative comparison of the contribution of variables of
233 both kinds to the precise determination of the sequence of a biomolecule. For
234 example, mutations to any of the many well-characterised intronic splicing
235 enhancer (ISE) or silencer (ISS) regions change the probability that one or
236 more exons will be removed from the resulting transcript [27]. We could
237 introduce this process into our toy model by replacing the variable DNA
238 with two variables, INT and EXO, whose values would be the intronic and
239 exonic content of the original DNA sequences respectively. The existence of
240 intronic splicing control regions would be represented by the specificity of
241 INT for RNA. This is an absolutely natural extension of the moves Crick
242 himself made in his 1958 paper in the light of what we now know about how
243 biomolecules are synthesized from the genome. There is sequence specificity
244 in non-coding regions.

245 Our approach has vindicated the idea that biological information is not
246 restricted to the coding regions of the genome, but can be found in other

⁷The entropy of RNA is $H(RNA) > 2$, we have just seen that $I(\widehat{POL}; RNA) = 1$, and DNA accounts for all the remaining entropy: $I(\widehat{DNA}; RNA) = H(RNA|\widehat{POL}) > 1$

247 functional regions as well. But we can go further. Our measure can be ex-
248 tended to variables representing epigenetic (narrow sense, see Box 1.) modi-
249 fications of DNA, insofar as they make a difference to the precise sequence of
250 biomolecules through their role in the regulation of transcription and post-
251 transcriptional and post-translational processing.

Box 1. Definitions of epigenetic. From [8] [p. 112]

Epigenesis: the idea that the outcomes of development are created in the process of development, not preformed in the inputs to development; epigenetic can be used in these senses:

Epigenetics (broad sense Waddington 1940): the study of the causal mechanisms by which genotypes give rise to phenotypes; the integration of the effects of individual genes in development to produce the epigenotype.

252 **Epigenetics** (narrow sense Nanney 1958): the study of the mechanisms that determine which genome sequences will be expressed in the cell; the control of cell differentiation and of mitotically and sometimes meiotically heritable cell identity.

Epigenetic inheritance (narrow sense): the inheritance of genome expression patterns across generations (e.g. through meiosis) in the absence of a continuing stimulus.

253 **Epigenetic inheritance** (broad sense): the inheritance of phenotypic features via causal pathways other than the inheritance of nuclear DNA. We refer to this as exogenetic inheritance (West and King 1987).

254 Numerous mechanisms have been suggested by which epigenetic marks
255 could determine which exons will be included in a mature mRNA. RNA splic-
256 ing is frequently co-transcriptional, either by splicing actually occurring while
257 the pre-mRNA is still being transcribed or by the recruitment of factors that
258 determine later splicing whilst the pre-mRNA is being transcribed. This cre-
259 ates many opportunities for interaction between the splicing machinery and
260 chromatin. The strongest direct evidence to date of epigenetic determination
261 of alternative splicing is by alternative methylation states of histones. Indi-
262 rect evidence suggests multiple significant roles for chromatin in determining
263 alternative splicing [28][29][30].

264 Epigenetic regulation of splicing is another missing variable in the toy
265 model described above. If we extended the model to include it, variable(s)
266 representing the methylation and acetylation state of histones would have

267 some specificity for the RNA product variable. So, by a direct application of
268 Crick’s original reasoning, there is both genetic and epigenetic information in
269 Crick’s original sense: both genes and epigenes can have sequence specificity.

270 Epigenetic modifications of chromatin can have sequence specificity. This
271 will seem unsurprising to many biologists, given the number of papers that
272 described the discovery of such mechanisms as the discovery of ‘missing in-
273 formation’ for splicing [27][30]. This way of speaking need not be regarded
274 in the deflationary manner described in Section 1. The approach to infor-
275 mation outlined here shows that it can be taken literally as a step towards a
276 unified theory of biological information. Sequence specificity is a measurable
277 quantity that plays a causal role in the production of biomolecules, namely
278 the precise determination of their linear structure.

