INFORMATION IN DEVELOPMENT AND EVOLUTION

By

RONALD J. PLANER

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Brian P. McLaughlin

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

May 2015

ABSTRACT OF THE DISSERTATION

Information in Development and Evolution

by RONALD J. PLANER

Dissertation Director:

Brian P. McLaughlin

Biologists make regular use of a large family of informational concepts. Examples include “signaling”, “coding”, “programming”, and more. Indeed, many biologists would go so far as to claim that biology is inherently an information science. The application of informational concepts in biology raises some intriguing philosophical questions, particularly when they are applied to low-level entities such as genes and other biomolecules. This dissertation looks at a subset of such applications in low-level biology. More specifically, I consider in detail the notion of a genetic program, of genetic representation, and of positional information (information about a cell’s location in the embryo). In the final part of this dissertation, I switch gears and consider some issues that arise in connection with information in a different biological domain, namely, human evolution.

ACKNOLWEDGMENTS

I would like to extend a warm thank you to all of my dissertation committee members: Frances Egan, Charles R. Gallistel, Peter Godfrey-Smith, and Brian P. McLaughlin. I owe a special debt of gratitude to Brian and Peter. Without Brian's guidance and encouragement throughout my years of undergraduate and graduate study, I would not be where I am today. His advice, concerning matters both philosophical and practical, has proved invaluable. Peter, too, has been a truly remarkable mentor. I cannot thank him enough for the countless hours he has spent reading and discussing my work, despite having his own students' needs to attend to. His influence is apparent throughout this dissertation. Finally, I want to thank my close personal friend, and fellow Rutgers graduate student, E.J. Green, for his constant willingness to listen to my ideas and provide excellent feedback.

Chapter 2 of this dissertation appears in *Biology and Philosophy*: Planer R. (2014) "Replacement of the 'genetic program' program," *Biology and Philosophy,* 29 (1): 33-53.

Chapter 3 is forthcoming in *The British Journal for the Philosophy of Science*: Planer, R. (forthcoming) "Are Genetic Representations Read in Development?" *The British Journal for the Philosophy of Science,* DOI: 10.1093/bjps/axu043.

Chapter 5 is forthcoming in *Biological Theory*: Planer, R. (forthcoming). "On the free-rider identification problem." *Biological Theory*, DOI: 10.1007/s13752-015-0206-2.

DEDICATION

For Ashley and Zoe

TABLE OF CONTENTS

Abstract ii

Acknowledgments..... iii

Dedication iv

1. Introduction 1

2. Replacement of the “genetic program” program 6

2.1 In support of the “genetic program” program 8

2.1 Against the “genetic program” program 15

2.3 Towards a better framework: signaling games 23

2.4 Genetic inheritance 35

3. Are genetic representations read in development? 41

3.1 Reading for Shea 49

3.2 Gene expression as reading 53

3.3 Are genetic representations read in development? (Part 1) 58

3.4 The manipulability theory of causation and difference-making causation 60

3.2 Are genetic representations read in development? (Part 2) 65

3.3 Conclusion 76

4. On the nature and explanatory relevance of positional information 79

4.1 Embryonic geography 81

4.2 Signaling molecules 83

4.3 Information 85

4.3 Positional information 87

4.4 On the explanatory relevance of positional information 88

4.5 Towards an improved account of the explanatory relevance of positional information 92

4.6 Conclusion 107

5. On the free-rider identification problem 109

5.1 Bowles and Gintis' on the the free-rider identification problem 110

5.2 Sterelny's criticism of Bowles and Gintis 118

5.3 Sharing social information 127

LIST OF ILUSTRATIONS

2.1 Genetic instruction table 13

2.2 Payoff matrix for genes 24

4.1 The distribution of signaling molecules in the embryo 85

4.2 Alternating rows of light and dark cells 91

4.3 Forelimb development 98

4.4 Development of a black and yellow ventral patch 100

4.5 A possible pattern of rods and cones in the eye 102

Chapter 1

Introduction

Biologists make extensive use of informational concepts in their day-to-day work. For example, genes are described as "coding for" proteins, and sometimes even for whole-organism traits, cells are described as sending and receiving "signals", developmental processes are described as the result "genetic programs" or "developmental programs", inheritance systems are described as "channels" through which "hereditary information" flows, and so on. Indeed, it is not uncommon to hear it said that biology is essentially an information science. Such a pro-information attitude is vividly illustrated by a passage from Richard Dawkins in his *The Blind Watchmaker* (1996). Dawkins writes:

> There is nothing special about the substances from which living things are made. Living things are collections of molecules like everything else … What is special is that these molecules are put together in much more complicated patterns than the molecules in non-living things, and this putting together is done by *following programs, sets of instructions for how to develop*, which the organisms carry around inside themselves … What lies at the heart of every living thing is not a fire, not a warm breath, not a 'spark of life'. It is *information, words, instructions*. If you want a metaphor, don't think of fires and sparks and breath. Think, instead, of a billion discrete, *digital characters* carved in tablets of crystal. If you want to understand life, don't think about vibrant, throbbing gels and oozes, think about *information technology*." (1996, p. 112) [italics mine]

While Dawkins' view is perhaps extreme, a strong enthusiasm for informational description is certainly evident in mainstream biology. Not all biologists share this enthusiasm, however, and some are vocal critics. Consider the following remarks from Richard Francis, for example:

> Preformationism got a new boost, as genes seemed to provide an attractive preformationist mechanism, eventually seduced through two seductive metaphors—genetic recipes and genetic programs. Because it is purely metaphoric … preformationism is [hard] to put to the test.
>
> … [A]ll [talk] of programming and reprogramming is a diversion from the concrete cellular interactions that are the root of … normal development. (2011, p.136-7)

Here we see Francis taking issue with the programming idea in particular, but his skepticism in fact extends across the board. For Francis, talk of codes, programs, and so on, is nothing more than empty metaphor. At best, the use of these concepts in biology serves to distract us from the mechanistic details of the developmental process.

"Information talk" in biology has attracted a good deal of attention from philosophers of biology over the last few decades, and continues to be a hotly-debated topic. Several important questions have been put on the table. First, and perhaps mostly simply, we can ask: which applications of informational concepts do real theoretical work and which are superficial (or even misleading)? Second, as regards those applications which *are* useful (if indeed there are some), we can ask: in virtue of *what* are they useful? How is it that they serve to increase or deepen our understanding of the relevant biological systems? A third question concerns the *naturalizability* of the informational properties low-level entities such as genes and other molecules are apparently attributed by information talk. It has long been a central aim of the philosophy of mind and language to reduce semantic properties to non-semantic ones. Assuming that the application of informational concepts in biology yields genuine theoretical insights (at least sometimes), does this mean that there is a whole family of semantically evaluable entities with which philosophers have not generally concerned themselves? If so, can the same or similar strategies for naturalizing semantic properties at the level of things like

mental states be called upon to naturalize the semantic properties of these low-level entities, or will some radical new ways of thinking about semantic information be required? Yet another question is whether there might be some substantive applications of informational concepts which biologists have not themselves considered as of yet.

This dissertation engages with all of the above questions, though at times more implicitly than explicitly. My primary focus is on the nature and legitimacy of a certain subset of applications of informational concepts. Chapter 2 deals with the idea of a genetic program, Chapter 3 with the idea of genetic representation, and Chapter 4 with the idea of positional information (information about position in the embryo). In the fifth and final chapter of the dissertation, I switch gears and consider some issues arising in connection with information in a different domain, namely, human evolution. I assume that information talk has a clear and non-controversial application in this domain; what I am instead interested in is how a disposition to share information (in particular, social information) evolved in humans. Below is a summary of the main work done in each chapter. Each serves as a stand-alone unit.

In Chapter 2, I argue that the genetic program idea is on the whole misleading. I offer an improved way of thinking, one which preserves the intuitively attractive features of the program description, while avoiding those features which are problematic. On the account I present, genes are recast as senders and/or receivers in Lewis-Skyrms-style signaling games (Lewis 1969; Skyrms 1996, 2003, 2010). Accordingly, genes turn out to be the producers and consumers of developmental or regulatory information, rather than instructions symbolically encoding such information. I then show how this account of genes and gene action throws new light on the nature of genetic inheritance.

In Chapter 3, I consider the most well-worked out account of genetic information to date according to which genes possess semantic information: Nicholas Shea's infotel account (Shea 2007, 2014). Shea claims that developmental systems receive genetic representations which are read over the course of individual development. I argue that a close look at the process of reading for Shea reveals that, if there were genetic representations of the sort he has in mind, they would not be ones that are actually read in development. I then use this fact to cast doubt on Shea's claim that there are genetic representations (again, in his sense of the term) at all.

In Chapter 4, I examine the concept of positional information with an eye towards accomplishing two things. The first is to simply get clear on what positional information is. Despite the prominence of this concept in developmental biology, exactly what is meant by "information" here, and why intercellular signals are legitimately understood as being associated with it, is often left implicit. I then turn to matters concerning explanatory role. What we see is that the standard way of thinking about when positional information is explanatorily relevant to cell behavior, and when it is not, is entirely lacking in motivation. Thus, the second aim of the chapter is to sketch an improved account of the explanatory relevance of positional information. I show how we can appeal to a key distinction between the information that is *carried* by a signal and the information that is *communicated* by that signal to make progress in this area.

Finally, in Chapter 5, I turn to a debate in human evolution with ties to information. Samuel Bowles and Herbert Gintis (2011) have argued that individual-selection accounts of human cooperation flounder in the face of the free-rider identification problem. Kim Sterelny (2012) has responded to this line of argument for

group selection, arguing that the free-rider identification problem in fact poses no theoretical difficulty for individual-selection accounts. Here I set out to clarify Bowles and Gintis' argument. As I see matters, the real crux of their argument is this: solving the free-rider identification problem, even in modestly-sized social groups, requires that group members are disposed to *share social information* with one another. Having clarified their argument, I turn to Sterelny's criticism, and argue that he underestimates the challenge being raised by Bowles and Gintis. More specifically, I argue that it is unclear whether the expected benefits of having a disposition to share social information would have outweighed the expected costs for an individual belonging to a Pleistocene social group. Importantly, this is not to say that I am persuaded by Bowles and Gintis' argument; on the contrary, what I claim is that more theoretical (and in particular) empirical work is necessary before the issues under discussion can be settled. I formulate some specific questions which I think future research in this area should aim to address.

References

Bowles S, Gintis H (2011) A Cooperative Species: Human Reciprocity and Its Evolution. Princeton University Press, Princeton NJ

Dawkins R (1996) The Blind Watchmaker. W. W. Norton & Company: New York, NY

Francis R (2011) Epigenetics: The Ultimate Mystery of Inheritance. Norton & Company: New York, NY

Lewis D (1969) Convention: A Philosophical Study. Harvard University Press, Cambridge

Skyrms B (1996) Evolution of the Social Contract. Cambridge University Press, Cambridge

Skyrms B (2003) The Stag Hunt and the Evolution of Social Structure. Cambridge University Press, Cambridge

Skyrms B (2010) Signals: Evolution, Learning, and Information. Oxford University Press, Oxford

Shea N (2007) Representation in the genome and in other inheritance systems. Biol Philos 22:313–331

Shea N (2014) Inherited Representations Are Read In Development. Br J Philos Sci 64:1–31

Sterelny K (2013) The Evolved Apprentice. MIT Press, Cambridge

Chapter 2

Replacement of the ‘‘genetic program’’ program

Appears in *Biology and Philosophy*, 29 (1): 33-53

Open any textbook on biology today and it won't be long before you encounter the claim that genes carry or encode information. Moreover, it isn't as if talk of genes carrying information is supposed to be a useful fiction or simplification meant solely to introduce newcomers to otherwise complicated facts about genes. Rather, even professional biologists are inclined to treat this claim as literally true. But closer inspection of this claim in recent years, particularly by philosophers of biology, has raised a number of complex issues. The extension of the concept of information, and other related concepts (e.g., representation), to genes remains controversial.

A related issue that has received less critical attention concerns the biological reality of ‘‘genetic programs’’. Among those biologists and philosophers who embrace the idea that genes carry information, many would also want to say that (at least some) genes not only carry information about which proteins to make, but also about the conditions in which to make them. For this reason, the genome is often described as containing a program that is read by the cell during development. In this paper, I argue that the program description, while accurate in some respects, is ultimately misleading and should be abandoned. An alternative framework is sketched that is better suited to capturing the full informational nature of genes. This framework is centered on the notion of a signaling game, as originally developed by David Lewis (1969), but expanded upon considerably by Brian Skyrms (1996, 2003, 2010) in more recent years.

But what is a genetic program anyway? Although frequently mentioned by biologists and researchers in other fields, the concept itself is almost always left vague. In addition, the existence of such a program tends to be conflated with a number of distinct ideas from which it should be kept separate, such as that genes play some kind of privileged role in development, that genes code for whole- organism traits, and so forth. A notable exception in both respects can be found in the work of Gary Marcus (2004). Briefly put, Marcus suggests that since genes both code for proteins and control their own expression conditions, ‘‘each gene acts like a single line in a computer program,’’ realizing an ‘‘IF that controls when [the] gene will be expressed,’’ and a ‘‘THEN that says what protein it will build if it is expressed’’ (p. 60-1). Ironically, rather than vindicating the idea that genomes encode programs, I think Marcus’ insight paves the way for replacement of the genetic program idea with an altogether different and superior one. On the view I will develop, genes turn out to be the producers and consumers of regulatory or developmental information, rather than entities encoding such information. In what follows, I first expand upon Marcus’ basic proposal, filling in some additional details, biological ones in particular. Having done this, I then make the case that the program concept is not the best way to conceive of the functional architecture of the genome underlying control of development. The alternative framework that is provided recasts genes as fundamentally different kinds of entities, so the issue at stake is not merely terminological. Finally, I conclude by considering one important consequence of my view for a broader debate in the philosophy of biology concerning inheritance systems. More specifically, I argue that the nature of genes and gene action that emerges on my view puts pressure on the claim that genetic inheritance should simply be seen as

one among many systems for transmitting information down the generations. I take this to be one form of theoretical payoff that results from applying the signaling games framework to genes.

2.1 In support of the ‘‘genetic program’’ program

A crucial discovery to come out of modern molecular biology is that virtually all genes have two functional parts, one that codes for a protein, and another that controls the expression of the gene.[1] The former is sometimes called the gene's 'protein-template', the latter its 'promoter'. The protein-template just is a long sequence of nucleotide bases that specifies the amino acid sequence of a protein. The function of the promoter is more complicated. Although the promoter is also composed out of nucleotide bases, this region is not transcribed into RNA, and so does not code for a protein. Instead, it serves as a binding site for certain proteins (called 'transcription factors') that are attracted to the promoter due to its DNA sequence properties. The state of the promoter interacts either directly or indirectly with RNA polymerase molecules inside the cell, determining the rate at which the gene's protein-template is transcribed into RNA. In general, transcription can occur only at a very low rate, or else not at all, unless specific combinations of proteins are present (or absent) at the binding sites located in the promoter.

[1] To be sure, many different proposals have been offered through the years regarding what genes are. In this paper, I make the by no means uncommon assumption that a gene includes not only a protein- template but also the proximal DNA region adjacent to the protein-template, the protein-template's promoter region. On this view, genes are not merely coding stretches of DNA, but rather have ‘‘parts’’. It is quite standard I think to now use the term 'gene' to cover at least the protein-template and its promoter region. Here, for example, is the definition of 'gene' cited by Pearson (2006) as having emerged from a recently held genetics symposium: ‘‘A locatable region of genomic sequence... which is associated with regulatory regions, transcribed regions and/or other functional sequence regions’’ (p. 401).

Recent research has revealed that promoters can have an extremely complex internal structure. For example, the promoter of the Endo16 gene of Strongylocentrotus purpuratus (the purple sea urchin) contains six individual cis-regulatory modules, each of which contains multiple protein-specific binding sites for transcription factors (Yuh et al. 2001). There appears to be nothing special about Endo16, so it is believed that this kind of complexity is characteristic of most of the promoters in higher organisms. Strikingly, many of the regulatory modules found in gene promoters resemble the logic gates of modern computing devices: there are modules that positively regulate, that is, increase the rate of transcription, if and only if each of its binding sites is occupied (AND-gate), other modules that positively regulate if and only if one or more of its binding sites is occupied (OR- gate), and still other modules that can positively regulate in the absence of some protein, but which switch to an inhibitory effect once that protein is bound (NOT-gate) (Istrail et al. 2007). It is largely through the functioning of such modules that signal or stimuli-specific gene expression is achieved in cells.

In addition, these modules often interact with each other in complex ways. For example, a protein bound to one module may ‘‘cancel out’’ the effect of a protein bound to another module, in a process called ‘activation site masking’. Alterna- tively, proteins bound at several or more sites may form a ‘‘protein bridge’’ along which other factors literally slide to directly interact with RNA polymerase at the transcription start site. Obviously, the molecular details of these processes can be very complicated, and discussing them would take us well beyond the scope of this paper. Suffice it to say, then, that the internal structure of promoters can be extremely sophisticated, containing many interconnected parts, with each part generally recognizing different sets of proteins.

In sum, the vast majority of an organism's genes are associated with specific, unambiguous conditions under which they will be actively transcribed, resulting in the protein products coded for by the genes (whereby 'actively transcribed' I simply mean transcribed at a rate much higher than when this condition is not satisfied). It would be an idealization to characterize gene expression in terms of Boolean functions, given that transcription is a noisy, probabilistic affair, but I won't worry too much about that here. Each gene can be seen as realizing a certain IF–THEN instruction, just as was suggested by Marcus in the quote above, but strictly speaking these instructions are probabilistic in nature, not all or none. (That is, a gene's IF condition being satisfied will significantly raise the probability that its protein-template will be transcribed in the immediate future, without necessarily guaranteeing it.) Having made this caveat, I will treat these instructions as non- probabilistic in what follows so as to simply the exposition. It's important to note that the dispositions expressed by such IF–THEN instructions are not emergent properties of the cell as a whole in the way that memory or mental representation is sometimes said to be an emergent property of whole brains; rather, the IF half of each instruction is realized by concrete biochemical machinery in the promoter, operating on discrete inputs (proteins), while the THEN half of the instruction is realized by the spatial relation the protein-template bears to a particular promoter. I will return to the significance of this point in due course.

Another relevant feature of genes is that many of them encode transcription factors for other genes. In this way, certain genes can control the activation level of other genes, either up- or down-regulating them. But the genes thus regulated may also happen to encode transcription factors for still other genes giving rise to truly staggering results.

To illustrate this with a toy example, consider a gene G with 4 binding sites for 4 distinct proteins all of which must be occupied for G to be expressed (G thus implements an AND-function). Now suppose that each transcription factor for G is encoded by a gene which itself contains 4 binding sites for 4 other proteins, and that each of these genes also implements an AND- function. In a hierarchy two levels deep, G's expression would be made to depend on the presence of $4^2 = 16$ distinct proteins. Continuing this pattern: in a hierarchy just 4 levels deep, G's expression would be made to depend on the presence of $4^4 = 256$ distinct proteins, while in one 8 levels deep G's expression would be made to depend on the presence of $4^8 = 65{,}536$ distinct proteins! All this despite the fact that G's promoter has only 4 binding sites.

Going in the other direction, we can also imagine such a hierarchy being turned on its head. For example, suppose G encodes a transcription factor for 4 other genes, each of which encodes a transcription for 4 other genes, and so on. Now gene G sits ''on top'' of a gene hierarchy. Assuming that transcription of each of these genes only required a single transcription factor to be present at its promoter, then in a hierarchy 8 levels deep, expression of G would result in the expression of $4^8 = 65{,}536$ distinct genes!

There is again an element of idealization present in these examples. Among other things, gene cascades are much noisier phenomena, with low-level expression of certain genes occurring regularly. Perhaps more importantly, the regulation of transcription factors commonly occurs at one or more post-transcriptional stages (e.g., RNA splicing, differential transport of mRNAs out of the nucleus, modification of pre-existing proteins in the cell, etc.) rather than at the level of transcription itself. Effectively, this means that the idea of a single gene either directly ''turning on'' or ''being turned on by'' a number

of other genes is often misleading or even false. That having been said, the examples certainly do illustrate something important about the nature of gene expression and regulation. What we see is that the system of promoters and transcription factors in the genome allow for individual instructions to be linked together so as to form more complex ones. In practice, there will be a definite limit to the number of instructions that might usefully be composed in this way, though in principle the number is unbounded. One thing this kind of structure supports is the creation of what might be called 'virtual instructions'. Of particular interest, a system of interacting genes can together realize an instruction which takes a larger number of inputs or ''arguments'' than any individual gene can accept. If a gene Ga realizes IF P_1 & P_2 THEN P_3, Gb realizes IF P_4 & P_5 THEN P_6, and finally Gc realizes IF P_3 & P_6 THEN P_7, then the system comprised of all three genes realizes the more complex instruction IF P_1 & P_2 & P_4 & P_5 THEN P_7, even though there is no single gene in the system that is capable of taking more than two inputs.[2] This is none other than a process of functional composition, an ingredient essential to the logic underlying the programs run on modern computing devices. A typical program consists, in large part, in series of sophisticated virtual instructions that have been decomposed into series of less sophisticated (and at the most fundamental level, utterly mindless) ones whose ordered execution realizes the execution of the virtual or high-level instructions.

Thus, we see two features of the genome that serve to give it the general ''flavor'' of a program. Genes can be seen as detailed IF–THEN instructions capable of combining together in various ways so as to form even more complex instructions. Surely, some will disagree that these two features alone are really sufficient for the genome to encode a

[2] Subscripted 'P's stand for proteins. I am here setting aside the complication that some protein-templates can give rise to several or more proteins due to alternative RNA splicing. Nothing hangs on this simplification.

program in anything like a literal sense, and as will become clear shortly, I am sympathetic to that idea. My purpose at present is simply to focus attention on those features of the genome that I think have made the program description appear to be a good one, at least initially.

I propose, then, to understand the structure picked out by the program description as the kind of thing that might be expressed with a long instruction table of the following form:

IF	THEN
Chromosome 1:	
1. $C(P_{i, ..., } P_j)_1$	$(P_1)_1$
2. $C(P_{i, ..., } P_j)_2$	$(P_2)_2$
3. $C(P_{i, ..., } P_j)_3$	$(P_3)_3$
. . .	. . .
n. $C (P_{i, ..., } P_j)_n$	$(P_n)_n$
Chromosome 2:	
1. $C(P_{i, ..., } P_j)_{1'}$	$(P_{n+1})_{1'}$
2. $C(P_{i, ..., } P_j)_{2'}$	$(P_{n+2})_{2'}$
3. $C(P_{i, ..., } P_j)_{3'}$	$(P_{n+3})_{3'}$
. . .	. . .
n'. $C(P_{i, ..., } P_j)_{n'}$	$(P_{n+n'})_{n'}$
. . .	. . .

Figure 2.1

Each row in this table corresponds to a single (coding) gene. The numbers heading the rows indicate the gene's position in the order of genes on a single chromosome. Subscripted '$C(P_i, ..., P_j)$'s are used to stand for the compound expression

realized by the corresponding gene's promoter region. (The (P_i, ..., P_j) portion tells us which combinations of proteins are sufficient for expression of the gene.) Subscripted '(Pi)'s are used to stand for the protein the gene will make if it is expressed. Finally, primes are here used to indicate the chromosome the gene is located on (0 for Chromosome 1, 1 for Chromosome 2, and so on).

The instructions shown above would be the base level ones. By examining the rows, we could see all of the genes that a given Pi acts as transcription for (we just look and see which IF conditions contain P_i). This set might be empty (if P_i is a ''worker protein'') or very large. We could also see which sets of proteins are normally sufficient for the synthesis of P_i (we just look at the C(P_i, ..., Pj) corresponding to the gene coding for P_i and see which combinations of proteins satisfy the expression). Pulling these two strands together, then, we could (in principle!) restate portions of the table using fewer but more sophisticated instructions. The above table would then be seen as a high-level program that had been decomposed into the biological equivalent of machine language. This would be ''no different'' from the way in which a set of machine language instructions stored in computer memory might be the implementation of a familiar word processor program, for example. But just as in a computer, it is the machine language program that is actually read by the containing system (the central processor, in the case of a computer; the cell, in the case of an organism). I find this idea deeply fascinating, but I do not think it remains standing in the face of further scrutiny.[3]

[3] There may be other ways of formulating the genetic program idea, however. One possibility, which I will not consider here, is that the ''division of labor'' we observe in the genome, with some genes encoding transcription factors and others encoding ''worker proteins'', might warrant extending the stored program concept to the genome. The stored program concept is really a name for a particular kind of computational architecture in which both the data symbols and the instructions for manipulating the data symbols are explicitly represented for the machine and stored in the same memory mechanism. This proposal faces its own set of challenges, but my arguments in this paper should be understood as applying only to the

2.2 Against the ‘‘genetic program’’ program

Although the program description may seem like a natural fit in light of the foregoing, I think it is ultimately misleading and should probably be abandoned. In particular, there are at least two crucial respects in which the analogy breaks down. The first is that there is no order in which these instructions can be properly said to be retrieved and executed by the cell during development. Put otherwise, there is no beginning, middle, or end to this ‘‘program’’. Instead, the cell must simply be understood as retrieving and executing every one of these instructions simultaneously or in parallel. This is conspicuously unlike the sorts of programs that computers store in memory and execute. In a computer, the instructions in a program are contained in a memory stack and are read by the central processor in a serial manner. Changing the physical order of the instructions in the memory stack (or jumping to another memory location) would almost certainly alter the end product of the program.[4] There is no analogous process that goes on in the cell.

It is easy to overlook this fact for a number of reasons. For one thing, there is a strong temptation to see an instruction as being executed only when its corresponding protein is being actively synthesized in the cell, but that isn't correct. To see why, note that each instruction also says when not to make some protein. As regards an instruction of the form IF P_i THEN P_j, for example, it isn't the case that this instruction is executed by the cell only when P_j is being made; rather, the instruction is also being executed

---

formulation of the genetic program idea given above. The two are really quite different.

[4] Something like this is true even of multi-threaded programs, whose parts may be executed in parallel by different processors. Although different parts of the program are executed in parallel, it's still the case that the instructions making up those parts are executed in a particular order, and that changing the order of the instructions in those parts would almost certainly alter the outcome of the entire process.

whenever P_j is not being made because $\sim P_i$ is true (i.e., when the antecedent of the instruction is false or does not hold). Therefore, an instruction that results in no protein product over some period of time must be understood as being executed in the same way that an instruction resulting in many copies of its protein being made is executed by the cell. Every gene in the genome will be either involved in protein production at some time t during development or not, and so it follows that every instruction is executed in parallel at t.

Another reason we can be misled is this: the instructions in the genome certainly do have a physical order to them. In fact, they are arranged linearly (at least in eukaryotes) along the chromosomes that contain them, much like bit patterns in a computer memory. However, unlike computer memory, this order does not appear to be causally efficacious in its own right, precisely because each instruction is simultaneously accessed and carried out by the cell. That is why inverting the order of two or more instructions, of two or more genes along a chromosome, does not in general have an effect on development unless that inversion is also associated with a direct change in protein coding information (as in gene fusion events) or the pairing of some protein-template with a new promoter (or other regulatory region). To be sure, we as onlookers can see the instructions as having a certain order to them, an order that is reflected in the table given above, but this order does not generally matter from the perspective of the cell. An interesting and important exception here might be Hox genes, whose linear order along chromosomes mirrors the regions of the body in which they are expressed. That is, the genes at the one end of the Hox cluster are expressed only in tissues located in anterior parts of the organism, those at the other end are expressed only in tissues located

in posterior parts, and those in the middle of the cluster are expressed only in tissues located somewhere in between these regions (central regions). I leave the connection between Hox genes and the exploitation of spatial order for computational purposes for a future work, however.

Just to be clear, I am certainly not denying that the order in which proteins are synthesized (the order in which protein-templates are actively made use of) during development can matter to development; on the contrary, this order can matter a great deal, especially as we move from shorter to longer time scales (e.g., protein P_i was actively synthesized 10 s before protein P_j was actively synthesized, versus protein P_i was actively synthesized 100 days before protein P_j was actively synthesized.) An extremely simple illustration of this point can be seen in the control of metamorphosis in a wide-range of creatures. In *Xenopus laevis* (the African clawed frog), for example, the protein TRα acts as a potent inhibitor of genes involved in metamorphosis (those contributing to the death and reabsorption of larval tissues, the growth of new tissues, etc.) (Laudet 2011). As the production TRα begins to wane, X. laevis begins to undergo metamorphosis. Obviously, if this protein were not synthesized at high levels early on in development, most of the cells in the organism would turn out to be very different. The whole organism would undergo the changes characteristic of its metamorphosis at a much earlier point in time, giving rise to a kind of heterochronic evolutionary change. What I am claiming is that it is not the order in which the IF–THEN instructions are executed (they are all executed simultaneously) that is the determinant of the order in which proteins are synthesized during development. Rather, it is the patterns of cellular conditions over time, interacting with the dispositional properties of genes, that is the

primary determinant of the order of protein synthesis. These two processes are not the same thing.

Now it may not be literally wrong to call a list of instructions with little or no functional order to them a program, but it certainly strikes me as awkward. More to the present point, however, in a typical computer program the end result of the computation is achieved by making each step in the program depend on the result of certain earlier steps. This is precisely what serves to distinguish a genuine program for computing the product of two integers, say, from a mere look-up table that simply specifies in advance the product of every pair of integers (within some range), as might be printed on a child's flash card. Development could very well have been controlled by a program in the more ordinary sense, but this is not the design that evolution arrived at. Instead, the genome's role in development would seem to more closely resemble the flash card model.

The second way in which the program analogy breaks down is related to the first, but raises some distinct issues. Recall that the primary motivation for the program description was to capture the idea that genes not only carry or encode information about which proteins to make, but also about the conditions in which to make them. The main reason that the instructions turn out to lack a functional order is that there is no reader in the cell whose job it is to read each of the instructions one at a time like the central processor in a computer. I now want to argue that the sequence properties of gene promoters are not really read at all inside the cell: there is no physical mechanism that uses the DNA sequences of these regions to determine whether or not the protein coded for by the gene's protein-template should now be made. Admittedly, the term 'read' is an informal one, and a bit vague. I do not propose to offer a definition of the term here. Instead, I will

first sketch what I take to be a clear or paradigm example of a mechanism inside cells that would count as reading the sequence properties of promoters, and then compare it with how we know things to actually be. The important point, I think, is that a bona fide reader of promoters would enable the cell to do things that just aren't possible given its current set-up, so there certainly is a difference.

Consider a thought experiment. Imagine that each cell contained a type of mechanism (of which there might be many token instances) that read the sequence of nucleotide bases making up each gene's promoter. This mechanism would ''recognize'' each such pattern as standing for a certain condition that could obtain within the cell. (The sense in which the mechanism would read this pattern need not be too unlike the way in which ribosomes read the sequence of RNA triplets in mRNAs.) Having formed some kind of short-term ''memory'' of the sequence properties of the promoter region, the mechanism would then check to see if this condition was presently satisfied, say, by taking a sample of the cell's protein population at that time. If the condition was found to obtain, the mechanism would transcribe the protein-template of the gene; otherwise, it would simply move on to another gene. In such a case, I think it would be entirely accurate to say that the IF half of each instruction is explicitly represented and is actually read during development.

Needless to say, this is not how things actually work inside cells. Instead, different types of proteins simply accumulate on the binding sites of a gene's promoter, and as a consequence, the rate at which RNA polymerase molecules are able to stably bind and transcribe the gene's protein-template is modulated. Some of these proteins do directly interact with RNA polymerase molecules, but equally often, the rate of

transcription is determined by interactions among the proteins bound to the promoter itself. In addition, recall the point from above that each instruction must also count as being executed when no protein is being made. Thus, it seems we would have to say that the instruction is read in such cases by virtue of no RNA polymerase interacting with the promoter, which strikes me as highly contrived. In any case, this is quite unlike a computer in which the instruction is explicitly represented for a physical device that decodes it and then acts in accordance with the instruction's content. More generally, if RNA polymerase can be reasonably said to read anything other than the protein- template of the gene, it is the presence or absence of proteins that are now bound to the gene's promoter, not the promoter's DNA sequence properties. Of course, the collection of proteins that are bound to a promoter at some specific point in developmental time is not something that is genetically inherited, but is rather the product of large-scale activities of the entire cell as well as other cells in the organism, and is itself regulated by a huge range of other processes (e.g., signal-dependent transport of proteins into the cell or the cell's nucleus by other molecular machines). Alternatively, one might claim that the sequence properties of promoters are not actually read by RNA polymerase but instead by the proteins that are capable of binding to promoters. This suggestion faces similar problems (e.g., when no protein is being made by the gene, it's the absence of proteins at its promoter that is reading the IF half of the instruction), but we would in addition have to posit a distinct reader for virtually every promoter in the genome. The question that needs asking is whether we have any independent reason to do so other than to simply save the program description.

I do not think this is a mere verbal quibble; there are important differences

between the two kinds of systems being compared here. Note, for example, that in the case described in our thought experiment, it would be appropriate to talk of a ‘‘promoter code’’ on par with the standard genetic code linking DNA and amino acids. And as with the genetic code, the existence of a combinatorial, arbitrary mapping between the representation of the cellular condition and the condition itself would provide the cell with an additional resource to help optimize its performance (in this case, the control of gene expression). For example, with 4 nucleotides to choose from (ACGT), a sequence of just 9 nucleotides would provide the cell with enough symbols to refer to over 200,000 proteins (a number greater than some current estimates of the total number of distinct protein types in a human cell). To encode combinations of proteins, additional 9 nucleotide-long sequences could simply be reserved in the promoter region, together with one additional nucleotide for signaling the logical operation to be performed on the proteins (AND, OR, or NOT) leaving one symbol to spare. This would reduce the number of base pairs necessary to implement one of the IF conditions in our above table by more than an order of magnitude[5], meaning that the energy normally expended upon replication and repair of these DNA regions could be put to some other use.

To be sure, we are here abstracting away from lots of low-level chemical details that would likely render this particular coding schema infeasible or even impossible, but that does not undermine the point illustrated by the example. The existence of a promoter code would add a great deal of flexibility and economy to gene expression, but such a code would imply the existence of a bona fide reader of these DNA sequences (one who ‘‘understands’’ the code). But there is no such mechanism inside cells. What we instead find is a highly complex set of interactions between the DNA sequences of promoters,

[5] The promoter region of the *Endo*16 gene, for example, is about 2,300 base pairs.

proteins, and RNA polymerase that cannot be neatly described as part of an ‘‘encoded instruction’’ that is passed down the generations. To ignore these biochemical details and continue to talk of a program encoded in our genes would, I think, be to engage in the very kind of metaphorical talk that is often criticized by those biologists and philosophers who are skeptical of the notion of genetic information in general.

Here is where we’ve gotten to then. At least within the framework that is standardly used for thinking about genetic information, there is no message or representation in the absence of a receiver of that message. As Peter Godfrey-Smith has put it, this does not mean that ‘‘the physical vehicle of the message has some strange dependence on the existence of a receiver, but for a physical object to count as a message, there must be a receiver present in at least normal circumstances’’ (Godfrey-Smith 2011, p. 178). What we’ve seen is that there does not appear to be a reader of gene promoter regions in cells, or at best the readers are of a highly dubious nature. There is reason then to think that talk of genes carrying information about the conditions in which to make various proteins is misguided. This does not impugn the claim that protein-templates carry information about proteins, as there is indeed a bona fide reader of the nucleotide bases making up the protein-template, namely, RNA polymerase (and a reader of the resulting mRNAs, ribosomes). The picture that emerges is one on which only the THEN halves of the instructions wind up containing information, so that the ‘‘program’’ that is actually encoded by genes and read during development is simply a set of commands of the form ‘make protein p’, ‘make protein q’, and so on. But this is no more a program than a grocery list containing items like ‘get milk’ and ‘get bread’ is a program.

And yet, the complex internal structure of gene promoters should give us pause. Recall

that the typical promoter will contain a number of cis-regulatory modules, many of which bear a striking resemblance to the logic gates in modern computing devices. Moreover, these modules are often interconnected in various ways leading to further functional complexity. There certainly appears to be something ‘‘informational’’ about these DNA regions, it’s just that the program concept does not seem suited to capturing it. Fortunately, there is a better framework for thinking about the full informational nature of genes.

2.3 Towards a better framework: signaling games

Earlier I pointed out that the IF half of each instruction is realized by discrete biochemical machinery (circuits of cis-regulatory modules) operating on discrete inputs (proteins), while the THEN half is realized by the spatial relation the promoter bears to a specific protein-template. Thus, the dispositions expressed by the IF–THEN instructions on our table are not emergent properties of whole cells, but are rather realized locally and concretely by individual genes. I think the patterns implicit in gene expression together with how they are realized supports the application of a particular kind of framework to genes. The framework is centered on the notion of a signaling game, as originally developed by Lewis (1969), but extensively elaborated upon and ‘‘evolutionized’’ by Skyrms (1996, 2003, 2010) in more recent years.

A basic signaling game involves two players and an ‘‘external’’ source that takes on certain states in accordance with a fixed probability distribution. First, a state w from the set of possible source states W occurs with probability Pr(w).[6] The sender then observes the state and chooses a signal s from a set S to send to the receiver. Finally, the

[6] Think w as in state of the world.

receiver observes the signal and chooses an act a from a set A which has payoff consequences for both the sender and the receiver. Thus, the sender can see the state of the source but cannot act, while the receiver can act but cannot see the state of the source. Both the sender and receiver have a preference ordering over the acts for each state. These preferences can be in total or partial agreement, or else in total disagreement. The nature of the sender and receiver in such a game is characterized by the strategies they each implement or follow. The sender implements some strategy that maps source states to signals, while the receiver implements some strategy that maps signals to actions. For example, suppose W contains two equiprobable states, w1 and w2, S contains two signals, s1 and s2, and finally A contains two actions, a1 and a2. And suppose that the payoffs of the sender and receiver for each (state, act) pair are as shown in the matrix below:

	a1	a2
w1	1,1	0,0
w2	0,0	1.1

Figure 2.2

The first entry in each cell of the matrix is the payoff to the sender when the act heading the column is performed in the state heading the row, the second entry is the payoff to the receiver. Even in this simple game, there are a total of 16 possible combinations of sender-receiver strategies. A small subset of these combinations realizes a one-to-one mapping from states to signals, and from signals to acts, and also ensures that the act resulting in the highest payoff in each state for the sender and receiver is

performed (there are two combinations with these features in the above game). Such combinations of strategies are known as signaling equilibria. An example would be the pair of sender-receiver strategies: S1 = {w1 -> s1, w2 -> s2}, R1 = {s1 -> a1, s2 -> a2}. Only signaling equilibria support perfect or pure information transmission, so it's as if the receiver can directly observe the state of the source and the sender can directly act. Skyrms suggests that both the quantity of information as well as the informational content or meaning of a signal reduces to facts about how that signal moves the probabilities of various source states. To see this, suppose that the sender implements strategy S1, for example. It follows that she sends signal s1 if and only if the state of the source is w1, so the conditional probability of w1 given that s1 is sent is 1, and the probability of s1 has been moved by a factor of 2 (i.e., from 0.5 to 1). The quantity of information carried by s1, measured in bits, is given by the binary logarithm of this number, $\log_2(2) = 1$ bit, and the meaning of s1 just is that state w1 obtains. In contrast, were the sender to implement a strategy according to which she sent the same signal, s1 say, no matter what state the source was in, then her signals would be devoid of information content. This makes intuitive sense, but formally speaking, it's because the probability of each state conditioned upon the signal that is sent is equal to the unconditional probability of the state (i.e., $Pr(W|s1) = Pr(W)$).[7]

But what do the interests of the sender and receiver have to do with information transmission exactly? One reason the notion of a signaling game is so useful is that it

[7] Although not in this simple game, there will often be intermediate cases between these two extremes (pure information vs. no information) where the probabilities of some states will be raised but none will go to 1. One way to gloss information content in such cases is to see it as the set of states left with a probability greater than 0. Matters grow more complicated when no state is left with probability 0. See Godfrey-Smith (2012a) for a discussion of these issues. Such details are not of immediate relevance for present purposes, however.

provides a clear backdrop for posing and answering questions about the information carried by signals, as well as an explanation for how information transmission can evolve out of essentially random behavior and remain stable in a population. The interests of the sender and receiver are crucial to the latter explanatory project. When a signaling interaction involves rational agents capable of entering into explicit agreements beforehand (e.g., lantern signaling between the Sexton of the North Church and Paul Revere), or when the interaction is one that has been designed by such agents (e.g., a computer network), there will be no deep mystery surrounding the origin of information transmission in that system. However, when it is a population of unsophisticated agents (e.g., bacteria) simply evolving by natural selection, the fact that those agents have interests that can be differentially promoted by different combinations of sender and receiver strategies can explain the emergence of information transmission in the absence of reason or foresight. For example, in a reoccurring situation accurately modeled by our simple signaling game above, the fact that the sender's payoff is maximized by S1 when the receiver implements R1 provides the sender with continued reason to use S1, and vice versa. Importantly, we need not suppose that either player knows they have such a reason, or knows that the other has such a reason. It's enough to suppose that the strategies are heritable, and that higher payoffs correspond to higher biological finesses. If, by contrast, the sender were to always send the same signal regardless of the state of the source, the receiver would have no reason to pay attention to those signals; and likewise, if the receiver were always to perform the same act regardless of the signal observed, the sender would have no reason to discriminate the states of the source and send different signals for each such state. Questions about the evolution of gene signaling

must be left for another day, however; for now, I simply want to make the case that genes are indeed accurately modeled as agents in signaling games.

Finally, it is worth emphasizing that the game presented above is the simplest of all signaling games that have been developed and studied. Much of Skyrms' recent work has been focused on adding complexity to signaling games in various ways. As regards gene expression and regulation, the most relevant class of variations are those involving multiple senders and receivers, or what Skyrms calls 'networking' [see especially Skyrms (2009)]. Here we might have a signaling game with more than one sender, more than one receiver, individuals that are at once both a sender and a receiver, or any combination of these themes. Not surprisingly, patterns of information transmission and the evolutionary or learning dynamics supporting them grow more complicated in these settings, but are not fundamentally different.

Here is the heart of my proposal then. It is possible to reinterpret each of the IF–THEN instructions given in our table as expressing a particular sender or receiver strategy, or both (as will become clear shortly). More precisely, each such instruction can be seen as a pair of sender or receiver rules of the form:

(ia) $C(P_i, \ldots, P_j) \rightarrow P_k$;

(ib) $C(P_i, \ldots, P_j) \rightarrow P_k$

that together make up a complete strategy. Rule (ia) says that the gene is to make protein P_k when condition $C(P_i, ..., P_j)$ obtains, and rule (ib) says that the gene is to do or make nothing when condition $C(P_i, ..., P_j)$ does not obtain. There might be other, more complicated ways of interpreting the output behavior of each gene, specifically as some (continuous) rate of protein-production per some unit of time, but I here stick to the

simpler idea that the behavior is either transcription or no transcription. The interests of each gene are to be understood in terms of the biological fitness of each gene. A payoff to some gene is simply an increase or decrease (or possibly no change at all) in its fitness, and results from the protein- production behavior of both itself and other genes with which it interacts.