279 **5. Why epigenetic information cannot be reduced to genetic infor-** 280 **mation**

281 A common thought about why epigenetics cannot be a distinct source
282 of information is worth considering, because it throws light on why Crick
283 needed to introduce the idea of information. The thought is that, because
284 the machinery that creates epigenetic modifications consists of molecules
285 transcribed from the genome, the information in the epigenetic marks must
286 ultimately be derived from the genome.

287 “The problem with this kind of hair splitting is that ultimately
288 the extra information (e.g. methylation) is provided by enzymes
289 (methylases) encoded by genes in the genome. Epigenetics, per
290 se, doesn’t add any new information. It’s just a consequence, or
291 outcome, of the information already in the DNA.”⁸

292 This informal comment is significant precisely because it is a typical first
293 response to the idea that epigenetic marks contain information that supple-
294 ments the information in the genome. This response makes it clearer why
295 Crick needed to distinguish “the flow of energy, the flow of matter, and the
296 flow of information.” (1958, 144) The concept of specificity is a causal con-
297 cept, not a material one, and identifying the sources of biological specificity

⁸Larry Moran, Sandwalk Blog: <http://sandwalk.blogspot.com.au/2016/10/extending-evolutionary-theory-paul-e.html> Accessed 2016-12-08. This was a response to the abstract of the conference presentation from which this article is derived.

298 requires measuring causal control, not material contributions. Once we look
299 at the matter in this light it becomes clear that some epigenetic modifications
300 are specified by genomes whilst others are not.

301 To see why the ‘matter and energy’ side of how epigenetic marks are
302 created is not relevant, consider a case in which epigenetic marks are a site
303 of conflict between multiple genomes. In cases of parental imprinting of
304 genes it is biological common-sense that the parent, not merely the offspring,
305 is a source of the biological information expressed in offspring phenotype. If
306 this genetic conflict is mediated by epigenetic mechanisms that contribute
307 to the precise determination of the sequence of gene products, for example
308 by affecting which exons are included in a transcript [31], then it makes no
309 sense to say that the information specifying the splice variant all comes from
310 the *offspring* genome. The fact that the coding sequences for the enzymes
311 involved in establishing and maintaining the methylation pattern are in the
312 offspring genome is irrelevant. The relevant issue is where causal control
313 is being exercised over the transcription and processing of those sequences.
314 When parental imprints are established, the offspring provides the efficiency
315 of the reaction, but the parent provides at least part of the specificity of the
316 reaction.

317 Now consider a case where the epigenetic mechanism that contributes to
318 the precise determination of phenotype is influenced by the offspring’s en-
319 vironment. For example, regulation of alternative splicing by temperature
320 seems to be an important mechanism for maintaining circadian rhythms in
321 a wide range of species [32][33]. It seems reasonable to describe this as a
322 mechanism for conveying environmental information to the genome, so that
323 genome expression can be correctly matched to the environment. After all,
324 the adaptive problem facing the organism is to reduce its uncertainty about
325 where it is in the diurnal cycle and it does this by responding to an environ-
326 mental cue. Our account of information vindicates this idea - we could, at
327 least in principle, measure the contribution of the environmental variable to
328 the precise determination of sequence, just as we did the contribution of the
329 epigenetic marks further along in the causal graph. The fact that the coding
330 sequences for the enzymes involved are in the genome is irrelevant. The real
331 issue is where causal control is being exercised over the transcription and pro-
332 cessing of those sequences. In this case, evolution has designed a mechanism
333 which detects and responds to information from the environment.

334 In this section we have seen that our measure can be used to identify
335 sequence specificity in both coding and non-coding sequences, in epigenetic

336 marks, and in the causes of those marks, whether that is other genomes in
337 cases of genetic conflict, or the environment in cases of plasticity. Information
338 in Crick's sense is about precise determination. We have expanded the class
339 of things that do the determining beyond those Crick originally envisaged.
340 In the following section we will also expand the class of things that get
341 determined.