One might wonder how the concept of interests is supposed to apply to genes inside somatic cells. A gene located in a reproductive cell has interests in a perfectly straightforward sense: it stands to gain or lose a certain number of descendent copies of itself in future generations. Some ‘‘choices’’ will increase the chance that it will be passed along in a propagule sent out by the organism, and others will lower this chance. But genes in somatic cells are evolutionary ‘‘dead ends’’. Even so, these genes still have interests in an evolutionarily important sense, for they are identically related to genes in the reproductive line (somatic mutants aside). While the individual fitness of a somatic gene is effectively nil no matter how it behaves during development, the same is not true as regards its inclusive fitness (Hamilton 1964a, b).

What these remarks tell us is that the interests of genes are relevant to the emergence and stability of signaling interactions over the course of evolutionary or phylogenetic time, and only indirectly over the course of ontogenetic time. When the genes inside a cell engage in informative and cooperative signaling, this can be explained either in terms of those genes seeking to maximize the number of individual copies of themselves that will be passed on to future generations (in the case of genes in reproductive cells), or in terms of those genes seeking to maximize the number of copies produced by clonal ‘‘siblings’’ in the reproductive line (in the case of genes in somatic

cells).

Let us now consider the details of how genes can act as senders and receivers. First, consider a gene coding for a protein which we might call a ‘‘worker protein’’, that is, a protein which does not serve as a transcription factor for any other gene but rather performs some structural or metabolic function, for example. This gene is to be interpreted as implementing a particular receiver strategy. The rationale for this interpretation is just that the gene does not produce a protein that is ‘‘consumed’’ or used as a signal by other genes in the cell. A paradigm example here would be a gene coding for a specific antibody that functions to combat viruses inside the cell. Suppose that the IF–THEN instruction realized by this gene was of the form IF P_i & P_j THEN P_k. Put into signaling terms, this gene would thus implement a strategy according to which it is to make (‘‘perform’’) P_k when it observes P_i & P_j, and make nothing when it observes anything other than P_i & P_j (i.e., (P_i & P_j) -> P_k, ~(P_i & P_j) -> ~P_k). Of course, what we are here treating as an act on the part of the gene, namely, directing the synthesis of protein P_k, runs through an intermediate mRNA stage. I do not see this as posing a problem for my interpretation, however, as the translation of mRNA into protein is certainly a reliable effect which can be counted on by the gene.

Next, consider a gene that does code for a transcription factor for some other gene or genes inside the cell. It is natural to interpret this gene as implementing a particular sender strategy. The rational here is again simple: the protein product of this gene is used as a signal by other genes in the cell. Suppose the IF–THEN instruction realized by this gene is of the form IF P_i v P_j THEN Pk. Put into signaling terms, this gene would thus implement a strategy according to which it is to send P_k when it observes either P_i or P_j

and send nothing when it observes neither P_i nor P_j (i.e., $(P_i \text{ v } P_j) \rightarrow P_k$; $\sim (P_i \text{ v } P_j) \rightarrow P_k$).

And finally, consider a gene located somewhere in the middle of a gene hierarchy or cascade. Such a gene is to be interpreted as implementing both a sender and a receiver strategy. To see why, note that relative to the genes ‘‘above’’ it in the hierarchy, that is, genes that encode transcription factors for our focal gene, it will function as a kind of receiver, implementing a strategy that maps signals to acts, whereas relative to genes ‘‘below’’ it in the hierarchy, that is, those that our focal gene encodes a transcription factor for, it will function as a kind of sender, implementing a strategy that maps states of some source to signals. There is nothing mysterious or unusual about a single individual having a ‘‘dual role’’ in this sense. In any complex signaling network, there will be many such individuals. The linguistic label we attach to the protein(s) observed by such a gene (‘state’ vs. ‘signal’) and the protein it produces (‘signal’ vs. ‘act’) will simply vary with the perspective we take on the signaling interaction.

All of these remarks are intended to be taken quite literally. The complex structure of genes, and in particular of gene promoters, makes the application of the signaling games framework anything but metaphorical. For example, a typical promoter will be capable of receiving an enormous number of distinct input states. By ‘input’, I simply mean a combination of proteins serving as signals that can be either present or absent at the binding sites along the promoter. A promoter with just 4 binding sites per module, and 8 such modules altogether, for example, could receive 1 of 2^{32} possible distinct inputs at any given time! And this promoter will be physically positioned along the DNA molecule so that a particular protein will be synthesized only when some of these inputs are encountered by the promoter. Of course, not all of these different

promoter states will be made use of at any point in evolutionary time (i.e., not all of the states will be linked to a particular transcriptional response on the part of the gene because that response is adaptive given that state), but this kind of organization obviously allows for an extraordinary degree of sensitivity to cellular conditions.

Now if the application of the signaling games framework to genes is to be of real interest, at least two things should be true. The first is that, as regards some gene in the role of sender, the inputs that lodge in the gene's promoter must be capable of indicating the state of some source external to the promoter itself. The second is that, as regards some gene in the role of receiver, the inputs (the collections of signals) that lodge in its promoter must be capable of moving the probabilities of states of that external source. These two conditions actually boil down to the same one upon closer inspection. The crucial issue is whether the inputs to a gene's promoter will reliably tend to correlate with states of some external source, and it seems undeniable that they will. Before moving on, I want to present just a few toy examples that serve to drive home this point and in the process show how an application of the signaling games framework to genes can help to throw new light on the structure of genomes. I would not go so far as to claim that this structure is only visible through the lens of signaling games, but only that taking this perspective tends to draw attention to facts that we might otherwise miss or lack an appreciation for.

If we are interested in whether some gene receives information about some particular external source, the question to ask is whether and to what extent the inputs to this gene change the probabilities of the various states this source can be in. First, let's think of the source as something like the chemical composition of the cell at location L,

where L is the vicinity in which the gene is located in the cell's nucleus. There is a large number of possible states this source can be in at any given time, and receiving a specific input can reasonably be expected to make some of these states much more probable than their unconditional probabilities and some much less probable. (The input may, of course, leave the probabilities of a wide- range of other states unchanged.) For example, receiving an input containing three proteins of type P_i might significantly raise the probability that there is currently a dense concentration of P_i at location L. However, P_i may itself be inversely correlated with protein P_j in which case the input will at the same time significantly lower the probability that there is a dense concentration of P_j at location L, and so on. If actively making its protein in this state of the cell increases the (inclusive) fitness of the gene (either directly, or by inducing another gene to make or stop making its protein product), then we would expect for this input to facilitate transcription of the gene's protein-template. In other words, we would expect for the gene to be structured so as to map this input to the signal/action that is synthesis of the protein it codes for. On the other hand, just the opposite might be true, in which case we would expect the gene to block the synthesis of its protein when it receives this input.

When hierarchical coding is added to this picture, there is a huge increase in the information-processing power of certain genes. We implicitly touched upon this idea when we considered the sensitivity of a gene 8-layers deep in an AND hierarchy, but I now want to give a different example. Suppose there are two genes, Ga and Gb, respectively realizing the strategies: Sa = $\{P_1 \rightarrow P_2, \sim P_1 \rightarrow \sim P_2\}$, Rb = $\{(P_2 \,\&\, P_3) \rightarrow P_4, \sim(P_2 \,\&\, P_3) \rightarrow \sim P_4\}$. It is apparent that Ga encodes a transcription factor for Gb. When P_2 is part of the input to Gb's promoter, the input signals to Gb that P_1 is present at Ga's

promoter. But as we saw above, there is good reason to think that P_1's being present at Ga's promoter itself carries information about the state of the cell where Ga is located. This information can be directly transmitted to Gb. Now recall that Gb may itself encode a transcription factor for some other gene Gc, and so on.

Note that Gb can't bind P_1 to its promoter, so any information which is carried by inputs containing P_1 would not be available to Gb in the absence of hierarchical coding. It may seem strange for Ga to encode a transcription factor for Gb given that Ga can receive just 1 of 2 possible inputs (P_1 present/absent). Why wouldn't evolution outfit Gb with a binding site for P1 and dispense with Ga altogether? This misleading feature of the example can be corrected by noting that the condition sufficient for the synthesis of Ga's protein would really be a compound expression containing many distinct types of proteins. It is by virtue of this complex biochemical machinery (or what we might even describe as perceptual machinery) in promoters that genes come to realize one or another sender/receiver strategy. Replicating binding sites for each of these proteins in Gb's promoter would not be a trivial task, and that is to say nothing of replicating the functional connections between the modules making up Ga's promoter that confer upon Ga its discriminatory power. Obviously, adding in all of this detail would make our example too complicated to be helpful, so I have omitted it. The thing to appreciate is that, with hierarchical coding, information that is available to one gene in virtue of its promoter's structure can be transmitted to other genes without having to replicate that complex structure over and over again in the genome. This is how the expression of a gene in our 8-layer AND-hierarchy could be made to depend on the presence of 65,536 different proteins at particular places in the cell despite having only 4 binding sites on its

promoter.[8]

With more information being pooled together at each step in a gene hierarchy, it's easy to see how some genes might receive inputs conveying information about the global state of the cell. Indeed, it's clear how some genes might receive information about extra-cellular or even extra-organismal sources. All that is required is that the signals arriving at the promoter of the gene move the probabilities of the states of such sources. In this way, certain signals might reliably indicate to a gene in some cell that a part of the body is under attack or has sustained damage, or perhaps that there is a high density of predators in the environment or that food is scarce. This information could then be propagated around the cell to other genes, and passed on to genes in other cells via hormones or some other biochemical means. Given this possibility, I think it's worrisome that more attention has not been paid to the idea that at least some forms of mental representation and computation are realized at the sub-cellular level rather than at the level of the neuron or groups of neurons.[9]

Although I feel we have already given some compelling reasons for abandoning the ''genetic program'' program, the foregoing considerations serve to bring out yet another shortcoming. Specifically, the idea that genes encode a program does not sit at all well with the actual logic of gene signaling networks of the sort just described. In a computer program, it's just not the case that one instruction receives or processes

---

[8] Earlier I said that there is an element of idealization to these hierarchies owing to the fact that regulation of gene expression often involves some post-transcriptional processes, e.g., allosteric changes to transcription factors. I don't think this fact undermines the present line of argument; if anything, it actually strengthens it in the majority of cases. Transcription factors that must be ''activated'' by signal- dependent changes to their shape by other enzymes, for example, serve to integrate two distinct sources of information, namely, the information initially carried by the inactive form of the factor (that the promoter of the gene coding for it is ''activated'') and the additional information implied by the factor's assuming the modified form (that some other signal-dependent enzyme has made contact with it, say).

[9] See Gallistel and King (2009) for a discussion of sub-cellular representation and computation.

information from another instruction, which might then be routed to still another instruction, and so on. Rather, the instructions are entirely passive entities that are simply there to be read by the central processor.

The picture that emerges is one on which genes do not themselves encode regulatory or developmental information, but rather condition their output behavior on signals carrying such information that are received either directly or indirectly from other genes. Accordingly, genes themselves are the readers of this regulatory information, not some entity external to them. Unlike those who favor the program description, then, we do not wind up positing unread messages or information contents. Marcus has suggested that the IF–THEN structure of genes makes them like single lines in a computer program, but I think these dispositions are better thought of as particular strategies of senders and receivers whose behavior simultaneously causes, and is caused by, other genes in the genome. It seems to me that there remains much work to be done in unpacking the implications of this shift in perspective. In the remainder of the paper, I seek to take a first step in that direction by briefly considering how the present account of genes and gene action links up with a broader debate in the philosophy of biology concerning inheritance systems.

2.4 Genetic inheritance

As was mentioned in the introduction, it is certainly the orthodox view that genes are essentially information-bearing entities. There may be a number of reasons why informational descriptions of genes strike us as so natural (see Godfrey-Smith 2000), but whatever the reasons, this mode of thinking has permeated much of modern biology. Of

course, since genes are the stuff of genetic inheritance, being passed from parent(s) to offspring during reproduction, the process of genetic inheritance thus appears to us as a form of information transmission. Genes are signals or signs that are passed down the generations.

In their book The Major Transitions in Evolution (1995), John Maynard Smith and Eors Szathmary argued that the evolution of life from its simplest to its most complex forms can ultimately be understood in terms of innovations (and subsequent modifications on those innovations) for transmitting information trans- generationally. This idea has been quite influential in recent years and has been developed in various ways by different biologists and philosophers. In particular, I here have in mind Eva Jablonka and Marion Lamb (2006) who have sought to develop something like a general theory or framework for thinking about inheritance systems and their role in evolution. On their view, genetic inheritance is just one specific ‘‘channel’’ for communicating information between generations. Other examples of such channels include various epigenetic ones, behavioral or social learning channels, and finally symbolic channels associated with natural languages. They claim that information and its transmission is the ‘‘common denominator’’ shared by all inheritance systems (2007, p. 382). Thus, all inheritance essentially involves information; what differs is the particular channel along which it is sent. Nicholas Shea has also developed a similar view but adds to the above picture the idea that proper inheritance systems are distinguished from other means of bringing about parent-offspring similarity by the sending of signals with genuine semantic content (i.e., signals with correctness and/or satisfaction conditions) (Shea 2007, forthcoming). Each inheritance system for Shea involves an entity whose job it is to

produce a signal that carries correlational information of some kind and another entity who uses that signal to adaptively control its behavior (one in the next generation in cases of trans-generational inheritance). An attractive feature of these accounts is that they offer a unification of genetic inheritance with more familiar forms of communication between a sender and a receiver (such as animal signaling).

To be sure, there are a number of questions that can be raised about the above project. Perhaps the most obvious issue concerns the reality of the producers and consumers of information that is sent along the genetic (as well as some of the epigenetic) inheritance channels. For example, Shea claims that in the case of genetic inheritance the producer entity is a ‘‘whole, temporally-extended episode of natural selection’’ operating on a gene pool with multiple consumer entities each comprised of ‘‘the long complex process [of development] taking a selected genotype—through interaction with other parts of the organism, and with the environment, via the expression of multiple genes—to a phenotypic trait for which it was selected’’ (2013, p. 6–7, see also 2007). Godfrey-Smith (2007, 2011) has been critical of this idea. He suggests that genetic inheritance is better conceived of as a case in which genes are not sent by a producer entity at all, and so are more like cues than signs, and that the consumer of these cues is not a whole developmental system in Shea’s sense but rather the individual cell (Godfrey-Smith 2012b).

However the above debate pans out, I think it is clear that the view developed here raises a separate issue for those who would seek to treat genetic inheritance as essentially on par with other forms of information transmission or communication. The account of genes and gene action that emerges from our application of the signaling

games framework is one on which genes are not passive carriers of information, but rather active producers and consumers of information. This change in perspective throws completely new light on the nature of genetic inheritance. On this view, what organisms actually genetically inherit from their biological parent(s) is a rich and diverse network of information processing agents that function to adaptively control development by exchanging signals with one another in real time. They can engage in representing various cellular and extracellular states, request information from one another, and perform logical inferences allowing individual genes to ‘‘deduce’’ information that is directly behaviorally relevant to them. Thus, similarity between parent and offspring is genetically brought about by giving one's offspring some of the same information-processing mechanisms (genes) that guided one's own developmental trajectory. Assuming that the environment does not change too much, and that these inherited machines don't malfunction, offspring will indeed tend to resemble their parents.

That having been said, I do not think of the view developed in this paper as incompatible with the claim that the protein-templates of genes carry or encode information. There are indeed bona fide readers of these DNA regions in the cell, namely, transcriptional and translational machinery, and one might hold that this is enough to confer upon these regions information content of some kind. (By contrast, when the sender-receiver perspective is taken on genes, the protein-templates turn out to be the physical basis of the gene's ability to either send some signal or to perform some act). Thus, I do not take myself to have shown that the information- based approach to genetic inheritance is inherently wrong or flawed. Genetic inheritance can be conceived of as the transmission of information, realized by protein-templates, down the generations. But I would then

argue that it is unclear why that conception is superior or should be favored to one based on the idea of transmitting information processors down the generations. Interpreted in the latter way, genetic inheritance turns out to be fundamentally different from the other inheritance systems mentioned above, certainly as regards the behavioral and symbolic ones, but likely the epigenetic ones as well. Only in the case of genetic inheritance do organisms pass on entities that can be viewed as engaging in patterns of information production and consumption. This fact strikes me as something that should not be ignored.

References

Gallistel CR, King P (2009) Memory and the computational brain. Blackwell Publishing, Malden

Godfrey-Smith P (2000) Information, arbitrariness and selection: comments on Maynard Smith. Philos Sci 67(2):202–207

Godfrey-Smith P (2007) Innateness and genetic information. In: Carruthers P, Laurence S, Stich S (eds) The innate mind: foundations and the future. Oxford University Press, Oxford, pp 55–68

Godfrey-Smith P (2011) Senders, receivers, and genetic information. Biol Philos 26(2):177–181.

Godfrey-Smith P (2012a) Signals: evolution, learning, and information, by Brian Skyrms. Mind 120:1288–1297

Godfrey-Smith P (2012b) Sender–receiver systems within and between organisms. Presented at PSA 2012, San Diego, as part of a symposium on ‘‘Signaling Within the Organism.’’ Available online: www.petergodfreysmith.com/SRWithinOrgs_PGS_PSA_2012_G.pdf

Hamilton WD (1964a) The genetical evolution of social behaviour. Part I. J Theor Biol 7:1–16

Hamilton WD (1964b) The genetical evolution of social behaviour: Part II. J Theor Biol 7:17–52

Istrail S, De-Leon S, Davison E (2007) The regulatory genome and the computer. Dev Biol 310:187–195

Jablonka E, Lamb MJ (2006) Evolution in four dimensions. MIT Press, Cambridge

Jablonka E, Lamb MJ (2007) Pre´cis of evolution in four dimensions. Behav Brain Sci 30:353–392

Laudet L (2011) The origins and evolution of vertebrate metamorphosis. Curr Biol 21:726–737

Lewis D (1969) Convention: a philosophical study. Harvard University Press,

Cambridge
Marcus G (2004) The birth of the mind. Basic Books, New York Maynard Smith J, Szathmary E (1995) The major transitions in evolution. Oxford University Press, Oxford
Pearson H (2006) Genetics: What is a gene? Nature 441:398–401
Shea N (2007) Representation in the genome and in other inheritance systems. Biol Philos 22:313–331
Shea N (2013) Inherited representations are read in development. Br J Philos Sci 64:1–31
Shea N (forthcoming) Two modes of transgenerational information transmission. In: Calcott B, Fraser B, Joyce R, Sterelny K (eds) Signaling, commitment, and emotion. MIT Press, Cambridge
Skyrms B (1996) Evolution of the social contract. Cambridge University Press, Cambridge Skyrms B (2003) The stag hunt and the evolution of social structure. Cambridge University Press, Cambridge
Skyrms B (2009) Evolution of signaling systems with multiple senders and receivers. Phil Trans R Soc B 364:771–779
Skyrms B (2010) Signals: evolution, learning, & information. Oxford University Press, Oxford Yuh C, Bolouri H, Davidson E (2001) Cis-regulatory logic in the endo 16 gene. Development 128:617–629.

Chapter 3

Are genetic representations read in development?

To appear in *The British Journal for the Philosophy of Science*

DOI: 10.1093/bjps/axu043.

Talk of "signals", "codes", "programs:, and the like has come to permeate much of molecular and developmental biology. But do these terms really have a literal application to low-level biological phenomena? This question has attracted a good deal of attention from philosophers of biology in recent years. In particular, the claim that genes encode information about whole-organism traits has been hotly debated. While it is clear that genes carry information in a purely statistical sense about such traits, the same is true of at least some other cellular and environmental factors. The contentious claim is that genes carry *semantic information* that is about or relevant to whole-organism traits, whereby "semantic information", it is meant that the information has correctness and/or satisfaction conditions. Nicholas Shea's infotel account of genetic representation is the most detailed development of this idea to date (2007a, 2011, 2013).

Shea's account is based on his infotel theory of semantic content. This theory is an elaboration of the teleosemantic theories first developed by Ruth Millikan (1984) and David Papineau (1987). According to this family of theories, the content of a representation is determined by the evolutionary function of the representation. Shea maintains that a state cannot have the function of carrying information, since only the effects of a state are visible to natural selection; hence, he argues, teleosemantic theories do not entail that representations carry information about the things they represent

(2007b). Shea sees this as a flaw in the standard teleosemantics program and seeks to remedy it by adding an explicit information-carrying condition. The main motivation he offers for this condition is that, where success in behavior is explained by the having of a representation, it should be explained by the having of a true or correct representation, something which theories focused solely on the effects of a representation cannot deliver.

In order for Shea's infotel theory to apply to some target system, that system must contain: (i) a range of intermediate states or tokens serving as *inputs*, (ii) a *consumer* mechanism that receives those inputs, and finally (iii) a range of *behaviors*, performed by the consumer, in response to those inputs. In many cases, the system will also contain a *producer* mechanism. When present, the role of the producer is just to produce the intermediate states upon which the consumer acts. Provided that a target system has this kind of general structure (and that it satisfies several other requirements presupposed by the conditions below), the infotel theory tells us that the (indicative) contents of the intermediate states in the system are determined in the following way:

Tokens of type R have content C if:

(a) there is a consumer system which is caused by a range of tokens, including tokens of type R, to produce a range of outputs, with a specific evolutionary function for each type of output, and where every token satisfies (b) to (d) with respect to some content;

(b) Rs carry the correlational information that condition C obtains;

(c) an evolutionary explanation of the current existence of the representing system adverts to Rs having carried the correlational information that condition C obtains; and

(d) C is the evolutionary success condition, specific to Rs, of the output of the consumer system prompted by Rs. (2013, p. 5)

Shea has provided a detailed explanation of each of these conditions in several places, so I will not do so here. Shea's information-carrying condition is (b), which requires that a representation carry correlational information about the evolutionary success condition of the behavior the representation causes the consumer to perform. By 'correlational information', Shea has in mind what is commonly referred to as 'Shannon information', the concept at the heart of Claude Shannon's mathematical theory of communication (1948). A token of state type R carries information in this sense about the obtaining of a condition C just when the probability that C obtains, conditional upon R being tokened, is greater than the unconditional probability that C obtains. Or in symbols: tokenings of R carry correlational information about C if and only if Pr(C|a token of R) > Pr(C). Correlational information is a clearly ubiquitous feature of the world; that's one reason why several further conditions must be satisfied in order for states to carry semantic content or information.

Shea argues that genes will often satisfy the infotel conditions, in which case it is appropriate to ascribe semantic or representational content to them. (Strictly speaking, Shea thinks content can only be ascribed to individual genes in sexually reproducing organisms; in asexual ones, whole genomes wind up being the bearers of content. In this paper, we focus solely on genes in sexual reproducers.) Genes thus turn out to be the bearers of messages that can be either true or false, as well as of directives that can be satisfied or go unsatisfied. I think it's doubtful that the infotel conditions capture what biologists have in mind by the "gene for …" locution, but that is no objection to Shea's

account. The infotel theory provides an explanatorily powerful notion of content, one that is applicable in a wide-variety of contexts. If the theory applies in the case of genes, then this is an important and intriguing result in its own right.

Here's how a gene can come to satisfy the above conditions, according to Shea. Different genes (alleles) populate the gene pool of a species over evolutionary time. These genes are consumed by developmental systems associated with individual organisms, which in turn produce whole phenotypic traits as outputs. Suppose a specific gene G produces a phenotypic trait P that confers a fitness advantage on its bearer. G will then be naturally selected for, and thereby come to have the evolutionary function of producing P. The reason G was selected for, however, must be because some environmental condition E obtained (where 'environment' is construed broadly enough to include internal features of the organism). What this means is that the trait produced by G has an evolutionary success condition, namely, E. Because an increase in the frequency of Gs in the gene pool is due to the presence of E, it follows that Gs carries the correlational information that E obtains. Finally, in order to explain why G is transmitted and acted upon by developmental systems in the way that it is, we have to cite the fact that Gs carry correlational information about E over evolutionary time. Gene G will thus come to have the semantic content, *the environment is in condition E*. Shea writes: "zygotic DNA transmitted down the generations has semantic content. It indicates that environmental conditions are conducive to the phenotype, the production of which caused the genotype to be selected" (2013, p. 7).

For quite similar reasons, the infotel theory will wind up ascribing imperative content to G and other selected-for genes, though I will not present the details of how that

works here. The full semantic content of a gene, then, will have something like the form: *the environment is in condition E, make phenotypic trait P*. Genes are "pushmi-pullyu" representations, to use Millikan's terminology (1996).

Recent criticism of Shea's infotel account of genetic representation has focused on the nature of Shea's consumer mechanism, what Shea calls a 'developmental system'. Positing a consumer of genetic representations is one of the most distinctive features of Shea's account, serving to set it apart from earlier teleosemantic accounts such as those presented by Sterelny, Smith, and Dickson (1996) and John Maynard Smith (2000). But unless there really is a consumer of genetic representations, Shea's infotel theory won't actually apply to genes, and so won't vindicate the claim that genes have semantic content.

What is a developmental system, exactly? Shea conceives of such a system as the collection of "cellular machinery" and "environmental factors" that construct a whole organism on the basis of DNA (2007a, p. 325). More precisely, he says:

> When a [developmental system] act[s] on a selected genotype, transcription and translation are both part of the process of consumption, as are the complex interactions in which proteins fold, build structures, and carry out metabolic functions. The consumer is the long complex process taking a selected genotype—through interaction with other parts of the organism, and with the environment, via the expression of multiple genes—to a phenotypic trait for which it was selected. (2013, p. 6-7)

Peter Godfrey-Smith rejects Shea's claim that a developmental system in this sense is a genuine consumer mechanism (2007, 2011). Godfrey-Smith is not a critic of the notion of genetic information in general; he simply doubts that the infotel theory applies in the case of genes. The following passage nicely summarizes Godfrey-Smith's view on the issue:

> [Shea] thinks that a "developmental system" is the reader or "consumer" of the genetic message. This, for Shea, is not a situation where we have a separate "developmental system" in each cell. Rather, visible across the whole organism there is a single developmental system which reads all of the DNA that was present in the zygote, and which has been copied and distributed across many cells. This system reads a message which represents whole-organism phenotypic characteristics. I think, in contrast, that this "developmental system" is not a genuine reader of a message. It is an abstraction, and a reader should be a physical device. There are reader-like mechanisms in organisms, visible within cells, but putting a lot of these cell-level readers reading cell-level messages together does not yield an organism-level reader reading a message specifying phenotypic features found at the whole-organism level. (2011, p. 324).

I am sympathetic with Godfrey-Smith's conclusion, but I find his premises to be too strong. For example, I'm inclined to think that one can sometimes get a high-level message being read by a system from several or more low-level messages being read by part or parts of the system. A good case to think about here is the modern computer. In a modern computer, it is the ordered execution of many low-level instructions (machine language instructions) that actually enables the computer to carry out the operations familiar to lay users. Imagine, then, some computer system in which multiplication is realized by successive applications of an addition procedure, and moreover, that parts of the overall sequence of addition instructions are carried out in parallel by several processors within the system. Even though there wouldn't be any machine language instruction of the form *multiply x and y* read by a part of the system, it doesn't seem problematic to say that the system executes high(er)-level multiplication instructions through or by way of its processors executing sequences of low-level addition instructions. Also, Godfrey-Smith's insistence that the reader be a "physical device" rather than an "abstraction" may well be too strong, depending on how the notion of abstraction is understood here. Consider computers again. At "bottom", computers are

nothing more than systematically connected collections of logic gates. (In some cases, the computer might contain just a single type of gate, e.g., a NAND gate.) Each such gate receives current from a small number of input wires. In even talking about a 'central processor', then, we are already engaging in a kind of abstraction; we are lumping collections of low-level physical devices together and treating them as realizing a single consumer or reader. Yet, presumably, Godfrey-Smith wouldn't want to deny that the central processor is a *bona fide* consumer mechanism. I don't think either of these points actually work to Shea's advantage, but they do show that Godfrey-Smith's criticism needs an additional layer of detail in order to be convincing.

My criticism of Shea's account of genetic representation is related to Godfrey-Smith's, but approaches matters in a different way. Even bracketing issues about the reality of Shea's consumer (I will have more to say on this topic below, however), there is a different problem facing Shea's account. Shea not only claims that organisms contain genetic representations in his sense, but that these representations are "read in ontogenetic time, in the course of individual development" (2013, p. 1). For this reason, he thinks his account can shed light on "how the developmental processes itself unfolds" (2013, p. 9). In fleshing out these statements (the latter, in particular) Shea says such things as, "In organisms that pass through developmental stages, some of the information in the genome is not used until those later stages" (2013, p. 26), and also that development may have been adapted for "preserving genetic information until it is needed in development" (2013, p. 27).[10] In this paper, I want to argue that a close examination of the process of reading in Shea's sense reveals that acts of reading do not

[10] I here assume that by "information", Shea means *semantic* information, as all signs would seem to indicate.

actually occur over the course of developmental time at all. To make this vivid, I contrast the process of reading for Shea with another type of developmental process that is widely seen as a form of reading directed on inherited genes, and which certainly does occur over the course of developmental time, namely, gene expression. I suggest that this error in Shea's thinking can be traced back to an equivocation on Shea's part in the meaning of "reads", and also to a reliance on an invalid principle regarding the transference of representational content from one token gene to another.

As will become clear, the issues at play are bound up with questions about causation, and in particular, about causation over time. Thus, having first presented my arguments in a way that doesn't depend on any particular theory of causation, I then make use of Kenneth Waters' (2007) framework of difference making causation to conceptually sharpen and shed further light on matters. There are two main reasons I choose Waters' framework for this task. First, given that it is built upon the manipulability theory of causation (Woodward 1997, 2003, 2010), it is maximally permissive in what it recognizes as a real causal relationship between two things. I thus view this choice as being highly charitable to Shea in a domain where there continues to be much disagreement about the causal status of genes with respect to whole-organism traits. Second, Waters' framework is ideally suited to handling complex questions about causation over time. I emphasize, however, that the problems I raise for Shea's account do not depend on the correctness of Waters' view, or any other theoretical package.[11]

Finally, I conclude by discussing a consequence of the fact that genetic representations aren't read in development. I argue that this consequence casts additional

[11] I take this to be true even of the final critical point in the paper (presented towards the end of section 6), which builds upon our application of Waters' framework. That is, I think this point can also be made in a theory-neutral way, although I won't attempt to do so here.

doubt on Shea's claim that developmental systems are genuine representation consumers, the part of Shea's account with which Godfrey-Smith has taken issue.

3.1 Reading for Shea

How exactly does Shea understand the acts of reading performed by a developmental system? Numerous passages in Shea (for example, the one already presented in the introduction) paint the following picture. An act of reading begins at the time of fertilization with the insertion of a token gene into the zygotic cell of the organism. The act of reading this token then extends forward in time, encompassing all of the downstream consequences for the organism of that token having been inserted into the zygotic cell. Shea writes:

> … it is the zygotic DNA which has semantic properties, acting as a representational intermediate between generations. Those representations are consumed by the whole developmental system to issue in the phenotypes on which natural selection acts. ([2007a], p. 325)

As Shea himself repeatedly notes, such developmental processes are not naturally seen as forms of reading. For one thing, given that the majority of token zygotic genes (as well as their more proximal copies) are not ones that are actively expressed in development, most of the causal chains realizing acts of reading zygotic DNA will consist in nothing more than the replication of DNA and the distribution of copies to other organismal cells for prolonged periods. Why, then, does Shea conceive of the process of reading in this way? It certainly contrasts strongly with the way molecular and developmental biologists conceive of information being read off from inherited genes (more on this below).

Light can be shed on this issue by reflecting on the conditions of Shea's infotel theory of content, specifically, on condition (a): "there is a consumer system which is caused by a range of tokens, including tokens of type R, to produce a range of outputs, with a specific evolutionary function for each type of output, and where every token satisfies (b) to (d) with respect to some content". It's a requirement of the infotel theory that the intermediates or tokens, upon being read by the consumer mechanism, have to serve as causes of the output behavior of the consumer. (I take this to be equivalent to saying that acts of reading the intermediates or tokens have to serve as causes of the output behavior of the consumer.) Moreover, if the tokens are to have representational content, these output behaviors must be ones that were adaptive or successful, in the evolutionary past, because the condition indicated by the token obtained. In the present context (that is, given the sorts of representational contents Shea has in mind), this implies that the token genes, upon being read by the developmental system, have to cause whole-organism traits to develop. For example, in order for a token to come to possess the content *the environment is chilly*, it is necessary that the act of reading this token cause thick hair to develop at the whole-organism level. (This is one of Shea's own examples.)

Unfortunately, Shea's language at times obscures this causal requirement of the infotel theory. For example, in distancing himself from any form of genetic determinism, Shea writes that his account "only makes the uncontroversial assumption that there was a genotype-phenotype correlation at the time of selection" (2013, p. 7). This remark makes it sound like a causal relation between an act of reading some token that will serve as a

representation and some whole-organism trait isn't actually required by Shea's account; rather, all that's needed is a *correlation* between the act of reading and the whole-organism trait (indeed, a *past* correlation). But such a claim would be in direct conflict with the infotel theory. Moreover, I think it's safe to say that weakening condition (a) of the theory so that only a correlation is required would be sure to invite all sorts of counterexamples to the theory, although there isn't room to go into that issue here.

Having made this requirement explicit, it's no longer surprising that Shea conceives of acts of reading in the way that he does. First, provided that reading is understood in this way, acts of reading will, in general, satisfy the causal requirement (or so I here assume). This is because which traits develop at the whole-organism level can in general be changed by changing the DNA that is present in the zygotic cell at the time of fertilization. For example, supposing that there's a gene for thick hair in some population, then it may well be the case that whether or not an organism goes on to develop thick hair (as opposed to thin hair, or perhaps no hair at all) essentially depends on its receiving a copy of this gene, rather than a copy of some other gene, at the time of fertilization. Second, it's not at all clear that there's any other type of developmental process that this is true of, at least across the board (that is, for all of the whole-organism traits that Shea would wish to treat as outputs of the developmental system).

Thus, commitment to a highly counterintuitive conception of reading is likely the price Shea must pay in order to secure that claim that there are genetic representations of the sort he has in mind at all. But that is not what I wish to take issue with here. Rather, my objection concerns Shea's claim that these genetic representations are "read in ontogenetic time, in the course individual development" (2013, p. 1). If, as we've just

seen, all of the acts of reading performed by the developmental system begin at the time of fertilization (typically, involving nothing more than DNA replication and distribution of copies to daughter cells for extended periods), then it is false to say that genetic representations are read over the course of developmental time. Shea cannot hold, for example, that some representational contents aren't read until they are needed at later developmental stages, that certain features of the developmental process may be adaptations for preserving representational contents over developmental time, and so on.

Of course, where G_z is a zygotic gene having representational content, there will be many type-identical copies of G_z distributed throughout the organism. These copies will all have the same DNA-base sequences as G_z, barring mutation. However, I think it would be wrong to conceive of the non-zygotic copies as also having representational content, at least on Shea's infotel theory. For one thing, unless there's some very special story to tell here, such non-zygotic copies won't count as being read at all by the developmental system, but rather as mere constituents of a temporally-extended process of reading token G_z that began at the time of fertilization.[12] This is important because, on a user-based semantic theory like Shea's, there can be no representation without representation *use*. For another thing, it will in general be true that the vast, *vast* majority of non-zygotic copies of a zygotic gene will reside at far too low a level to count as representations. As regards a given non-zygotic gene in the organism, it's overwhelmingly likely that this token could be replaced by an alternative allele

[12] Could the process of reading G_z really contain as temporal parts millions or billions of other acts of reading type-identical copies of G_z? What could possibly be the point of such an exercise from an informational perspective, given that each copy will be associated with the same content?

(excluding super 'selfish' and/or lethal ones), or simply knocked out altogether, without bringing about a functional difference at the whole-organism level.[13]

In short: it is tempting to think that the representational content carried by a zygotic token will be transferred to all of the token genes in the organism that are sequence-identical copies of it, but such a *copy* or *transference principle* is in fact at odds with Shea's infotel theory. This is not to say that non-zygotic genes won't have associated with them the kind of *correlational information* Shea draws attention to; indeed, they will. But this information will simply be "along for the ride" in the way that trunk rings carry information about tree age that is (typically) "along for the ride".

3.2 Gene expression as reading

I now want to switch gears and consider an altogether different type of developmental process, namely, gene expression (i.e., the active transcription and translation of the protein-coding portion of a gene). Doing so will prove useful for several reasons. First, gene expression is naturally understood as a form of reading that is directed on inherited genes, one which certainly does occur over the course of developmental time. Thus, by juxtaposing acts of reading in Shea's sense with instances of gene expression, the fact that the former do not occur in development is brought into further relief. Another reason is this. Given the "credentials" of gene expression, so to speak, one may well wonder whether Shea's account can be modified so that it treats acts of reading as being realized by instances of gene expression (or at least as beginning with

[13] The latter, at least, occurs all of the time in development, as when a diseased cell is killed off by the immune system, bringing about the destruction of all of the cell's contents, token genes included.

instances of gene expression[14]). So, putting gene expression on the table at this point provides us with an opportunity to answer this question. Crucially, I think the answer is an emphatic "no". And finally, still another reason is that, while Shea pretty clearly doesn't identify reading with expression in explaining how genes can come to satisfy the conditions of his infotel theory, when he claims that genetic representations are read in development, the meaning of "reads" in play appears to have shifted. More specifically, it's hard to make any sense of this claim without assuming that, by "reads", Shea means express. This section discusses the first two points; in the next section, we turn to the third.

The process of gene expression is already widely viewed by molecular and developmental biologists, as well as many philosophers of biology, as essentially a form of reading (indeed, as *the* way in which information is read off from genes). Several features of the process contribute to it's being viewed in this way. First, there is a clear separation between what is read (the gene) and what does the reading (transcriptional and translational machinery in the cell), with independent motivation for recognizing both the message and its reader as real entities. Second, gene expression is localizable in both space and time, and is furthermore a repeatable process (the same gene can be read over and over in development). Third, expression is the normal means by which a gene is given causal efficacy within a developing organism. A gene that is nowhere expressed in an organism (think of a heavily methylated gene) will not ordinarily be able to affect the goings-on of development. Fourth, how the developmental process unfolds is highly

[14] In other words: on this view, acts of reading would still be (typically long) temporally-extended processes, but would *begin* with an instance of gene expression. In what follows, I leave this option aside, and simply talk of the possibility of Shea treating acts of reading as being realized by instances of expression.

sensitive to the order and conditions in which various genes are expressed. In all of these respects, the process of gene expression quite closely resembles processes of reading outside the biological domain, and in particular, those processes in modern computing devices that are commonly described as processes of reading.

But the most crucial point for present purposes is simply that gene expression does indeed occur over the course of developmental time. A token gene might not be expressed until a later stage of development, or not expressed at all. For this reason, it makes perfect sense to say that information carried by the token is merely preserved until it's needed at some later developmental stage, say, when a particular whole-organism trait is about to start to develop (and so that developmental mechanisms may exist which have been selected for preserving/protecting that information). In short: on this usage of "reads", a token gene doesn't count as read until that token is expressed. This isn't true on Shea's account.

Given these features of the process of gene expression, it's certainly natural to wonder whether Shea's account can be modified so that acts of reading are taken to be realized by instances of expression. (Actually, I think it's fair to insist that any adequate account of genetic representation will take expression to be the relevant form of reading, though I won't argue for that claim here.) To see why the answer is "no", we need only return to the causal requirement of the infotel theory mentioned above. Here I'll make the point using language that a molecular or developmental biologist might use. Setting aside some very unusual whole-organism traits, no individual instance of expressing a token gene will be *sufficient* to produce a trait at the whole-organism level. Indeed, in many

cases, neither will such an individual instance of expression be *necessary*.[15] What *is* true is that there being very many instances of expressing copies of the same type of gene will often be at least a necessary condition for the development of some whole-organism trait. For example, it's not at all unreasonable to expect for the expression of many copies of a gene, ancestral forms of which were selected for producing thick hair, to be required for the development of thick hair. But any instance of expressing a single one of these copies could likely fail to occur without there being a functional difference at the whole-organism level. Thus, were we to think of the development of thick hair as being (even partially) caused by the expression of a single token gene in the organism, we would be guilty of making a type-token error. Individual instances of expression reside at much too low a level in the organism.

A very special case here concerns the so-called "master genes" (e.g., eyeless). This term is used by molecular and developmental biologists to refer to genes whose expression is both necessary and sufficient for the development of a whole-organism trait, such as an eye or an antenna (see, e.g., Gehring 1998, Carroll 2006). However, upon closer inspection, one sees that it's not the expression of a single *token* of the master gene that causes the relevant whole-organism trait to develop, but rather a collection of tokens of the same type. (In other words, to "make an eye", it's not sufficient to express a single token of the eyeless gene; rather, a collection of copies of eyeless need to be expressed in

[15] Roughly speaking, an instance of expressing some token gene would be *sufficient*, on the present usage of 'sufficient', just if that instance of expression *alone* produces the whole-organism trait in all of the genetic, cellular, and environmental conditions that would be deemed 'normal' by biologists. What counts as normal here would be relative to a particular population of organisms, likely one in which the organisms share most, if not all, of the same genes. (The population might be a merely possible one.) And an instance of expressing some token gene would be *necessary*, on the present usage of "necessary", just if the trait does not develop in any genetic, cellular, and environmental conditions that would be deemed 'normal' by biologists unless that instance of expression occurs. The same remark about the relative nature of the normal conditions applies in this case as well.

a region of the body.) Moreover, in order to build a whole-organism trait having different functional parts, each token of the master gene has to be expressed in a subtly, but importantly, different cellular context. Only then will it be possible for the expression of one token of the master gene to trigger a cascade of gene expression events that will produce one part of the trait (or a part of a part), and for expression of another token of the same type to trigger a different cascade that will produce another part. Thus, even in the case of master genes, it's really the expression of a collection of genes of the same type that is necessary and sufficient for the development of the whole-organism trait.

It's worth pausing to respond quickly to an objection that might be raised at this point. In discussing the passage from Godfrey-Smith above, I suggested that in some cases multiple acts of reading low-level messages may well realize an act of reading a high-level message at the system level. Why doesn't this logic apply in the present context? I don't want to rule out that it does. That is, I'm not sure it would be wrong to say that messages at the whole-organism level are read in development, where acts of reading these messages are realized by many acts of reading low-level messages occurring both sequentially and in parallel. My point here is simply that none of the token genes in the organism will themselves be bearers of messages at the whole-organism level, if reading is understood as expression. Rather, the whole-organism messages would instead exist (if they exist at all) only at a level of abstraction, in much the way that the word-processing messages read by a desktop computer exist only at a level of abstraction. This is not the view that Shea holds, though; for Shea, at least some of the token DNA sequences in an organism are bearers of whole-organism representational contents.