342 6. Sequence specificity and other biological information

343 Crick used 'information' to label the distinctive relationship of precise
344 determination that holds between coding sequences of nucleic acids and the
345 order of elements in their products, a relationship which does not hold be-
346 tween those products and many of their other causes. However, in Sections
347 4 and 5 we saw that *some* other causes *do* have this relationship to the order
348 of elements in gene products. In this section we ask whether this distinc-
349 tive relationship of precise determination exists for phenotypes more distal
350 than the primary structure of RNAs or proteins. In this context we will not
351 talk of 'sequence specificity', reserving that term for the precise determina-
352 tion of sequence, which was Crick's original concern. We will use the more
353 general term 'biological information' to refer to the precise determination of
354 phenotypes that are causally downstream of the primary structure of gene
355 products, phenotypes such as the tertiary structure of proteins, and still more
356 distally, morphology, and behavior.

357 As we noted in Section 1, the expressive power of the genetic code is
358 limited to specifying the linear order of elements in a polypeptide. Changes
359 to DNA coding sequences *cause* a whole chain of events, but they do not *code*
360 for the more distal events in that chain [2]. The use of 'code' in this extended
361 sense is metaphorical, like saying that when Richard Nixon literally ordered
362 the Watergate cover-up he also 'ordered' his own downfall.

363 But while the genetic triplet code is limited in this way, the broader idea
364 of information as precise determination is not. The idea of information as
365 precise determination, whether measured using SPEC or another measure,
366 can be applied to any set of variables arranged in a causal graph. In principle,
367 therefore, our approach can be used to measure biological information in a
368 gene (or an epigene) with respect to any downstream variable affected by that
369 gene. In fact, a range of causal Shannon information measures related to the
370 one introduced here are already used in complex systems science to study a
371 wide spectrum of living and non-living systems [34]. Genes or epigenes may

372 not literally ‘code’ for morphology and behavior, but they do literally contain
373 biological information that specifies to some measurable degree morphology
374 and behavior.

375 It is now possible to extend our approach to biological information to
376 mechanisms of exogenetic heredity (broad-sense epigenetic inheritance, see
377 Box 1). We have already seen that environmental factors can have sequence
378 specificity, since they can be specific causes of epigenetic modifications of
379 chromatin and thus contribute to the precise determination of the structure
380 of biomolecules. But there are broader mechanisms of environmental hered-
381 ity, such as habitat or host imprinting, in which the phenotype of offspring
382 is influenced by parental phenotype but where no epigenetic mark is trans-
383 mitted through meiosis, so there is no epigenetic inheritance in the standard,
384 narrow sense. These broader mechanisms are still usually referred to as ‘epi-
385 genetic inheritance’ but we will refer to them as exogenetic inheritance to
386 avoid confusion. The question of whether such environmental variables con-
387 tribute information to development becomes the considerably more precise
388 question of how specific is the causal relationship between those variables
389 and variables representing morphology and behavior.

390 At this point we have something like a general theory of biological infor-
391 mation. Information refers to a distinctive relationship of precise determi-
392 nation, which we can identify with the older concept of biological specificity.
393 The phenomenon of biological specificity is explained by the existence of
394 causes through which organisms exercise precise determination of outcomes,
395 and the functional expression of this specificity is explained by natural se-
396 lection acting on those causes. Central to organisms’ ability to exercise this
397 highly specific control is the relationship of precise determination originally
398 identified by Crick between the sequence of DNA and the sequences of RNA
399 and protein. Heredity is the transfer of biological specificity from one gener-
400 ation to the next. Central to organisms’ ability to transfer specificity in this
401 way is the existence of coding sequences of DNA which contain the informa-
402 tion to determine the specificity of their products.⁹

⁹Comparison of causal roles need not be reduced to a simple ‘more or less specific’. For example, elucidating the distinction between permissive and instructive induction events in development requires a more complex application of the tools used here [35]

403 7. Development and evolution

404 We have seen that there can be genetic, epigenetic and exogenetic sources
405 of biological information in development. How significant the later two
406 sources are in development is an empirical question. But even biologists
407 who find it plausible that epigenetic and exogenetic factors are significant in
408 development are often sceptical about whether they are significant in evolu-
409 tion. The most common reason for this scepticism is that epigenetic marks
410 are relatively unstable when compared to genetic mutations.