3.3 Are genetic representations read in development? (Part 1)

When viewed side-by-side with instances of gene expression, it couldn't be clearer that acts of reading in Shea's sense do not occur over the course of developmental time. Shea cannot both hold that, for every genetic representation in the organism, the act of reading that representation begins immediately at the time of fertilization, *and* hold that some of these acts occur in development.[16] Moreover, we've just shown that Shea's account can't be modified so that it treats acts of reading as being realized by instances of gene expression.

As I mentioned earlier, I suggest that this error in Shea's thinking can be traced back to two things: first, an equivocation on Shea's part in the meaning of "reads", and second, a (tacit) appeal to what I above described as a kind of copy or transference principle. Shea relies on one sense of "reads", the sense on which acts of reading are temporally-extended processes beginning at the time of fertilization, to establish that

[16] This conclusion is quite similar to that reached by Paul Griffiths in his recent article "Lehrman's Dictum: Information and Explanation in Developmental Biology" (2012). As I read Griffiths, he offers two separate criticisms of Shea's account. First, Griffiths argues that Shea confuses two things, namely, providing examples where the semantic information carried by one or more genes matters to how a developmental process unfolds, and providing examples of features of developmental processes that appear adaptive when one takes an informational perspective on genes. Griffiths thinks that Shea (at best) succeeds in offering examples of the latter type, although Shea seems to understand himself as providing examples of the former type. I would argue that Shea has not even provided us with examples where taking an informational perspective illuminates the features of developmental processes, assuming that by 'information' we mean the sort of representational content Shea has in mind. Griffiths' second criticism is that teleosemantic information cannot even in principle play a role in explaining how development unfolds because two physically identical genes, one with selectional properties and one without, will make the same physical contribution to development, thus showing that "the teleosemantic content of a developmental cause makes no difference to what happens as a result of that cause" (2012, p. 25). (Griffiths notes that Shea himself makes this point in an earlier publication). While I think Griffiths may be correct as regards mechanistic explanations of development, I would in general disagree that teleosemantic information cannot make a difference to how development unfolds. To me, it's a mistake to think that the informational properties of a representation, and that representation's physical properties (the properties of its vehicle), are in competition with one another. Thus, to show that the vehicle of a representation carrying teleosemantic content or some other form of semantic content plays a certain causal role in development just is to show, on my view, that the information carried by the representation plays a causal role in development, that the information matters to development. To be sure, this is a complex issue and discussing it here would take us well beyond the scope of this work. The thing to note is just that, while Griffiths and I arrive at the same or a similar conclusion (i.e., genetic representations are not actually read in development), we arrive at it via quite different routes.

there are genetic representations. That reading be understood along these lines is very likely necessitated by the causal requirement of the infotel theory. Having then established that there are genetic representations in the organism, an appeal to the transference principle is made: where G_z is a zygotic gene having representational content, all of the sequence-identical non-zygotic copies of G_z, namely, G_{c1}, . . ., G_{cn}, will have the same representational content as G_z. Finally, noting that many token genes are only expressed later on in development, as the organism passes through different stages marked by different forms and/or amounts of uncertainty about how it should develop, Shea conceptualizes these later instances of expression as involving the reading off of representational contents that have been transferred from the zygotic genes to non-zygotic genes. Some evidence that Shea in fact relies on a principle of this kind is provided by a remark he makes about DNA replication: "DNA replication is the basis of representation transmission (from one collection of zygotic DNA to its descendants)" (2013, p. 7). And as for a shift in the meaning of "reads": I can make no sense of the idea that some representations aren't used until later developmental stages except by taking gene expression to be the relevant form of reading.

But as we've seen, this transference principle doesn't hold in the present context; nor can instances of gene expression be taken to realize the acts of reading performed by the developmental system. Note as well that, even if Shea were to concede that the sense of "reads" needed to establish that there are genetic representations at all isn't the same sense of "reads" on which it's correct to say that genetic representations are read in development, this would do little to improve his overall situation. The reason is that the failure of the transference principle here implies that the non-zygotic genes in the

organism will all be "empty hulls", semantically speaking. Consequently, the expression of these genes later on in development can't be taken to realize acts of reading representational content in any sense of "reads". These genes aren't representations, at least not ones of the sort Shea has in mind.

Having presented my arguments in a way that doesn't depend on any particular theory of causation, I now want to restate the core of my case within the framework of difference making causation developed by Kenneth Waters (2007). The next section sets out Waters' framework and the manipulability theory of causation it's built upon. In the section after that, we return to our discussion of Shea's account.

3.4 The manipulability theory of causation and difference making causation

Waters' difference making framework consists in a family of related causal concepts, namely, *a potential difference making cause*, *the actual difference making cause*, and finally, *an actual difference making cause*. These concepts are built upon the manipulability theory of causation developed by Woodward (1997, 2003, 2010). On a bare bones version of this theory, a causal relation between a variable X and a variable Y is analyzed as follows:

> [(MTC)] 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 except through changes in X) were to occur in B, then the value of Y would change. (Woodward 2010, p. 290)

For our purposes, most of the details of even this bare bones version can be safely set aside. The key notions are those of a *variable*, an *intervention*, and *background circumstances*. A variable is held to be any property or feature of an entity that is capable of taking on at least two different "values". More technically, it has been suggested that

variables are *determinable* properties of entities, the *determinates* of which are the variable's values (Woodward 2003).[17] An example of a variable would be the state of a switch which can either be "on" or "off". An intervention on some variable X with respect to some variable Y is a (possibly idealized) manipulation of X such that, if the value of Y changes, this change in Y occurs only through the change in X. For example, I might change whether or not my car is running (Y) by turning the key in the ignition (X). Finally, the background circumstances are variables as well, but not ones that are explicitly mentioned in the X-Y causal link. As the background circumstances consist in a set of variables capable of taking on values of their own, we can ask further questions about the range of values the X-Y link is stable or continues to hold under, revealing what is known as the *stability* or *invariance* of the causal relation (the larger this range is, the more invariant is the causal link).

Waters uses the above mentioned concepts to attempt to solve a long-standing philosophical puzzle about causation, which he lays out as follows:

> Mary's striking the match counts as a cause . . . because if Mary had not struck the match, then the match would not have lit. But the same reasoning leads to the conclusion that the presence of oxygen was also a cause; if oxygen had not been present, then the match would not have lit. There is no denying that oxygen was a real cause. Nevertheless, in ordinary contexts, observers would identify Mary's striking the match, and not the presence of oxygen, as the factor that made the difference. And in an important sense they would be right. The puzzle is, in what sense? (2007, p. 551–2)

We will oversimplify Waters' account somewhat, and focus only on what is essential for our purposes. Once the above mentioned concepts have been fleshed out,

[17] An anonymous referee for this journal pointed out to me that there remains a good deal of uncertainty about exactly what variables are. For example, saying that variables are determinables leaves many questions wide open, as there is still no consensus on what exactly determinable properties are. There is not space to go into these issues here. For more recent discussions of the determinable-determinate relation, see Johansson (2000), Denby (2001), Funkhouser (2006), Crane (2008), and Wilson (2012).

however, Waters' proposed solution to this philosophical puzzle will be obvious.

The concept of a potential difference maker is the simplest to grasp. Let us stick with Waters' own example of a match being caused to light. More precisely, let Y be a variable whose values are *lit* and *not lit*. Then we can ask, 'Is match striking a cause of match lighting?'. In this case, what we are doing is asking about how values of a variable X (the values of which are *strike* and *not strike*) are associated with changes in the value of Y. Put in these terms, it's immediately obvious that MTC yields an affirmative answer to our causal query. This is because there are indeed background circumstances B such that, intervening on variable X to change its value from *strike* to *not strike* would serve to change the value of variable Y from *lit* to *not lit*. That having been said, it's also the case that the presence of oxygen counts as a cause of the match lighting according to MTC. To see this, we simply treat the match being struck as part of the background circumstances, in which case whether or not the match is lit can be manipulated by intervening on the presence of oxygen. Thus, both striking the match and the presence of oxygen are causes of the match lighting on the present view, as there are indeed background circumstances in which both are associated with changes in whether or not the match is lit (hence the above puzzle). In this sense, they are both *potential difference makers* or *potential difference making causes* of match lighting.[18]

However, in actual cases, there will virtually always be an important asymmetry between these two causal factors. This asymmetry tracks the difference between a potential and an actual difference making cause, and to see it, consider the last time you

[18] I'll sometimes speak, not just of a causal relation between a variable X and a variable Y, but also of a causal relation between one value x of X and one value y of Y, when changing X to x would suffice to change Y to y. This way of speaking simplifies the exposition a great deal. One could always rephrase talk of variable values causing other variable values with the fuller, more technically correct formulation in accordance with MTC.

actually struck a match thereby causing it to light. It is certainly true that if oxygen had not been present on that occasion, then the match wouldn't have lit. But oxygen was present in your environment long *before* the match lit, and so it was not a change in the presence of oxygen that *actually* caused the change in the match. Thus, the reason (says Waters) it makes sense to single out the striking of the match as 'the' cause of the match lighting is that, among the many potential difference making causes of that effect, the striking of the match is the actual difference making cause, the cause that actually made the difference.

Waters explains that, to identify the actual difference maker of some effect (a change in an effect variable), the first step is to identity the actual variation in the effect variable that we wish to explain. If there is no actual variation, there can be no actual difference making cause of it. Waters describes this step as identifying variation within some *population*, where the term 'population' is used broadly to cover both collections of different entities as well as collections of different temporal stages or time-slices of a single entity, as in our match lighting example above (where we were comparing actual variation in the match stick over time). Here our focus will mostly be on populations of the latter sort. The second step is this. Of all of the variables that serve as causes of that population variation (i.e., the variables which, by intervening on them, one could change the value of that population variation), the actual difference making cause is the cause variable that actually varied in the situation, and in so doing, caused the change in the effect variable. Thus, just as with the effect variable, there must be actual variation in at least one of the cause variables if there is to be an actual difference making cause at all. This applies to our above example (i.e., the last time you struck a match) in the following

way: while both the presence of oxygen and match striking are potential difference makers of match lighting, only match striking actually varied on this occasion, and so match striking counts as the actual difference making cause. There was no actual variation in the presence of oxygen, recall; oxygen was present both (long) before the match was lit, as well as when the match was lighting. It is for this reason that Waters thinks we're 'right' in some sense when we single out the striking of the match as the cause of match lighting (this is Waters' solution to the puzzle).

The final concept is that of *an actual difference making cause*. In a nutshell, when some actual variation in an effect variable is accompanied by more than one form of actual variation among cause variables, then we must instead talk of the *set* of actual difference making causes, the individual members of which are each an actual difference maker. Here is one toy example of a case involving more than one actual difference making cause. Imagine we are interested in why Alice went from being healthy at time t (say, her teen years) to being unhealthy at time t+*n* (her later adult years). It is the actual variation in Alice's health over time that we would thus like to explain. Both eating a poor diet and smoking are causes of being unhealthy, and Alice unfortunately acquired both habits between t and t+*n* (say, during her college years). So there was actual variation in two distinct cause variables over the relevant time scale, i.e., she went from eating a good diet and not smoking at time t, to eating a poor diet and smoking between t and t+*n*, and this variation caused the decline in her health. However, as there is no single actual difference making cause of the effect in this case, we can only say that poor diet was an actual difference maker and that smoking was an actual difference maker. That having been said, there may still be important differences in how each actual cause bears

upon its effect (one causal relation might be more invariant and/or more specific than another).[19]

3.5 Are genetic representations read in development? (Part 2)

It's perfectly natural to view a developing organism as persisting though time in virtue of being composed of a set of temporal stages or time-slices. We thus have a population in one of the two senses specified by Waters. Moreover, for a typical multicellular organism, there will be much actual variation between the different temporal stages making up the organism. Regarding a whole-organism trait like a wing, for example, there will be stages at which it's clear that the trait has not yet begun to develop, others at which it's clear that the trait is in the process of developing, and still others where it's clear that the trait has reached its mature form. It is the main business of biologists focused on the developmental process to describe the relative time course, and the mechanisms underlying these changes in the organism's phenotype.

Next, let's think about how acts of reading token genes in Shea's sense might be represented within Waters' framework (or more specifically, on the manipulability theory of causation the framework is built upon). I propose the following (which is admittedly

[19] Skepticism about Waters' framework has been set aside in the present paper, but it is worth flagging two issues at this point. First, Waters argues that whether some change in a cause variable is an actual difference making cause of some variation in an effect variable is a purely ontological matter. Thus, Waters challenges longstanding philosophical orthodoxy which has it that the singling out of one cause among many in an explanation of an effect is guided by pragmatic considerations. However, Robert Northcott (2009) has recently argued that the selection of one population over another partially constrains which variables can be said to actually vary, and so the status of some change in a cause variable as an actual difference maker would appear to depend in part on certain pragmatic or interest-relative considerations after all. Second, Waters also argues that where biologists single out DNA as 'the' cause of some effect in what is really a complex web of causes, (activated) DNA will figure as an actual difference making cause of the effect within the population under consideration. This claim has recently been criticized by Ulrich Stegmann (2012), who provides what would appear to be counterexamples to this claim from the domain of molecular biology. For our purposes, neither one of these criticisms of Waters needs to be challenged. The machinery Waters develops is simply being used to highlight independently-existing problems for Shea's account.

oversimplified in several respects). Associated with some focal organism is a collection of variables of the form Read_Gz_1, Read_Gz_2, . . ., Read_Gz_n. (The "z" in the variable name indicates that the token being read is a zygotic gene.) More precisely, there will be a unique variable corresponding to each chromosomal locus of the zygotic chromosomes of the organism. The different values of a given variable Read_Gz_i are the different genes (alleles) that can reside at the corresponding locus, and so be read (in Shea's sense) by the developmental system. In addition, there is a collection of variables of the form Trait_P_1, Trait_P_2, . . ., Trait_P_n, one such variable for each whole-organism trait that counts as a type of output behavior of the developmental system associated with the organism. The values of these variables just are the possible character states of the corresponding trait, including the value *is absent*. (For simplicity, we can assume that at the time of fertilization, the value of each one of these variables is set to the value *is absent*.) The causal requirement of Shea's infotel theory, you will recall, is that the intermediates or tokens, upon being read by the consumer mechanism, have to serve as causes of the output behavior of the consumer. So, in the present context, the causal requirement comes to this: for a given variable Read_Gz_i, there exists a unique[20] variable Trait_P_i, such that, a causal relation exists between Read_Gz_i and Trait_P_i. That is, there must be background circumstances B such that the value of the latter variable can be changed by an intervention on the former variable.

I assume that there will typically be such a causal relation, or that there will be such a causal relation often enough for Shea's account to remain attractive. My objection

[20] Such a uniqueness assumption can likely be relaxed. I make the assumption here just to keep the presentation simple. Shea too typically makes such an assumption: "[On] [t]he 'basic representational model' [the] consumer system produc[es] a different type of behavior in response to each type [of token] on which it acts" (2013, p. 4).

instead concerns Shea's claim that acts of reading genetic representations occur over the course of developmental time. Here Waters' framework becomes quite helpful. We can say that an act of reading a token gene occurs in development if and only if (i) there is actual variation in whether the token is being read between at least two temporal stages making up the organism, and (ii) this variation consists in the token *not* being read at the earlier of the two stages and that same token being read at the later of the two stages. This analysis captures the intuitive idea that if a token gene is read in development (i.e., if there is an act of reading that token that occurs in development), then there exists some developmental stage prior to its being read at which the token is merely being carried forward in time. Having explicitly formulated this idea, it's easy to see what's wrong with the claim that acts of reading in Shea's sense can occur in development. Put simply, the problem is that *there can't be variation in whether some token is being read by the developmental system over developmental time at all*.[21] For every input received by the developmental system, the process of reading that input (that token) begins immediately upon fertilization, and extends forward in time to encompass all of the downstream consequences for the organism of that token having been inserted into the zygotic cell. It thus makes no sense to say that some token is merely carried forward in time for a period and is only read at a later developmental stage.

We can again make all of this quite vivid by contrasting acts of reading in Shea's sense with instances of gene expression. Associated with every token gene in the organism is a variable of the form Express_G, having as its values *is being expressed* and

[21] Note that, even if acts of reading for Shea were conceived of in a way that allowed for them to reach completion at early or intermediate developmental stages, this still wouldn't help. It still wouldn't be the case that there exist stages *prior* to the acts of reading at which the tokens are merely being carried forward in time.

is not being expressed.[22] Or were we to take gene expression to comprise a form of reading that's directed on inherited genes (as is the norm in biology), we might conceive of these variables as taking the form Read*_G, with values *is being read** and *is not being read**. (The "*" in the variable and value names indicates that this is a distinct usage of 'reads'.) Now, it's entirely accurate, both on conceptual and empirical grounds, to say that acts of reading in *this* sense occur over the course of developmental time. A given token will be read in development just when there exists at least two developmental stages making up the organism, the earlier of which has the variable associated with the token being set to the value *is not being read**, and the later of which has this same variable being set to the value *is being read**. In such a case, there will indeed be actual variation (of the appropriate kind) in whether this token is being read relative to the population of temporal stages making up the organism.

It's also possible to sharpen the claim that Shea's account can't be modified so that it treats acts of reading as being realized by instances of gene expression. Above I put the problem this way: any individual instance of expression will reside at too low a level for the token being expressed to count as a representation of the sort Shea has in mind. What *is* true, generally speaking, is that there being many instances of expressing token genes of the same type will serve at least as a necessary condition on the development of a particular whole-organism trait. Drawing on the tools provided by Waters' framework, we can now summarize this line of reasoning as follows. For a given variable Express_G_i (or Read*_G_i), it's generally not the case that there exists a (unique)

[22] It's worth pointing out that the way I'm using variables here departs from the determinable way of thinking about variables. For example, not being expressed isn't a specific way of being expressed, etc. The present formulation helps to keep matters as simple and clear as possible. As far as I can tell, one could re-conceptualize the variables associated with token genes so that *is being expressed* and *is not being expressed* turn out to be determinates of such variables. I set this issue aside in what follows.

variable Trait_P_i, such that, the value of the latter can be changed by an intervention on the former (where the intervention in question would serve to change the value of the former from *is not being expressed/read** to *is being expressed/read**). That having been said, it will often be the case that the value of Trait_P_i can be changed by intervening on a (typically large) *collection* of variables of the form Express_G_i, Express_G_j, . . ., Express_G_k, where the relevant token genes are all tokens of the same type. In such a case, it would be appropriate to say that the *aggregate* of these individual instances of expression not only serves as a cause of the whole-organism trait, but also as an *actual difference making cause* of the trait, relative to the population of temporal stages making up the organism. More precisely stated: there would be a temporal stage t of the organism at which Trait_P_i has the value *is absent*, and another (later) temporal stage t+*n* at which Trait_P_i has come to assume some other value (a particular character state), and finally, this variation in Trait_P_i will have been caused by variation in a collection of variables Express_G_i, Express_G_j, . . ., Express_G_k, where each variable has changed from *is not being expressed* to *is being expressed*, sometime between stage t and t+*n*. Or less tortuously: variation between temporal stages regarding the development of the trait will have been caused by variation in the expression pattern of a collection of token genes of the same type over the appropriate time scale. However, for our purposes, the crucial thing to note is that any individual instance of gene expression is neither capable of producing the whole-organism trait, nor is it likely to even be required for the trait to be produced. Thus, taking the process of gene expression to be the relevant sense of "reads" on Shea's account would run afoul of the causal requirement of the infotel theory. The tokens being read (i.e., expressed), wouldn't count as representations at all.

To summarize, we've seen that acts of reading in Shea's sense, while serving as causes of whole-organism traits, do not occur over the course of developmental time. For a token to count as being read over the course of developmental time, there must be variation in whether that token is being read between two or more developmental stages making up the organism, where this variation consists in the token not being read at the earlier stage and being read at the later stage. Such a situation simply isn't possible as long as reading is understood in Shea's way. Moreover, couching matters within Waters' framework, we see that acts of reading in Shea's sense do not serve as actual difference making causes of the traits the representations are posited to explain (or of anything else in development for that matter), relative to the population of temporal stages making up the organism. In this respect, such acts contrast markedly with aggregates of instances of gene expression in development which do serve as actual difference making causes of whole-organism traits, relative to the same population. Since, presumably, the developmental system is made up of the same set of temporal stages as the organism, it's also true that acts of reading for Shea do not serve as actual difference makers with respect to the output behavior of the developmental system, relative to the population of temporal stages making up the developmental system. In the remainder of this section, I want to explore this consequence further, and more specifically, argue that it casts additional doubt on the idea that developmental systems are genuine representation consumers or readers.

At least in clear or paradigm cases of entities that are consumers of representations (perhaps even in all such cases), acts of reading representations serve as actual difference making causes of the output behaviors the representations are posited to

explain with respect to the population of temporal stages realizing the consumer. Reading is a process that is performed by the entity (or more strictly speaking, by a part or parts of the entity), that varies over the course of time, and that serves to confer upon representations that have just been received, or else have been carried forward in time in the entity's memory, the ability to causally influence the entity's behavior in the present moment. An illustration of these features of the process of reading is provided by one of the clearest and most widely-discussed non-human cases of evolved representation consumers, namely forager honey bees (a case Shea himself foregrounds in showing how his infotel theory works). In this case, a forager honey bee returns to the hive and performs one of a range of different 'waggle dances', where each dance carries different correlational information about the location of food. The dance that is actually performed causes other forager bees that read the dance to fly a certain distance and direction from the hive. This interaction, it can be shown, satisfies the infotel conditions, and so the dance tokens are ascribed representational content by that theory. Now, when we ask what kind of cause an act of reading a waggle dance is of the output behavior of an individual forager bee, relative to the population of temporal stages of the bee, it's clear that such an act at least serves as *an* actual difference making cause of that behavior. That is, each consumer bee is made up of a population of temporal stages that exemplify some actual variation in output behavior (i.e., at time t, there is no flying behavior being output, while at time t+*n*, there is some specific flying behavior being output). And this variation between different temporal stages is caused by an actual change in the bee's reading a particular dance telling it that there's food at such-and-such a location (i.e., sometime between t and t+*n*, there's a change in whether the bee is reading a particular

representation). Thus, the act of reading the representation realized by the waggle dance is a process that actually varies over individual time, and in so doing, brings about a change in the output behavior of the individual bee.

Interestingly, I think the same is true in cases where the representations that are received or created by an entity are only used at some later time in the control of overt behavior. For example, we might imagine that, having read a particular waggle dance, a consumer bee doesn't immediately take flight, but instead creates a representation in memory that carries the same information as the dance, and that the bee will carry forward in time. Here the 'output' that is caused by reading the dance is actually an internal, symbolic one (i.e., creating a particular representation in memory). Then, when the bee becomes hungry, it (or again, more strictly a sub-organism part) reads the representation it has stored in memory, and this act of reading causes the bee to fly a certain distance and direction from the hive. What we see is that both acts of reading serve as an actual difference making cause of the outputs the bee performs, relative to the population of temporal stages of the bee. In the first case, the act of reading the waggle dance is an (perhaps even *the*) actual difference making cause of an internal, symbolic process that is conducive to the future success of the bee. In the second, the act of reading the representation it has carried forward in time in memory serves as an actual difference making cause of the bee's flying behavior at that time.

Note that this same point about acts of reading also extends to the low-level consumers Godfrey-Smith makes mention of in the passage presented in the introduction. Take a ribosome inside a cell, for example, which is standardly conceived of as a kind of

consumer of representations realized by mRNA scripts.[23] At time t, the ribosome is not yet engaged in the synthesis of a protein, while at time t+*n* it has just released some newly assembled protein. Thus, there is variation between the different temporal stages making up the ribosome. And this variation between temporal stages is caused by variation between t and t+*n* in whether the ribosome is reading a particular mRNA script. That is, the script is not being read at t (the script has not yet bound to the ribosome, say), but then it is read sometime before t+*n*, and this change in whether the ribosome is reading the script makes an actual difference in the ribosome's output behavior.

Having said all of that, it's important to note that acts of reading in Shea's sense, while not serving as actual difference makers with respect to the population of temporal stages making up the developmental system, may well serve as actual difference makers with respect to a population of *different developmental systems*. So, the claim that acts of reading representations for Shea aren't actual difference makers of whole-organism traits *full stop* would be too strong. In a population made up of different organisms (and so developmental systems) existing at the same time, for example, it may well be that differences in some trait across developmental systems (e.g., hair thickness) is caused by differences in the zygotic DNA each received upon fertilization. For example, assuming there's a gene for thick hair, variation in the "hair thickness output" among the different developmental systems making up the population may well owe to some of the systems receiving a copy of the gene for thick hair and others not receiving a copy. Provided that reading is held to begin at the time of fertilization, we can imagine that differences in hair thickness across the population is indeed caused by differences in the genetic representations that are read by members of the population (some developmental systems

[23] I here assume it is legitimate to refer to an mRNA script as a representation.

read the representation constituted by the zygotic copy of the gene for thick hair, other's don't). The same reasoning also extends to populations made up of different developmental systems existing over time, for example, a lineage of such systems.

But returning to the population of temporal stages making up an individual developmental system: As we watch a single, fertilized egg develop into a mature organism, we can identify numerous stages where more than one developmental trajectory remains open to the system. Then, observing the developmental system develop in one way rather than another, we can ask what it is that made the difference. *What has changed over developmental time so as to prompt the developmental system to behave in this way rather than that?* The point I wish to make here is just that, surprisingly, an act of reading one or another genetic representation *can't* be the (or an) answer to a question of this type. While acts of reading representations may serve as causes of the outputs of a developmental system (whole-organism traits), and may serve as actual difference making causes of such outputs with respect to a population of different developmental systems, it's not the case that acts of reading representations serve as actual difference making causes of such outputs, relative to the population of temporal stages making up a single developmental system. It is in this latter regard that developmental systems appear to be truly *sui generis* as compared with clear or paradigm cases of representation consumers at the very least. I would suggest that this awkward feature of Shea's account is symptomatic of the concept of a representation consumer being extended to an "entity" that it doesn't actually fit or apply to. (I would further suggest that a desire to avoid this strong contrast with paradigm representations consumers may be one reason that Shea's more recent work on the topic has explicitly

claimed that genetic representations are read over the course of developmental time. But as we've seen, at least that part of Shea's account is false.)

One move that might be made at this point to attempt to dissolve the asymmetry between developmental systems and paradigm representation consumers would be to conceptualize developmental systems, not as being associated with single organisms, but rather with whole lineages of organisms.[24] A developmental system would first read the zygotic DNA present at the birth of organisms in the parental generation, which would in turn prompt a suite of whole-organism traits to develop (to be output) in the population. After that, it would then read the zygotic DNA present at the birth of organisms in the first generation of offspring, prompting a suite of (presumably) different whole-organism traits to develop, and so on. On this view, acts of reading representations on the part of a developmental system would turn out to be actual difference making causes of output behavior with respect to the population of temporal stages making up the system. The individual developmental system would thus come to be like the individual bee or ribosome, for example. However, I think it's fairly obvious that this would be a pyrrhic victory at best. It would, for example, commit Shea to the claim that a single developmental system is instantiated by a collection of distinct organisms that are scattered through space and over time. Indeed, it's not clear that there would even be an upper bound on how spatially and temporally far-flung these developmental systems could be. But what's even more worrisome, I think, is that one would apparently have to buy into a kind of lineage selection in order to make the suggestion work (the most contentious type of selection by far). This is because the infotel theory requires that the outputs produced by the consumer mechanism have evolutionary functions (effects which

[24] An anonymous referee for this journal suggested this possibility to me.

have contributed to their being selected in the evolutionary past). Consequently, one would have to claim that these lineage-based developmental systems were subject to a process of natural selection in the past. More generally, though, one wonders (as Godfrey-Smith does even in the one organism-one developmental system case[25]) what motivation there is or even could be for believing in entities of this sort other than to bolster a particular account of genetic representation.

7. Conclusion

The status of genes as bearers of semantic content, as opposed to mere correlational information, has been among the most hotly debated topics in philosophy of biology in recent years. Nicholas Shea's infotel account of genetic representation is arguably the most well-worked out account that attempts to vindicate the idea that genes carry semantic content. For Shea, the contents that are carried by genes tell the "developmental system" associated with the organism what condition the environment is in, (and so) which whole-organism traits to develop. Moreover, at least in his most recent work on the topic, Shea makes clear that he takes these genetic representations to be "read in ontogenetic time, in the course of individual development" (2013, p. 1).

In this paper, I have argued that a close examination of the process of reading in Shea's sense reveals that acts of reading do not actually occur over the course of developmental time at all. Shea likely conceives of the *expression* of non-zygotic genes at later developmental stages as instances of representational contents being read in development, but he is not entitled to do so: for one thing, this involves an equivocation

[25] Godfrey-Smith writes: "Do we have any independent motivation to believe in these things [i.e. developmental systems] at all, other than their enabling us to say something extra and intuitive about the content of the genetic message?" (2007, p. 64).

on Shea's part in the meaning of "reads"; for another, such non-zygotic genes don't actually have representational content at all (just correlational information). I further argued that there is a consequence of genetic representations not being read in development that works to cast additional doubt on the status of developmental systems as genuine representation consumers or readers. At the very least, I hope to have raised several new questions for Shea's account of genetic representation that are clearly deserving of future attention.

References

Carroll S (2006). Endless Forms Most Beautiful: The New Science of Evo Devo, New York: W. W. Norton & Company

Crane T (2008) Causation and determinable properties: on the efficacy of colour, shape, and size. In J. Hohwy and J. Kallestrup (eds), Being Reduced: New Essays on Reduction, Explanation, and Causation, New York, NY: Oxford University Press, pp. 176–95

Funkhouser E (2006) The determinable-determinate relation. Noûs, 40:548–69

Gehring, W (1998): Master Control Genes in Development and Evolution, New Haven: Yale University Press

Godfrey-Smith P (2011) Senders, Receivers, and Genetic Information. Biology and Philosophy, 26:177–81

Godfrey-Smith P (2007) Innateness and Genetic Information. In P. Carruthers, S. Laurence, and S. Stich (eds), The Innate Mind: Foundations and the Future, Oxford, UK: Oxford University Press, pp. 55–68

Griffiths P (2013) Lehrman's Dictum: Information and Explanation in Developmental Biology. Developmental Psychobiology, 55:22–32

Johansson I (2000) Determinables as Universals. The Monist, 83:101–21

Denby D (2001) Determinable nominalism. Philosophical Studies, 102:297–327

Maynard-Smith J (2000) The Concept of Information in Biology. Philosophy of Science, 67:177–94

Millikan R (1984) Language, Thought, and other Biological Categories. Cambridge: MIT Press

Millikan R (1996) Pushmi-pullyu representations. In J. Tomberlin (ed.), Philosophical Perspectives, Atascadero, CA: Ridgeview Publishing, 9, pp. 185–200

Northcott R (2009) Is Actual Difference Making Actually Different? Journal of Philosophy, 106:629–34

Papineau D (1987) Reality and Representation, Oxford: Blackwell

Shannon C (1948) A Mathematical Theory of Communication. Bell System Technical Journal, 27:379–423

Shea N (2007a) Representation in the Genome and in Other Inheritance Systems. Biology and Philosophy, 22:313–31
Shea N (2007b) Consumers Need Information: Supplementing Teleosemantics with an Input Condition. Philosophy and Phenomenological Research, 75:404–35
Shea N (2011) What's transmitted? Inherited information. Biology and Philosophy 26:183–89
Shea N (2013) Inherited Representations are Read in Development. The British Journal for the Philosophy of Science, 64:1–31
Stegmann U (2012) Causal Control and Genetic Causation. Noûs, 00:1–18
Sterelny K, Smith K and Dickson M (1996) The Extended Replicator. Biology and Philosophy, 11:377–403
Waters K (2007) Causes that Make a Difference. Journal of Philosophy, 104:551–79
Wilson J (2012) Fundamental Determinables. Philosophers' Imprint, 12:1–17
Woodward J (1997) Explanation, Invariance, and Intervention. Philosophy of Science, 64:228–77
Woodward J (2003) Making Things Happen: A Theory of Causal Explanation. New York: Oxford University Press
Woodward J (2010) Causation in biology: Stability, specificity, and the choice of levels of explanation. Biology and Philosophy, 25:287–318

Chapter 4

On the nature and explanatory relevance of positional information

The concept of positional information was introduced into developmental biology by Lewis Wolpert in his seminal article, "Positional information and the spatial pattern of cell differentiation" (1969). This concept refers to the information carried by intercellular signals about position in the embryo. Positional information is now taken to explain a wide-range of developmental phenomena. However, everyone seems to agree that embryos can and sometimes do employ other means for controlling the unfolding of development; positional information isn't a "cure all".

Of course, information talk in biology is hardly limited to the concept of positional information. Genes are regularly described as "coding for" proteins and sometimes even for whole-organism traits. An organism's genome is often said to encode a "program" that is executed over the course of individual development. Inheritance systems are likened to "communication channels" for transmitting messages down the generations. The list goes on.

Although information talk in biology has been much discussed by philosophers in recent years, the concept of positional information, and its explanatory role in developmental biology, has received very little attention. This lacuna in the philosophical literature doubtlessly owes to a preoccupation on the part of philosophers with evolutionary and molecular biology, but my guess is that it also reflects the impression that there is nothing inherently problematic in conceiving of cell behavior as coordinated by intercellular information flows. Indeed, I am of that impression, but I nonetheless

think that there remains important philosophical work to do in connection with positional information.

In this paper, I set out to accomplish two main things. First, I simply want to sharpen the concept of positional information. Despite its place of prominence in developmental biology, exactly what is meant by “information", and why it is that intercellular signals can legitimately be said to be associated with it, is often left implicit. My strategy will be to reconstruct the concept within the framework afforded by information theory.

Having suitably clarified the concept, I then address the explanatory role it plays in developmental biology. What we immediately see is that the standard way of thinking about when positional information is explanatorily relevant to cell behavior, and when it is irrelevant, is lacking in motivation. Thus, my second objective is to sketch out a new way of thinking about the explanatory relevance of positional information. I do not, however, claim that my account is the only viable one.

Here is how we shall proceed. I begin by laying out some essential biological details for our discussion (sections 2 and 3). I then turn to the nature of information (section 4). After that, I offer my explicit statement of the sense in which intercellular signals can be said to carry positional information (section 5). I then switch gears and turn to matters concerning explanatory role. I first highlight what I think is wrong with the standard thinking in this area (section 6), and then offer my alternative account (section 7).

4.1 Embryonic geography

When biologists talk of positional information, it is virtually always in connection with *bilaterian* organisms (i.e., organisms such that there exists just one plane to make a cut through the organism and wind up with two [roughly] mirror-image parts.) Accordingly, my focus in the present work will be limited to such organisms, though I do think that the main ideas I develop (with suitable modifications) can be extended to other types of organisms.[26]

Bilaterians have a distinct head and tail region, as well as a top (back) and bottom (belly) region. In the jargon of developmental biology, these regions are called the "anterior" and "posterior" regions, and the "dorsal" and "ventral" regions, respectively. Of course, these animals aren't limited to two spatial dimensions only, but also extend outward from their centers. This gives rise to a third spatial dimension of animal form, one which developmental biologists call the "distal-proximal" dimension. An animal's limbs, for example, are more distally located than its heart.

Biltaerian embryos also have anterior/posterior, ventral/dorsal, and distal/proximal regions. Roughly speaking, an embryonic region is designated as anterior, say, because the cells making up that region will, assuming normal development, come to make up the anterior part of the adult (likely accompanied by many offspring cells). There is considerable variation among animal groups in when the relative orientation of these regions in the embryo comes to reflect the relative orientation of these regions in the adult, but that's a complication I'll set aside in the present paper. No matter what species we're dealing with, there will be an earlyish developmental stage at which the layout of the embryo spatially corresponds to the layout of the adult.

---

[26] The reader should assume that any bear instance of "organism", "animal", "embryo", etc., in what follows means *bilaterian* organism/animal/embryo, unless otherwise specified.

Having reached such a stage, one can envision the embryo as having a kind of biologically meaningful "geography" (Carroll, 2006). To flesh out this idea, we map the various biological terms onto geographic ones in the following way: north/south goes with dorsal/ventral, east/west goes with posterior/anterior, and finally, proximal/distal goes with greater depth/lesser depth (or alternatively, with lesser/greater altitude). A north-south (NS) axis exists which connects the dorsal and ventral regions of the embryo, as does an east-west (EW) axis, connecting the posterior and anterior regions. The point at which the NS and EW axes intersect defines the embryo's center or origin. In addition, various lines of latitude and longitude exist which intersect the NS and EW axes of the embryo, respectively. Finally, depth can be conceived of as distance from the embryo's center. Below, I will mostly follow this geographical convention for ease of exposition.

These poles, latitudes, longitudes, and distances from the embryo's center together comprise a *global spatial coordinate system*. This is the coordinate system structuring the main body of the animal. However, the same reasoning can be extended to bodily parts that protrude outwards from the main body. These parts will have their own spatial coordinate systems. An arm, for example, will have its own top (i.e., dorsal or northern region), its own back (i.e., posterior or eastern region), and so forth. In turn, an arm or a leg will typically have parts that protrude outwards from it (e.g., digits), giving rise to yet another coordinate system. In short: embryos are associated with multiple coordinate systems existing at a variety of different scales (main body, bodily part, part of a bodily part, etc.).

4.2 Signaling molecules

The genome of a typical animal encodes a large and diverse array of *signaling molecules*.[27] Whether a signaling molecule is able to bind to some receptor depends on the biochemical properties of both entities. For present purposes, we will mostly be interested in signaling molecules that show a high-degree of specificity for receptors in the sense that the molecule is able to bind to only one or a small number of receptors.

Some signaling molecules are regularly deployed only in certain bodily regions (e.g., in the liver), while others are made use of in many different parts of the body. But even as regards signaling molecules of the latter sort, the *combinations* they form with other molecules so as to prompt cell behavior may well be specific to a limited number of spatial regions. The use of combinations of signaling molecules is an important developmental theme. Another important theme is this: often times, the same type of signaling molecule will form a *gradient* in virtue of diffusing from one area of the embryo to another. So, even if the molecule is present throughout some large region, a certain *concentration* of that molecule may well be specific to a limited area.

---

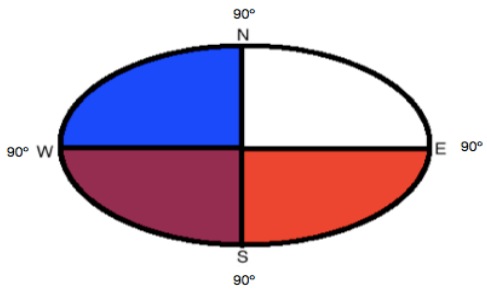
[27] What is a signaling molecule, exactly? As it turns out, this question is deceptively simple to answer. One thing we might say is that a signaling molecule just is any molecule which is capable of binding to a cellular *receptor*. However, this is only progress if we have some account of how receptors differ from other sorts of molecular entities. If we turn around and define a receptor as any molecular entity which is capable of binging a *signal*, then we will have drawn a very tight circle around ourselves indeed. (I will simply mention that it is not uncommon to see biologists define 'signal' and 'receptor' in such a circular fashion.)

A more promising approach is to define a signaling molecule as a molecule which has been naturally selected for *transmitting information* to cells. (Of course, such a definition should not be accepted until we have said what we mean by 'information', but as I explain below, I think there is a perfectly straightforward way of understanding talk of information in the present context.) Admittedly, this is more like a *schema* for an account of what signaling molecules are, than it is a definite view, but examining more specific formulations of the idea at this point would take us too far afield. To provide just a bit more detail, though: the guiding idea is that signaling molecules are distinctive in virtue of having been retained in the evolutionary past, at least initially, because they enabled receiver cells to detect the presence of behaviorally relevant states of the world. Such molecules may have had other functions too, but transmitting behaviorally relevant information was at least part of the reason why they were selected for.

In any case, I propose to set aside this morass of issues in the present work, and assume—along with the biologists—there is a real distinction to be drawn between those molecules which play a signaling role in the organism, and those that play quite another role (e.g., providing food to cells).

We can think of individual types of signaling molecules, concentrations of such molecules, and also combinations of such molecules (where part of the combination may well be the concentration of some type of signaling molecule) as being associated with a certain set of regions within the spatial coordinate systems structuring the embryo. To give just a toy example here: suppose that one type of signaling molecule S1 is present in the whole of the southern hemisphere, and only in the southern hemisphere, i.e., S1 is present from 90° W to 90° E and from 0° S to 90° S. At the same time, suppose another type of signaling molecule S2 is present in the whole of the western hemisphere, and only in the western hemisphere, i.e., S2 is present from 90° W to 0° W and from 90° N to 90° S. These two signaling molecules would thus overlap in the south-western quadrant of the main body. (See Figure 1 below.)

To get a sense of where we're heading, note that upon receiving a signal of type S1&S2, for example, a cell would be able to "infer" that it is located in the south-western quadrant of the main body. Crucially, this is not because S1 and S2 co-occur in the south-western quadrant of the *particular* embryo which the receiver cell belongs to, but rather because S1 and S2 *generally* co-occur in this region in normal embryos of the species.[28]



[28] In this paper, I will re-use the terms like "S", "S1" and "S2", and so on to stand for entirely different entities. One should not assume that such a variable refers to an entity from an earlier example unless explicitly noted in the text.

Figure 4.1. The distribution of signaling molecules S1 and S2 in the embryo. S1: red, S2: blue, S1&S2: purple.

4.3 Information

With these biological details on the table, let us now turn to the nature of information. Here I will borrow from some recent work by Brian Skyrms (2010). In many ways, Skyrms' treatment of information closely follows Claude Shannon's, the father of information theory (Shannon 1948). The main difference is that Skyrms offers an explicit account, not just of *information quantity*, but also of *information content* or *meaning*. For Skyrms (as for Shannon), the quantity of information carried by an event *e* about a state *w* of a source *W* is given by the extent to which *e* moves the probability of *w* off of its unconditional probability, measured in bits (although another measurement could just as easily be used). So, if *w*'s unconditional probability is 0.5, say, and *e*'s obtaining moves the probability of *w* to 1, then the probability of *w* has been moved by a factor of 2 (0.5*2 = 1), yielding $\log_2(2) = 1$ bit of information. We might also be interested in the amount of information *e* carries about source *W* as a whole. The most common measure of this quantity is known as the *Kullback-Leibler distance*, and is gotten by taking a weighted sum of the information carried by the event about each source state (i.e., $\sum_i \Pr(w_i \mid e_i)*\log_2[\Pr(w_i \mid e_i)/\Pr(e_i)]$).