411 The key point is that if epigenetic states are important to evolu-
412 tion, they are important through stable changes in these states,
413 namely transmissible epimutations. And if epimutations are not
414 transmitted with reasonable stability over generations, they can-
415 not have any long-term evolutionary potential (Slatkin 2009). If
416 an epimutation is to have evolutionary importance, it must per-
417 sist. [36] [p. 391]

418 The stability of epigenetic marks is certainly an important question. But
419 whether their evolutionary significance turns on their stability depends on
420 what is meant by ‘evolutionary significance’. In at least one important sense
421 of that phrase, epigenetic marks do not need to be stable to be significant. It
422 is surely reasonable to regard a biological phenomenon as having evolutionary
423 significance if it has widespread and substantial impact on the dynamics
424 of evolution, or, to put it another way, if models that do not include this
425 phenomena are unlikely to correctly predict the course of evolution. But we
426 already know that this is the case from work on the evolutionary genetics of
427 maternal effects [37]. Maternal effects can be defined as the causal influence
428 of maternal genotype or phenotype on offspring phenotype independent of
429 offspring genotype [38], which is in line with the approach taken here to
430 defining epigenetic and exogenetic information. Maternal effects may be
431 either epigenetic or exogenetic, depending on the specific causal pathway by
432 which maternal influence is exerted.

433 Maternal effects, and parental effects generally, are recognised as a sig-
434 nificant factor in evolution [39]. But any form of epigenetic or exogenetic
435 heredity that is a significant source of biological information in the sense
436 defined above will be significant in the same way because it substantially
437 alters the mapping from parent phenotype to offspring phenotype. In this
438 sense, epigenetic and exogenetic heredity is significant for evolution for the

439 same reason that Mendelian models of heredity were significant. The pri-
440 mary significance of Mendelism for the theory of natural selection was that
441 it specified the form of the transmission phase. Epigenetic and exogenetic
442 heredity change this form, and even in the most conventional cases, where
443 maternal effects are simply a one-generation time-lag in the expression of an
444 allele, this has substantial impact on the dynamics of natural selection.

445 Since Wilkins is well aware of all these points we can infer that this is *not*
446 the sense in which he is asking ‘if epigenetic states are important to evolu-
447 tion.’ Another valid sense of that question is whether epigenetic or exogenetic
448 mutations can be the basis of cumulative adaptation. It is plausible that an
449 unstable inheritance system cannot play this role, but that does not mean
450 that it cannot play an important role in a process of cumulative adaptation
451 that also involves the genetic heredity system [40]. Finally, an important per-
452 spective on the relative evolutionary significance of genetic, epigenetic and
453 exogenetic heredity is that they may play complementary roles. For example,
454 it is plausible that genetic and epigenetic heredity allows organisms to adapt
455 themselves to changing environments on different timescales [41].

456 Other authors have argued that to suppose epigenetic inheritance implies
457 anything for evolutionary theory is to conflate ‘proximate’ or mechanistic
458 with ‘ultimate’ or evolutionary biology. Scott-Phillips et al [42] draw a useful
459 comparison between the discovery of epigenetic inheritance and the discovery
460 of Mendelian genetics. In the first years of the 20th century some Mendelians
461 saw Mendelian inheritance as a theory of evolutionary change and presented
462 it as a challenge to the Darwinian theory of natural selection. They suggest
463 that authors who present epigenetic inheritance as a challenge to conventional
464 neo-Darwinism are like those early Mendelians: they are confusing a proxi-
465 mate, mechanistic theory of heredity with an ultimate theory of the causes
466 of evolutionary change. Scott-Phillips et al are engaged in a wider dispute
467 with authors who question the value of the proximate/ultimate distinction
468 [43] and I will not address that wider dispute here. However, with respect
469 to the specific issue of whether epigenetic inheritance has implications for
470 evolutionary theory, their analogy seems to establish exactly the opposite of
471 their intended conclusion. The founders of modern neo-Darwinism did not
472 dismiss Mendelism as a merely proximal mechanism, they used it to derive
473 the form of the transmission phase in the process of natural selection. As I
474 pointed out above, epigenetic and exogenetic heredity shows up in quantita-
475 tive genetics as parental effects, and the incorporation of parental effects into
476 evolutionary models has a significant effect on evolutionary dynamics. In this

477 way both Mendelian heredity and epigenetic heredity are part of ultimate,
478 not merely proximate biology.

479 An interesting aspect of Scott-Phillips et al’s argument is their insistence
480 that, “Put simply, if we wish to offer an ultimate explanation for the exist-
481 ence of some trait, we must make reference to how that trait contributes
482 to inclusive fitness.” [42] [p 40]. They base this conclusion on the results of
483 Grafen’s ‘formal Darwinism’ project [44] which seeks to show that evolution-
484 ary dynamics are in important respects equivalent to the maximisation of
485 inclusive fitness. But what is done in this very impressive program of work
486 is to rigorously compare optimisation models to population genetic models,
487 where the latter models simply assume that there is no epigenetic heredi-
488 ty. This is not a problem for the formal Darwinism program.¹⁰ But it is a
489 problem for Scott-Phillips et al, who are effectively arguing that epigenetic
490 inheritance cannot contribute to ultimate explanation because maximising
491 (genetic) inclusive fitness fully represents evolutionary dynamics in models
492 which assume there is no epigenetic inheritance.

493 Dickins and Rahman [46] suggest that, while epigenetic inheritance may
494 play a role in evolution, those who present it as a challenge to conventional
495 neo-Darwinism have only presented evidence that it is a significant proxi-
496 mate mechanism. They have failed to present evidence that it is significant
497 in ultimate biology. Once again, this seems to overlook the way that epige-
498 netic and exogenetic heredity show up in conventional, quantitative genetic
499 models, namely as parental effects, and the known impact of such effects on
500 evolutionary dynamics.

501 8. Conclusion

502 We set out to define a sense of ‘information’ that can make sense of the
503 idea that development is the expression of information that accumulated
504 during evolution and that heredity is the transmission of this information.
505 Whilst compelling at a metaphorical level, this is surprisingly hard to cash
506 out in serious, scientific terms. We began with a simple conception of infor-
507 mation that plays a straightforward role at the heart of molecular biology and
508 explored how many other aspects of biology can be clarified using this sense

¹⁰Lu and Bourrat [45] have recently discussed how this program can be extended to include epigenetic inheritance and suggest that *because of this* epigenetic inheritance does not require any radical revision of conventional neo-Darwinism.

509 of information. Our starting point was the sense of information introduced
510 by Francis Crick in 1958. We identified two aspects of Crick's conception
511 of information (1) precise determination and (2) the transfer of biological
512 specificity from one molecule to another. This paper concentrated on the
513 first aspect. We analysed the idea of precise determination using an informa-
514 tion theoretic measure of causal specificity. Using this measure we showed
515 that coding sequences of DNA have a distinctive relationship of precise de-
516 termination to RNAs and polypeptides. This distinguishes coding sequences
517 from many other causes of the same outcomes, such as the presence of an
518 RNA polymerase. This is what Crick meant when he identified coding se-
519 quences as containing information and the other causes as not doing so. His
520 distinction is closely related to the distinction between the specificity and
521 efficiency of a biochemical process.