What about the *information content* of an event? Skyrms takes this to be given by a vector listing how the probability of each source state has been moved by the event's obtaining. So, if we're dealing with a source that has just two equiprobable states, w_1 and w_2, and if *e*'s obtaining moves the probability of state w_1 to 0 and of state w_2 to 1, then the

information content vector would look like this: $<-\infty, 1>$[29]. Among other things, this vector tells us that the state of the source is w_2, or equivalently, that it's *not* w_1. This is the simplest possible case. A more complicated case would be one involving four equiprobable states, w_1-w_4, and an event that moves the probability of the first two states, say, to 0 and leaves the other two with a probability of 0.5. Then we would have a information content vector that looked like this: $<-\infty, -\infty, 1, 1>$. Here the event would be telling us that true state of the source is contained in the set $\{w_3, w_4\}$, or equivalently, that it's *not* in the set $\{w_1, w_2\}$.

In both cases, the event's obtaining has reduced the probabilities of some source states to 0. Skyrms describes such cases as ones in which the event has *propositional content*. However, sometimes a event may fail to move the probability of even a single source state to 0 (such an event will not have propositional content). And even when some source state *has* been reduced to 0, the important thing from the perspective of the receiver(s) involved may be how much the probabilities of the "live options" have changed, or simply what the new probabilities of those states are. At any rate, propositional content is a special case of something more general for Skyrms.

In short: for a given event e, we can ask both how much information e carries (either about a particular source state w, or about the source W as a whole), as well as what the information content of e is.

4.4 Positional information

To flesh out the sense in which intercellular signals can be said to carry information about position in the embryo, let is return to our toy example from section 2.

[29] The negative infinity here results from taking the binary logarithm of 0.

Each cell in the embryo can be thought of as having associated with it a source whose possible states are the positions making up the spatial coordinate system of the main body. Let us call this source $Position_{main_body}$. We assume that a given cell has an equal chance of being located at any one of these positions, and so all of the states of $Position_{main_body}$ have the same unconditional probability.

Now, some signal received by a cell, which, you will recall, might take the form of an individual signaling molecule, a concentration of some molecule, and/or a combination of such molecules, will carry information about position in the main body just in case the signal serves to move the probabilities of some of the states of $Position_{main_body}$. And the signal will carry information about some *particular* position in the main body just in case it moves the probability of that position's corresponding source state. Thus, a signal of the form S1&S2 in our example from section 2 carries positional information that might be glossed as "located in the south-western quadrant" because it moves the probability of every position *not* contained in the south-western quadrant to 0.

When a cell belongs to an embryo that is structured by multiple spatial coordinate systems, it will have multiple sources associated with it. Each such source will have as its possible states the positions of the corresponding coordinate system, and all will have equal unconditional probabilities. Thus, the information a signal carries about position in the *right forelimb*, say, is measured in the same way as it is for the main body.

Whether and to what extent a signal carries positional information, then, simply boils down to how that signal changes the probabilities of various positional states of the embryo. The *quantity* of information associated with a signal, as well as the information

content of the signal, will depend on how much probabilities of this sort have been moved by the signal.[30]

4.5 On the explanatory relevance of positional information

Having sharpened the concept of positional information, I now want to consider the explanatory role this concept plays in developmental biology.

It seems to me that the primary explanatory work done by positional information in developmental biology is to offer an explanation for *why* cells behave as they do (in particular, why it is that they differentiate, or adopt the specific cell fates, that they do), and by extension, *why* certain embryonic regions come to instantiate the specific patterns they do over the course of developmental time. Observing some initially homogenous population of cells change so as to come to realize a heterogeneous pattern, we want to know what it is that has prompted this change. (Sometimes, the pattern in question might consist in a subset of cells performing some type of behavior, while the others "do nothing".) In a wide range of cases, the preferred answer to this kind of question is that the cells received messages informing them of their whereabouts within the relevant region (positional information), and this new information was then exploited by the cells to develop in ways that "match" their positions. However, as noted in the introduction, everyone seems to agree that embryos can and sometimes do employ non-positional information based means for patterning the embryo, means that nevertheless still draw heavily upon intercellular signaling. So, what is it exactly that distinguishes cases where positional information is explanatorily relevant from those where it is not? When is positional information *the answer*, or at least part of the answer, and when isn't?

[30] See Calcott (forthcoming) for a similar analysis of positional information.

Before we go on to offer our own take on this question, let us first consider what some developmental biologists have had to say. Below is a passage from Sean Carroll, a leading developmental biologist who frequently draws upon the concept of positional information in his work. Carroll writes:

> One of the most striking truths about animal body patterns is their regularity at all scales, from their overall body plan to the fine details of an individual structure or body part. The tiling scales on a butterfly wing and the regular spacing of feathers on a bird are good examples of the latter. While the location of cells can be specified fairly precisely, this is not the only means of achieving regularity of patterns. The spacing of many individual elements in a larger array is often accomplished by a process dubbed lateral inhibition. …
>
> The net effect of lateral inhibition is a regular pattern—of hairs on insect bodies; feather, scale, and fur patterns on birds, reptiles and mammals; and the beautifully compound eyes of the arthropod. All of these patterns are generated locally by cell interactions, not specified by global coordinates (2006, p. 104-5, italics mine).

And here is a passage from Wolpert himself:

> It is clear that some patterns do not depend on a mechanism involving positional information. If, for example, each cell in a developing system has a unique identity this does not mean that they have positional information. Thus in the early development of the ommatidium in the *Drosophila* eye, each of the eight cells has a unique identity, but this depends on the cell's position in relation to each of the other cells. Moreover, each of these identities is established sequentially during development. Positional information is about graded properties … with its implications for a coordinate system. (1994, p. 4)

I am inclined to think that both sets of remarks are misguided. Carroll, for example, appears to think that if some developmental process involves *local interactions* among cells, then positional information is not in play. (Exactly how local is too local anyway?) And Wolpert, who would presumably agree with Carroll on this point, apparently sees a *sequential interaction* among cells as precluding an explanation citing positional information. Surely, there are plenty of paradigm cases of information

transmission and processing outside of biology that involve both local and sequential interactions between senders and receivers, so it is hard to see why these features *per se* would imply that positional information is not explanatorily relevant. But perhaps more importantly, given the details about how signaling molecules come to be associated with information about position, it strikes me as plain arbitrary to draw the line in the way Carroll and Wolpert suggest.

To bring this point out, let us consider a simple hypothetical case. Suppose that in the embryos of some species, region R has to develop into alternating rows of light and dark cells. More specifically, suppose that the southern-most row (row 1) must come to consist solely in light cells, the next row to the north (row 2) solely in dark cells, and so on.

Now suppose that cell differentiation in this region is under the control of two signaling molecules, S1 and S2. S1 activates gene G1 which encodes a protein prompting differentiation into a light cell. S2 has a more complex role. First, it represses G1. Second, it represses the gene coding for it, G2. (That is, G2 codes for S2, and S2 represses G2.) Third, S2 activates a gene, G3, which encodes a protein prompting differentiation into a dark cell. Crucially, S2 does not enter the nucleus of the cell synthesizing it and it only enters (or binds to) those cells to the immediate north. Finally, we suppose that S1 is gradually introduced into region R from the southern-most area so that the cells in row 1 encounter S1 first, followed by the cells in row 2, and so on.

Assuming that development proceeds normally, this setup will give rise to the pattern shown in figure 2. Is it a case in which positional information is explanatorily irrelevant? I claim that it need not be. If, for example, S1 is unique to region R, then

receiving a signal containing S1 will at a minimum tell the cell that it is located in R. In addition, note that S2 is only every present in even-numbered rows of S2. (S2 might be used elsewhere in the embryo; but as regards R, it's limited to even-numbered rows.) Consequently, receiving a signal of the form S1&S2 will move the probabilities of all positional states to 0 except for those belonging to even-numbered rows of R.

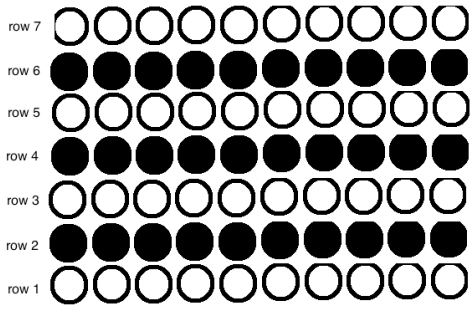


Figure 4.2. Region R must come to contain alternating rows of light and dark cells, with light cells making up odd-numbered rows, and dark cells making up even-numbered rows.

This is a case involving both local and sequential interactions among cells. Yet, the signaling molecules clearly do possess positional information. Moreover, this information certainly seems as though it could be relevant to explaining why the cells in R do what they do, and by extension, why region R comes to assume the banded pattern that it does. Why would we want to rule out such a possibility *a priori* (as Carroll and Wolpert would apparently have us do)?

Let us go back to basics and see if we cannot offer a more principled way of thinking about the explanatory relevance of positional information in development.

4.6 Towards an improved account of the explanatory relevance of positional information

The behavior of a cell in the present moment can be thought of as being determined by two kinds of causes, namely, *intrinsic* and *extrinsic*. So far, we have been concerned only with extrinsic causes, in particular, intercellular signals. Intrinsic causes, by contrast, concern such things as, e.g., the set of activated transcription factor proteins contained within the cell's nucleus, epigenetic modifications that have been made to the cell's DNA, the existence of particular signaling pathways in the cell, and so on. This set of intrinsic causes will in general reflect the cell's *developmental history*, that is, the cell's ancestry, together with the signals and/or other stimuli which the cell has encountered so far over the course of development.

We can think of the set of intrinsic causes of cell behavior at a given time as determining the *developmental state* of the cell at that time. By "developmental state", I mean a (multiply-realizable) structural state of the cell that determines the input-output character of the cell as regards the various signals it may encounter. Thus, the output prompted by a type of signal may change from moment to moment as the cell transitions from one developmental state to another. A given developmental history will correspond to exactly one developmental state, though multiple developmental histories can presumably map to the same developmental state.

Now, we can note that developmental states will in general be associated with information about position in just the way that signals are. This is because there are bound to be population-level facts about the correlation between being in a particular type of developmental state and being at a particular position, or set of positions, in the

embryo. In this way, a given developmental state may considerably move the probabilities of certain positional states. To give an extremely simple example: if, whenever a cell is disposed to perform behavior B in response to signal S, the cell is located in the future liver, then *given* that the cell now instantiates this disposition (or better: given that it is in a developmental state that prescribes such input-output behavior), the probability that the cell is located in the future liver is 1.

In ascribing information, not just to signals, but also to developmental states, I am here taking my lead from computer scientists, as well as cognitive scientists more generally. Computing systems are often described as having two forms of knowledge: symbolic and procedural. Symbolic knowledge is information which is explicitly represented for the system, information which can be stored, transported, and operated upon. It is information which is physically *read*, at least potentially, by some part of the system. In contrast, procedural knowledge is a matter of *implicit representation*, and one important way a computing system can be said to possess knowledge of this sort is by having reached a particular *processing state*, where a processing state is a structural state of the system determining its input-output profile at the present moment. For example, a (binary) adding machine might transition into a certain processing state if only if the last input it received consisted of two '1's. If so, then in virtue of being in that processing state, the system can be said to know that the last input consisted in two 1s. (When the procedurally-known information concerns the past sequence of inputs to the system, it is described as *state memory*.)[31]

As will by now be obvious, our notion of a developmental state has been directly modeled after the processing state of a computing system. It is the "wetware" equivalent,

[31] See Gallistel and King (2009) (particularly, p. 100-1) for an excellent discussion of these ideas.

so to speak. But this means that the information associated with a cell's developmental state is to be understood as being possessed in an implicit or state-based form, rather than in a symbolically-encoded form.

With the concept of a developmental state on the table, we can recruit a second distinction into service, one dating back to Shannon's original work on information. The distinction is between the information that is *carried* by a signal, and the information that the signal *communicates*. The information that is carried by a signal simply depends on how that signal moves the objective probabilities of some source. The information that the signal communicates, however, also depends on what the receiver can already be said to know about the state of the source, that is, what the receiver's prior knowledge is.

I claim that the prior knowledge of a cell about its position is constituted by the cell's developmental state. More precisely, the developmental state of the cell will imply a probability distribution over the set of positional states belonging to one or more spatial coordinate systems of the embryo. As signals come in, they can move these probabilities to a greater or lesser extent, or in the extreme, not at all. The information that the signal succeeds in communicating to the cell depends on how these probabilities are moved. (It is *Bayes' Rule* which tells us exactly how these probabilities are moved, and so what the posterior distribution is.)

Finally, it would appear that we need some notion of utility or interests in order to make sense of the idea that cell behavior can be *explained* by positional information, or put differently, that cells can *use* such information in the control of behavior. The main reason is that we want to be able to distinguish between behaviorally relevant and irrelevant information from the perspective of the cell. This can seem to pose a thorny

issue for the following reason. As far as non-cognitive biological entities go, talk of interests is almost always a short-hand for fitness. That is, the interests of such an entity is understood in terms of the entity's expected survival and reproduction. Yet, the vast majority of cells in a multicellular organism are somatic or non-germ-line cells, and consequently, represent evolutionary "dead ends". So, in what sense can such cells have interests that might be differentially promoted by behavior?

This issue can be side-stepped by appealing to the notion of the inclusive fitness of somatic cells. Inclusive fitness takes into account, not just one's own survival and reproduction, but also that of genetically related individuals (Hamilton, 1964). Since under normal developmental conditions, the degree of relatedness between somatic and germ-line cells will be 1, it follows that the former will be indifferent between having *n* offspring of their own and helping germ-line cells to have *n* offspring. In a word, then, somatic cells promote their interests by promoting the interests of germ-line kin, and typically the most efficient way to do this will simply be to perform behaviors in space and time that result in the construction of a fit organism.[32]

Accordingly, there is no obstacle to saying that the different behaviors in a cell's repertoire might be associated with different expected utilities for the cell. How would this work? Each behavior would have a certain payoff for the cell at each position in the body which the cell might be located. The expected utility of some behavior is the weighted sum of the utility of that behavior in each possible position, where the weights are the probabilities implied by the cell's developmental state. It then makes perfect sense for us to ask how an incoming signal changes these expected utilities in virtue of the

[32] Planer (2014) discusses this idea at greater length, though in relation to somatic genes, not cells.

positional information it communicates (which will depend on how the signal moves the probabilities of positions, the weights).

Of interest to us will be whether the signal increases the expected utility of the behavior it prompts. Or more precisely put, we will want to know whether the information communicated by the signal serves to increase the expected utility of the behavior. Setting aside bizarre cases, this will be true simply when the signal increases the probabilities of positions in which the behavior the signal prompts is adaptive for the cell (and so for the organism, presumably), *relative* to the probabilities implied by the cell's current developmental state.

By pulling these threads together, we can offer a new way of thinking about the explanatory relevance of positional information in development. I suggest that cell C's performing behavior B in response to signal S is explained by positional information if and only if:

(a) S communicates positional information to C (i.e., S moves at least some of probabilities assigned to positional states by C's current developmental state); and

(b) this information increases the expected utility of B for C (i.e., the way in which S moves these probabilities leads to an increase in the expected utility of B for C).[33,34]

---

[33] Actually, whenever a signal increases the expected utility of the behavior it prompts, it follows that the signal will communicate some information to the cell, so having two conditions here is technically redundant. However, I find it is useful on conceptual grounds to separate them out.

[34] There is at least a partial analogy here with the way Dretske (1988) was thinking. Put roughly, Dretske argued that informational-explanation of behavior gains traction when a type of internal state S has been *recruited* as a cause of some type of behavior B because S *indicates* the presence of some other state F (one which, when the behavior was performed in it in the past, led to the retention of the causal link between S and B). The present account differs from Dretske's in a few key respects, but it certainly has a Dretskean flavor to it.

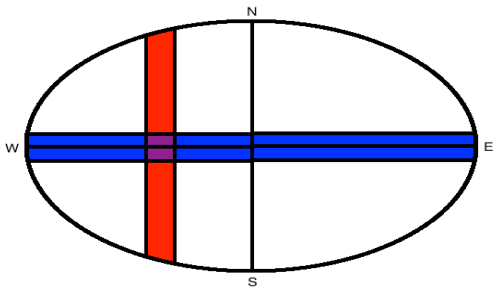
With an answer to this question in hand, we can then go on to say what it is for a whole developmental pattern—which I take to be an aggregate of cell-level behavior—to be explained by positional information. Such a pattern is so explained to the extent that the cell-level behavior realizing the pattern is itself explained by positional information. Thus, on my view, it is the behavior of individual cells that first and foremost receives an explanation in terms of positional information; whole patterns are explained in this way only indirectly.

Let us illustrate how this account works with an interesting, but still relatively simple, hypothetical case. Suppose that in the embryos of some species, the forelimbs have to be centered at 30° W and 0° N, and come to occupy a region having a 10° width and a 10° length. (We here ignore the distal-proximal dimension for simplicity.) Suppose further that, at the start of the process of forelimb placement, all of the cells located between 40° W and 20° W and 10° N and 10° S are in a developmental state D, where D is equally compatible with being located anywhere within this region, and incompatible with being located anywhere outside of this region (i.e., D assigns equal probability to each of the positional state between 40° W and 20° W and 10° N and 10° S and 0 to all other positional states).

A subset of the cells within this region then receives a signal composed of two signaling molecules, S1 and S2, which prompts differentiation into a forelimb cell by causing a certain gene G to be expressed. (We can assume that cells within this region that *don't* receive a signal of this type instead eventually assume a different cell fate, say, "scapula cells".) More specifically, let us imagine that S1 is present in a vertical band which stretches from 35° W to 25° W, and which just about reaches 90° N (at its most

northerly point) and 90° S (at its most southerly point). And in addition, let us imagine that S2 is present in a horizontal band stretching from 90° W to 90° E and from 5° N to 5° S. (The distribution of these signaling molecules is shown in Figure 3 below.)

Now, it's clear that tokens of the signal type (S1&S2) communicate information about position in the main body to the cells that receive them (condition (a)). Once a signal of this type has been received, a whole chunk of positional states that were previously "live options" from the perspective of the receiver cells now have probability 0, as only those states corresponding to a position between 30° W and 25° W and 5° N and 5° S now have positive probability. And finally, we assume that differentiating into a forelimb cell is the optimal cell behavior, given that the cell is located between 30° W and 25° W and 5° N and 5° S (condition (b)). Thus, the information about position that is communicated to a receiver cell by a signal of the type S1&S2 serves to increase the expected utility of the behavior the signal prompts from the cell. It is for this reason that I suggest that the behavior of these cells, and so the pattern as a whole, is explained by positional information.[35]



[35] It will by now be clear that the cell behavior involved in patterning region R in our example from section 6 could easily turn out to be explained by positional information, given certain assumptions about the developmental states of these cells.

Figure 4.3. The distribution of signaling molecules S1 and S2 in the embryo; S1: red, S2: blue, S1&S2: purple. Purple marks the spot where the forelimb must develop. (Adapted from a drawing in Carroll 2006.)

Can we give a case in which cell behavior clearly isn't explained by the communication of positional information? Indeed, we can. Here is a very simple case. Suppose that in the embryos of some species, a population of cells that will go on to form a patch of skin on the animal's belly must differentiate into two differently colored cell-types, black and yellow. Suppose that this population presently occupies a region spanning from 5° W to 5° E and from 85° S to 90° S of the main body. (Again, we are ignoring the proximal-distal dimension for simplicity). Cells occupying the western half of this region must develop into black cells, while those in the eastern half must develop into yellow cells.

Now suppose that, at the start of the process of coloration of this region, the cells in the western half are in a developmental state D1 which is equally compatible with being located anywhere within the western half of this region, and incompatible with being located at any other position. In contrast, cells within the eastern half of this region are in a different developmental state, D2, which is equally compatible with being located anywhere within the eastern half of the region, and incompatible with being located at any other position. Put intuitively, then, all of the cells in question already know that there are in this ventral patch, and moreover, which half of this region they are in, but that's it.

Finally, suppose a single type signaling molecule, S1, is now introduced into this region, and comes to be present at every position with in it. If a cell is in developmental

state D1, and it receives S1, it differentiates into a black cell, whereas if a cell is in D2, and it receives S2, it differentiates into a yellow cell. The mechanistic difference underlying this difference in output here as regards an input of type S1 might consist in nothing more than cells in D1 possessing a signaling pathway linked to an activator of a gene G1 that prompts differentiation into a black cell, while those in D2 possess a different signaling pathway, one linked to an activator of a gene G2 that prompts differentiation into a yellow cell. Assuming development proceeds normally, this region will come to take on the pattern shown in Figure 4.

According to the present account, this pattern is *not* explained by positional information. The reason is *that signals of type S1 fail to communicate any information to receiver cells about position*. At most, S1 would be capable of telling these cells that they are somewhere between 5° W to 5° E, and from 85° S to 90° S, but this is something the cells already know. Indeed, they know something more specific than this, namely, which half of the region they are located in. Consequently, signals of this type do not serve to move the probabilities of positional states, relative to either developmental state D1 or D2. To be sure, signals of this type might *carry* a great deal of information about position (for example, S1 might only be used within this small patch of ventral skin), but they do not succeed in *communicating* anything about position to our focal group of cells.

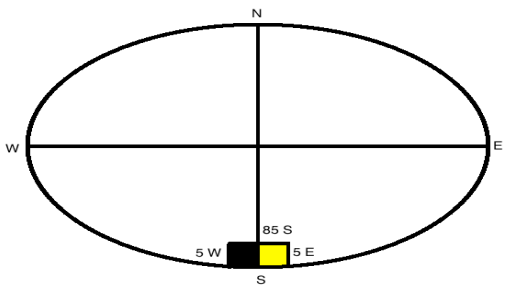


Fig. 4.4. Distribution of black and yellow cells within a patch of ventral skin. The relevant regions have been blown-up for visibility.