522 Since 1958, however, a great deal has been learnt about the production
523 of biomolecules. We saw that Crick's approach to information has natural
524 extensions to non-coding regions of DNA, to epigenetic marks, and to the
525 genetic or environmental upstream causes of those epigenetic marks. Any
526 of these variables may have sequence specificity, that is, they may con-
527 tribute substantially to the precise determination of the linear structure
528 of biomolecules. Moreover, we saw that it is a mistake to suppose that
529 the sequence specificity of epigenetic marks must always derive from se-
530 quence specificity elsewhere in the genome, or in other genomes. Finally,
531 we generalised to a broader concept of 'biological information' that is ap-
532 plicable to more distal phenotypes, and not merely to the linear structure
533 of biomolecules. Relationships of precise determination can exist between
534 genetic, epigenetic and exogenetic factors in development and distal pheno-
535 types, such as morphology and behavior. This gives us a general theory of
536 biological information that can be used to restate more precisely the idea with
537 which we started. Development is the expression of biological specificity, or
538 biological information conceived as precise determination and measured using
539 causal information theory. In heredity, factors which are able to exercise this
540 precise determination are passed on from previous generations. These factors
541 may be genetic, epigenetic or exogenetic. In the penultimate section of the
542 article we argued that the existence of biological information in epigenetic
543 and exogenetic factors is relevant to evolution as well as to development.

544 **Competing Interests**

545 I have no competing interests.

546 **Funding**

547 This publication was made possible through the support of a grant from
548 the Templeton World Charity Foundation. The opinions expressed in this
549 publication are those of the author and do not necessarily reflect the views
550 of the Templeton World Charity Foundation.

551 **Acknowledgements**

552 I thank the members of the Theory and Method in Biosciences group
553 at the Charles Perkins Centre for feedback on an earlier draft, Arnaud
554 Pocheville for drawing Figure 1, and Stefan Gawronski for his assistance
555 with the preparation of the manuscript.

556 **References**

- 557 [1] Williams GC. Natural Selection: Domains, Levels and Challenges. New
558 York: Oxford University Press; 1992.
- 559 [2] Godfrey-Smith P. On the theoretical role of “genetic coding”. *Philoso-*
560 *phy of Science*. 2000;p. 26–44.
- 561 [3] Jenuwein T, Allis CD. Translating the histone code. *Science* (New York,
562 NY). 2001 Aug;293(5532):1074–1080.
- 563 [4] Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al. Deci-
564 phering the splicing code. *Nature*. 2010 May;465(7294):53–59. Available
565 from: <http://www.nature.com/doi/finder/10.1038/nature09000>.
- 566 [5] Sarkar S. Biological information: A sceptical look at some central dog-
567 mas of molecular biology. In: Sarkar S, R S C, editors. *The Philosophy*
568 *and History of Molecular Biology: New Perspectives*. vol. 183 of Boston
569 *Studies in the Philosophy of Science*. Dordrecht: Kluwer Academic Pub-
570 lishers; 1996. p. 187–232.
- 571 [6] Levy A. Information in Biology: A Fictionalist Account. *Nous*.
572 2011;45(4):640–657.

- 573 [7] Griffiths PE. Genetic Information: A Metaphor in Search of a Theory.
574 *Philosophy of Science*. 2001;68(3):394–412.
- 575 [8] Griffiths PE, Stotz K. *Genetics and Philosophy: An introduction*. New
576 York: Cambridge University Press; 2013.
- 577 [9] Stotz K, Griffiths PE. Biological Information, causality and specificity
578 - an intimate relationship. In: *From Matter to Life: Information and*
579 *Causality*. Cambridge and New York: Cambridge University Press; In
580 Press. p. 000–000.
- 581 [10] Maynard Smith J. The concept of information in biology. *Philosophy*
582 *of Science*. 2000;67(2):177–194.
- 583 [11] Shea N. Representation in the genome and in other inheritance systems.
584 *Biology and Philosophy*. 2007;22:313–331.
- 585 [12] Witzany G, Baluška F. Life’s code script does not code itself. *EMBO*
586 *reports*. 2012 Dec;13(12):1054–1056.
- 587 [13] Rosenberg A. Is epigenetic inheritance a counterexample to the central
588 dogma? *History and philosophy of the life sciences*. 2006;p. 549–565.
- 589 [14] Morange M. *A History of Molecular Biology*. Cambridge, MA: Harvard
590 University Press; 1998.
- 591 [15] Crick FHC. On Protein Synthesis. *Symposium of the Society for Ex-*
592 *perimental Biology*. 1958;12:138–163.
- 593 [16] Crick FHC. Central Dogma of Molecular Biology. *Nature*. 1970
594 Aug;227:561–563.
- 595 [17] Woodward J. Causation in biology: stability, specificity, and the choice
596 of levels of explanation. *Biology & Philosophy*. 2010;25(3):287–318.
- 597 [18] Waters CK. Causes that make a difference. *Journal of Philosophy*.
598 2007;104(11):551–579.
- 599 [19] Griffiths PE, Pocheville A, Calcott B, Stotz K, Kim H, Knight RD.
600 Measuring causal specificity. *Philosophy of Science*. 2015;82:529–555.

- 601 [20] Pocheville A, Griffiths PE, Stotz K. Comparing causes - an information-
602 theoretic approach to specificity, proportionality and stability. In: Leit-
603 geb H, Niiniluoto I, Sober E, Seppälä, Päivi, editors. Proceedings of
604 the 15th Congress of Logic, Methodology and Philosophy of Science.
605 London: College Publications; In Press. p. 000–000.
- 606 [21] Woodward J. Making Things Happen: A Theory of Causal Explanana-
607 tion. New York & Oxford: Oxford University Press; 2003.
- 608 [22] Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. Cam-
609 bridge University Press; 2009.
- 610 [23] Tononi G, Sporns O, Edelman GM. Measures of degeneracy and redun-
611 dancy in biological networks. Proceedings of the National Academy of
612 Sciences. 1999;96(6):3257–3262.
- 613 [24] Korb KB, Hope LR, Nyberg EP. Information-Theoretic Causal Power.
614 In: Emmert-Streib F, Dehmer M, editors. Information Theory and Sta-
615 tistical Learning. Boston, MA: Springer US; 2009. p. 231–265.
- 616 [25] Ay N, Polani D. Information flows in causal networks. Advances in
617 complex systems. 2008;11(01):17–41.
- 618 [26] Janzing D, Balduzzi D, Grosse-Wentrup M, Schölkopf B. Quantifying
619 causal influences. The Annals of Statistics. 2013;41(5):2324–2358.
- 620 [27] Wang Z, Burge CB. Splicing regulation: From a parts list of regulatory
621 elements to an integrated splicing code. RNA. 2008 Mar;14(5):802–813.
- 622 [28] Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. Epigenetics in
623 Alternative Pre-mRNA Splicing. Cell. 2011 Jan;144(1):16–26.
- 624 [29] Sorenson MR, Jha DK, Ucles SA, Flood DM, Strahl BD, Stevens SW,
625 et al. Histone H3K36 methylation regulates pre-mRNA splicing in *Sac-*
626 *charomyces cerevisiae*. RNA Biology. 2016 Apr;13(4):412–426.
- 627 [30] de Almeida SF, Carmo-Fonseca M. Design principles of interconnections
628 between chromatin and pre-mRNA splicing. Trends in Biochemical Sci-
629 ences. 2012 Jun;37(6):248–253.

- 630 [31] Cowley M, Wood AJ, Bohm S, Schulz R, Oakey RJ. Epigenetic control
631 of alternative mRNA processing at the imprinted *Herc3/Nap115* locus.
632 *Nucleic Acids Research*. 2012 Oct;40(18):8917–8926.
- 633 [32] Sanchez SE, Petrillo E, Beckwith EJ, Zhang X, Rugnone ML, Hernando
634 CE, et al. A methyl transferase links the circadian clock to the regulation
635 of alternative splicing. *Nature*. 2010 Nov;468(7320):112–116.
- 636 [33] Syed NH, Kalyna M, Marquez Y, Barta A, Brown JWS. Alterna-
637 tive splicing in plants - coming of age. *Trends in Plant Science*. 2012
638 Oct;17(10):616–623.
- 639 [34] Bossomaier T, Barnett L, Harré M, Lizier JT. *An Introduction to Trans-
640 fer Entropy*. Cham: Springer International Publishing; 2016.
- 641 [35] Calcott B. Causal Specificity and the Instructive-Permissive Distinction.
642 *Biology & Philosophy*. In Press;p. 000–000.
- 643 [36] Wilkins A. Epigenetic inheritance: Where does the field stand today?
644 What do we still need to know? In: Gissis SB, Jablonka E, editors.
645 *Transformations of Lamarckism: From Subtle Fluids to Molecular Biol-
646 ogy*. Cambridge, MA: The MIT Press; 2011. p. 389–393.
- 647 [37] Wolf JB, Wade MJ. Evolutionary genetics of maternal effects. *Evolution*.
648 2016 Apr;70(4):827–839.
- 649 [38] Wolf JB, Wade MJ. What are maternal effects (and what are they not)?
650 *Philosophical Transactions of the Royal Society B: Biological Sciences*.
651 2009 Apr;364(1520):1107–1115.
- 652 [39] Badyaev AV, Uller T. Parental effects in ecology and evolution: Mech-
653 anisms, processes, and implications. *Philosophical Transactions of the
654 Royal Society, Biological Sciences*. 2009;364:1169–1177.
- 655 [40] Badyaev AV, Hill GE, Beck ML, Dervan AA, Duckworth RA, McGraw
656 KJ, et al. Sex-Biased Hatching Order and Adaptive Population Diver-
657 gence in a Passerine Bird. *Science*. 2002;295(5553):316–318.
- 658 [41] Danchin E, Pocheville A. Inheritance Is Where Physiology Meets Evo-
659 lution. *Journal of Physiology*. 2014;592:2307–2317.

- 660 [42] Scott-Phillips TC, Dickins TE, West SA. Evolutionary Theory and the
661 Ultimate-Proximate Distinction in the Human Behavioral Sciences. Per-
662 spectives on Psychological Science. 2011 Jan;6(1):38–47.
- 663 [43] Laland KN, Sterelny K, Odling-Smee J, Hoppitt W, Uller T. Cause
664 and Effect in Biology Revisited: Is Mayr’s Proximate-Ultimate Di-
665 chotomy Still Useful? Science. 2011 Dec;334(6062):1512–1516. Avail-
666 able from: [http://science.sciencemag.org.ezproxy1.library.
667 usyd.edu.au/content/334/6062/1512](http://science.sciencemag.org.ezproxy1.library.usyd.edu.au/content/334/6062/1512).
- 668 [44] Grafen A. Optimization of inclusive fitness. Journal of Theoretical Bi-
669 ology. 2006 Feb;238(3):541–563. Available from: [http://linkinghub.
670 elsevier.com/retrieve/pii/S0022519305002559](http://linkinghub.elsevier.com/retrieve/pii/S0022519305002559).
- 671 [45] Lu Q, Bourrat P. The Evolutionary Gene and the Extended Evolution-
672 ary Synthesis. The British Journal for the Philosophy of Science. In
673 Press;p. TBC.
- 674 [46] Dickins TE, Rahman Q. The extended evolutionary syn-
675 thesis and the role of soft inheritance in evolution. Proc
676 R Soc B. 2012 Aug;279(1740):2913–2921. Available from:
677 [http://rspb.royalsocietypublishing.org.ezproxy1.library.
678 usyd.edu.au/content/279/1740/2913](http://rspb.royalsocietypublishing.org.ezproxy1.library.usyd.edu.au/content/279/1740/2913).