Here is another (more complicated) negative case. Suppose that the (left) eye in the embryos of some species must be centered at 80° W and 20° N, and have width and length of 4°. So, the total area occupied by the eye will reach from 82° W to 78° W and from 22° N to 18° N. This region must come to be populated by two types of cells, rods and cones, where each cone cell is surrounded by a group of rod cells. Unlike like the foregoing case, at the start of the process of eye formation, all of the cells within this region are in the same type of developmental state D. D assigns equal probability to every position within this region, and 0 to every other positional state. Put intuitively, the cells already know they are located in the future eye, but have no further "idea" as to where they are in within this region.

Now suppose that differentiation in this region is under the control of two signaling molecules, S1 and S2. S1 activates a gene prompting differentiation into a cone cell, G1. S2, by contrast, represses this same gene. S2 also represses the gene coding for S2 (G2), and activates a gene prompting differentiation into a rod cell, G3.

Suppose that S1 rapidly comes to be present at every position within this region, but that it always enters (or binds to) some particular group of cells in the region first. Crucially, exactly *which* group of cells S1 enters first varies from embryo to embryo—in fact, it's a perfectly random affair. In *this* embryo, it is *this* group of cells within the eye that receives S1 first; in *that* embryo, it is *that* group of cells within the eye that receives S1 first, etc. If some cell is among the first to receive S1, it will respond by expressing G1, thereby differentiating into a cone cell. At the same time, such a cell will synthesize

S2, which (we stipulate) the cell then passes on to *all and only* its nearest neighbor cells. (That's the only way S2 gets introduced into this region.) It follows that a neighbor of a cell that was among the first to receive S1 will receive *both* S1 and S2, in which case it will express *neither* G1 nor G2. Such a cell will, however, be caused to express G3, thereby differentiating into a rod cell. (So, just to translate all of this back in to talk of developmental states, we can say that if a cell is in D, and it receives S1, it differentiates into a cone cell, and sends signal S2 to all of its nearest neighbors; but if it instead receives S1&S2, it becomes a rod cell and sends nothing.) Assuming normal development, the eye will come to take on a "sprinkle" pattern, such as the one shown in Figure 5 below.

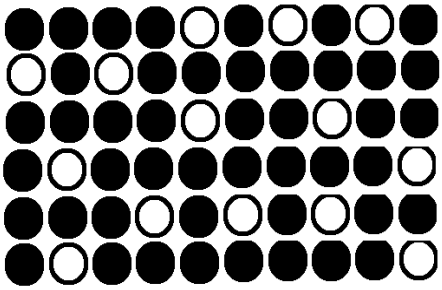


Figure 4.5. A possible pattern of rods and cones in the eye.

The present account again returns the result that this pattern is *not* explained by positional information. Why not? Recall that at the start of the process of eye formation, each cell already knew that it was somewhere within the eye, in virtue of being in developmental state D.[36] Since S1 comes to be present at every position within this

[36] If this assumption seems too much like "magic" for the reader: imagine that state D is partially constituted by a cell's containing an activated transcription factor that is necessary for the expression of both G1—the "cone cell gene"—and G3—the "rod cell gene". Then, given that this factor is only ever present in cells located where the eye will form (in which case the only cells that ever reach state D are

region, it follows that S1 could at most tell a receiver cell that it is located somewhere in this region. In contrast, signals of the type S1&S2 come to be present only at some subset of positions within this region, but exactly *which* set of positions this is will dramatically differ from one embryo to another. It is a randomly chosen set when viewed from the perspective of the population (which is the right perspective to take when it comes to understanding positional information, I claim). But this means that a cell is neither more nor less likely to be at any particular position, or set of positions, within the region, given that it receives a token of S1&S2. So, it is also true that signals of type S1&S2 could at most tell receiver cells that they are somewhere within the future eye. In either case then, the signal that is received will fail to move the probabilities of positional states within this region, relative to those that are implied by developmental state D. Once again, such signals might *carry* a great deal of information about position, but they do not serve to communicate positional information to the cells involved. *No cell within this region is being told anything about its position that it didn't already know*.

Both of the negative cases just given have involved a failure of cell-level behavior to satisfy our first condition (condition (a)) from above. But what about condition (b)? What work does it do?

Basically, condition (b) is required to rule out cases of the following sort. A signal prompting some type of behavior from a cell may communicate information about position to the cell, but the communicated information might be *behaviorally irrelevant* from the perspective of the cell. The most straightforward way for this to happen is for a signal to merely "take away" probability from one positional state and "add it" to

---

ones located at this region), it follows that cells which are in D can be said to know that they are where the eye will form.

another, where the behavior has the same utility for the cell in both states. Then the expected utility of the act will be exactly what it was *prior* to the cell's receiving the signal. For this reason, it strikes me as in some sense defective to take the positional information that is being communicated to the cell to *explain* why the cell performed the behavior it did in response to the signal. On the other hand, if the behavior came to have different utilities for the cell in these different positional states at some future time (say, as the environment changed, or as the behavior of other embryonic cells changed), then the cell's performing this behavior in response to the signal might well come to be explained by positional information. (The converse of this point is that cell behavior and so whole patterns that were once explained by positional information might cease to be so-explained over evolutionary time.)

Time to sum up. We have provided a sketch of an alternative way of thinking about the explanatory relevance of positional information in development. On the present account, what matters isn't the positional information that is carried by signals prompting cell behavior *per se*, but rather the positional information such signals *succeed in communicating* to receiver cells, where this takes into account what the receiver cells can already be said to know (albeit, in an implicit sense) about their position in one or another spatial coordinate system structuring the embryo. When the information about position that is communicated to the cell by a signal serves to increase the expected utility of the behavior that the signal prompts, then I claim that the cell behavior is properly seen as being explained by positional information. Failing this, we should search for another type of explanation, perhaps one involving a different kind of information (e.g., information about what one's neighboring cells are up to, information about what temporal stage

some developmental process has reached, etc.), or one not involving information at all (i.e., an explanation dealing solely in biochemical terms). (As regards questions about the explanatory relevance of other kinds of information to cell behavior, I would propose that we approach such questions in roughly the same way that we have for positional information in this paper. That is, we should ask what the relevant sources are, how incoming signals interact with the prior knowledge of receiver cells to determine what is communicated, and how the communicated information bears upon the expected utility of behavior.)

While the present account certainly does allow for there to be instances of cell-behavior, and indeed, whole developmental processes, to which positional information is explanatorily irrelevant, rather than thinking in all-or-none terms, it may ultimately prove more useful to view cases as falling along a spectrum here. Cases where no (behaviorally relevant) positional information is communicated by the signal prompting cell behavior merely constitutes one extreme, one end of this spectrum. At the other end of the spectrum are cases where the signal serves to communicate a very large quantity of positional information. Cases of this latter sort will inevitably involve receiver cells whose developmental states leave open a wide-range of more or less equiprobable positional states, and signals that carry a considerable amount of positional information. In this way, the signal can serve to drastically cut down on the set of "live options", leaving just a small number of positional states with positive probability. It seems to me that the instances of pattern formation which developmental biologists most readily cite positional information in connection with are ones that match this description. However,

the vast majority of cell behavior in development will presumably fall somewhere in between these two extremes.

On a more general note, quite apart from questions about whether and to what extent positional information is explanatorily relevant in a given case, I think the distinction between the information that is carried by signals, and that which is communicated is crucial. By attending to this distinction, we see how signals of the very same type can nevertheless communicate dramatically different amounts of positional information (or other kinds of information for that matter) to receiver cells. A token of signal type S might serve to merely "choose" between a small number of possible positional states (and so communicate a small amount of information), relative to a cell that is in developmental state D1, for example. Yet a token of the very same type, received by a cell that is in a different developmental state D2, might effect a huge reduction in uncertainty. This is so despite the fact that the two tokens are bound to *carry* the very same quantity of information about position. And even when two tokens of the same type communicate the same or roughly the same amount of information to receiver cells, there may well be crucial differences between them as regards the informational content they communicate. Two tokens, both of type S, might rule out a comparable number of live options, relative to two different developmental states D1 and D2, but the positional states that each token rules out may be totally different. Thus, tokens of the same type of signal can function to tell each cell something very different, despite the fact that each token will *carry* the same informational content. Bearing these facts in mind may prove very useful in seeking to understand why signaling molecules are "unfurled"

during development in the particular way that they are in a given species, though this is a topic I leave for another day.

4.7 Conclusion

The concept of positional information figures among the very core concepts of modern developmental biology. Yet it is often unclear exactly what developmental biologists mean by "information" in this context, and why intercellular signals are legitimately seen as being associated with it. In this paper, I have attempted to sharpen the concept of positional information by reconstructing it within the framework afforded by information theory. Doing so not only served to confer rigor upon the concept, but also helped to highlight some muddled thinking about the explanatory relevance of positional information on the part of developmental biologists. I have also attempted to provide an improved way of thinking about the explanatory relevance of positional information, though I remain open to the possibility that there are other fruitful ways of approaching this issue.

References

Calcott B (forthcoming) The Creation and Re-Use of Information in Gene Regulatory Networks. Proceedings of the Philosophy of Science Association

Carroll S (2006) Endless Forms Most Beautiful. Norton Publishing, New York

Dretske F (1988) Explaining Behavior: Reasons in a World of Causes. MIT Press, Cambrdige

Gallistel CR, King AP (2009) Memory and the Computational Brian. Wiley- Blackwell Publishing, West Essex

Hamilton WD (1964a) The genetical evolution of social behaviour. Part I. Journal of Theoretical Biology, 7:1–16.

Planer R (2014) Replacement of the "genetic program" program. Biology and Philosophy, 29:33-53

Shannon C (1948) A Mathematical Theory of Communication. Bell System Technical Journal, 27:379-423

Skyrms B (2010) Signals: Evolution, Learning, and Information. Oxford University Press, Cambridge
Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. Development, vol. 5, pp. 112-113
Wolpert, L. (1994). Positional information: revisited. Development, supplement 1-3

Chapter 5

On the free-rider identification problem

To appear in *Biological Theory*, DOI: 10.1007/s13752-015-0206-2.

In their book *A Cooperative Species* (2011), Samuel Bowles and Herbert Gintis launch a sustained attack against individual-selection accounts of human cooperation. Their attack is multi-faceted and is backed by an impressive range of modeling, experimental, and archeological evidence. Here my aim is to carefully consider but a single strand in their overall argument.

Bowles and Gintis argue that individual-selection accounts of human cooperation flounder in the face of the free-rider identification problem. This is the problem of identifying individuals who have defected (so as to secure personal gains) in a cooperative interaction, and making this information widely known throughout the group. In order for cooperation to arise and remain stable in even modestly-sized, non-kin based groups, it is imperative that free riders (and more or less only free riders) be punished. However, this cannot happen unless free riders are reliably picked out and collectively designated as such by the group.

Kim Sterelny (2012) has recently responded to this line of argument for group-selection, arguing that the free-rider identification problem is a "pseudo-problem" (p. 181). What exactly does Sterelny mean by this? We can be sure he isn't claiming that the free-rider identification problem, at least as I have just described it, is not one that actually exists. Instead, what I take Sterelny to be claiming is just that, *pace* Bowles and Gintis, the free-rider identification problem poses no real theoretical difficulty for

individual-selection accounts of human cooperation. There is no reason to think that groups in which this problem is solved are necessarily ones in which individuals have had their psychologies shaped by a process of group selection.

I find Bowles and Gintis' argument from the free-rider identification problem to group selection interesting but opaque. In this article, my first task is to clarify their argument. I suggest that the real crux of their argument is this: even in modestly-sized social groups, the free-rider identification problem can only be solved provided that group members are disposed to share social information. The difficulty individual-selection accounts confront, according to Bowles and Gintis, is that these accounts have no explanation for why individuals should be disposed to behave in this way. Having clarified their argument, I then consider its plausibility in light of Sterelny's criticism. While there is much in Sterelny with which I agree, I think he ultimately underestimates the challenge being raised by Bowles and Gintis for individual-selection accounts. More specifically, I argue that it is unclear whether the expected benefits of having a disposition to share social information would have outweighed the expected costs for an individual belonging to a Pleistocene social group. None of this is to say that I am ultimately convinced by Bowles and Gintis' argument. Rather, my view is that more theoretical (and in particular) empirical work must be carried out before this debate can be settled, and I conclude by formulating some specific questions which I think future research in this area should aim to address.

5.1 Bowles and Gintis on the free-rider identification problem

It is widely held that kin selection is a non-starter when it comes to explaining human cooperation. Reason being, the groups our Pleistocene ancestors lived in are believed to have contained too many non-kin for this selection mechanism to have been effective. Hence, in the view of most researchers, we are left with two main individual-selection mechanisms for explaining cooperation, namely, reciprocal altruism (Trivers 1971; Axelrod and Hamilton 1981) and indirect reciprocity (Sugden 1986; Alexander 1987; Nowak and Sigmund 1998), both construed broadly[37]. Bowles and Gintis claim that, for even modestly-sized social groups, explanations based upon either individual-selection mechanism (including some combination of the two) break down.

The following passage offers a nice summary of Bowles and Gintis' view on reciprocal altruism. They write:

> The reason for the ineffectiveness of reciprocal altruism for groups with several members is simple. In groups of two, a free-rider cannot go undetected because a player's payoff reveals the other player's behavior. Equally important, when one member defects in order to punish a Defector, the punishment is uniquely targeted on the Defector. But, in groups larger than two, a player cannot infer who has defected from the knowledge of his own payoff. Moreover, a retaliatory defection punishes not only the initial defector, but also all other members of the group. Moreover, other group members may not have observed the initial defection and hence may think that a retaliatory defection is a free-riding defection, inviting further retaliatory defections. (p. 64)[38]

There are two distinct issues being raised by Bowles and Gintis here. The first concerns the *identification* of free riders; the second, the *targeting* of free riders (and only free riders) for punishment. Although the second issue is also of critical importance to the

[37] By "construed broadly" here, I simply mean explanations revolving around repeated interaction (in the case of reciprocal altruism) and ones revolving around reputation (in the case of indirect reciprocity). Below, we will look at a third type of individual-selection mechanism, namely, partner choice, which also relies crucially upon reputation. In this regard, partner choice is closely tied to indirect reciprocity, and I wait until later on in this article to fully distinguish the two.

[38] All page numbers following block quotes in this section refer to Bowles and Gintis (2011).

more general debate between individual-selection and group-selection proponents in this area, I will set it aside in what follows.

Bowles and Gintis point out that when a cooperative interaction (with "simultaneous" moves) involves more than two interactants, there is no guarantee that a given interactant will know who has cooperated and who has defected, as this information cannot be ascertained from one's payoff. In this regard, such interactions are importantly different from dyadic ones. Bowles and Gintis also point out that uncertainty about who has cooperated and who has defected in the context of past interactions can have a kind of ripple effect. For example, if an onlooker A is uncertain as to whether B defected against C in the past, then there is no guarantee that A will be able to tell whether C's current defection against B constitutes a form of free riding or a form of punishment. Bowles and Gintis take these considerations to cast doubt on the claim that reciprocal altruism can offer an adequate explanation for human cooperation. But why, exactly?

Before delving into this question, let us have a look at what Bowles and Gintis say regarding indirect reciprocity. They write:

> [Indirect reciprocity models have] very demanding informational requirements. [Their] ability to target defectors makes [them] an improvement on the repeated game models … but this comes at a steep price. Each individual must know the current standing of each member of the group, the identity of each member's current partner, and whether each individual cooperated or defected against his current partner, since this information is necessary to ascertain the status of one's partner in the indirect reciprocity game. But real-world dyadic interactions are often private, and hence unlikely to be directly observed by more than a small number of others, vitiating the model for groups of any significant size. (p. 70)

In this passage, we see Bowles and Gintis' objecting to the (in their view) demanding "informational requirements" of indirect reciprocity. Surely, it is reasonable

to be more impressed by some of these requirements than others. For example, while knowing the identity of one's current partner(s) is probably trivial in the vast majority of real-life cases, knowing the status of one's partner(s) (whether they are in "good" or "bad" standing in the group), or the status of the individuals one is watching interact, is definitely not trivial. Bowles and Gintis are bothered by something else: they think that, even in small-scale societies, a significant number of cooperative interactions take place in the presence of few or no onlookers.[39] But again, why exactly do Bowles and Gintis take these considerations to undermine explanations of human cooperation based upon indirect reciprocity?

To be sure, there is a uncharitable way of interpreting both sets of remarks here. It can seem like what Bowles and Gintis are claiming is that the free-rider identification problem is effectively insoluble in even modestly-sized social groups. Or put more fully: Repeated interaction ceases to support cooperation in such groups as interactants will often be left not knowing who has behaved cooperatively and who has behaved selfishly. In addition, reputational considerations cease to support cooperation because there is just far too much for group members to keep track of. The representations group members possess of the past behavior of others are bound to be significantly incomplete at best, implying that free riders will often be treated as cooperators (and perhaps cooperators as free riders). Consequently, humans' disposition to cooperate cannot be explained by appeal to the individual benefits of such a disposition in the evolutionary past. An individual belonging to a Pleistocene social group would have been *worse off*, not *better*

[39] Ideally, one would like to see claims about the frequency of (more or less) private cooperative interactions buttressed by concrete ethnographic evidence. Unfortunately, there is not space to consider the ethnographic literature on this issue here. I set aside skepticism over Bowles and Gintis' claim going forward as my impression is that they are on relatively firm ground.

off, as a result of having a disposition to cooperate, as free riding would have often gone undetected, and so unpunished. Hence, to explain why humans nevertheless evolved such a disposition, we must appeal to selection operating at the level of groups.

The reason I say this would be an uncharitable interpretation of their argument is simple (although I do not think Bowles and Gintis do enough to discourage this reading of their argument). It is not just a problem for individual-selection accounts of cooperation if free riders cannot be reliably detected. That is, it is also crucially important, if not equally so, for a group-selection account such as Bowles and Gintis' that the free-rider identification problem can be solved in groups like the ones our Pleistocene ancestors lived in. A free rider in a group of cooperators will have a higher fitness, and so free riding will tend to spread through the group unless free riding is identified and punished. So, even if a group consisted solely in cooperators at some initial point in time (in which case, there wouldn't actually *be* any free riders to identify at that time) an inability to solve the free-rider identification problem would inevitably give way to what Richard Dawkins has colorfully described as "subversion from within" (Dawkins 1973). That is, "mutation" would eventually serve to introduce free riders into the group, and once introduced, such individuals would begin to overtake the group. Consequently, levels of cooperation would begin to decay and eventually collapse altogether. Bowles and Gintis are surely aware of this kind of argument against group-selection accounts of cooperation, as it has long been held up as group selection's Achilles' heel. In a word: group-selection proponents such as Bowles and Gintis need a viable explanation for how the free-rider identification problem gets solved every bit as much as their individual-selection opponents.

Fortunately, I think there is a much more plausible way to interpret Bowles and Gintis' argument, one that is suggested by what they have to say on the topic of *informational sharing* between group members. Still discussing the (purported) shortcomings of explanations of cooperation based upon indirect reciprocity, they write:

> … if individuals are entirely self-regarding, they have no reason to report truthfully [the behavior] they have observed. Though an active area of research, explanations of how private information could be converted to accurate public information in a population of amoral self-regarding individuals have not been presented.
>
> Because the truth-telling that is necessary to convert private to public information cannot be expected in the absence of social preferences and because public information is essential to the empirical plausibility of both the simple reciprocal altruism model and its indirect reciprocity variant, these models do not provide adequate explanations of cooperation among amoral self-regarding individuals. (p. 70)

To me, this passage lays bear the ultimate reason why Bowles and Gintis think that explanations for cooperation based upon either reciprocal altruism or indirect reciprocity are bound to be inadequate. The reason is *not* that the free-rider identification problem is insoluble in even modestly-sized social groups. Rather, it's just that solving this problem requires that group members are disposed to share (truthful) information they possess about one another's behavior.[40] I will call information about others' behavior (including others' dispositions to behave in various ways) "social information" in what follows.[41]

Consider reciprocal altruism first. As we saw above, Bowles and Gintis draw attention to the fact that, when a cooperative interaction involves more than two interactants, there is no guarantee that a given interactant will know who has cooperated

[40] In what follows, I drop the qualifier "truthful". The reader should assume that when I talk of sharing information, I mean sharing truthful or accurate information, unless I explicitly say otherwise.
[41] I use the term "social information" rather than "gossip" so as to avoid the negative connotations associated with the latter.

and who has defected. This is because the payoff they receive does not reveal such information. The importance of this point, I suggest, isn't that it will therefore be hard or costly for group members to come to know who has cooperated and who has defected; instead, it's that often one will have no choice but to rely on the reports of others to ascertain such information (for example, those who were in a position to directly perceive the act(s) of defection). At this stage of the dialectic, Bowles and Gintis will then ask (quite rightly in my view) why it is legitimate for individual-selection accounts of cooperation to simply assume that individuals are disposed to share such information. As the last passage makes clear, Bowles and Gintis' view is that individual-selection accounts have no explanation for why individuals should be disposed to behave in this way. Proponents of group selection, by contrast, will say that this disposition reflects selection at the level of groups (i.e., groups in which individuals were disposed to share social information with one another out-survived/competed those in which individuals were not disposed to behave in such a manner.[42])

---

[42] Ultimately, one will want to see a more detailed group-selection explanation for the evolution of the disposition to share social information. The motivation for such a desire is that there would seem to be another "subversion from within" problem lurking in the background. If this disposition is, on the whole, individually costly, then those who lack the disposition will have a higher fitness than those who possess it when each belong to the same group. What, then, is to prevent individuals who lack this disposition from overtaking the group? (I thank an anonymous referee for drawing my attention to this issue.)

I think there are a couple of things proponents of group selection can say at this point, though I do not deny that matters are complicated. Here, for example, is a sketch of one line of response which I find to be *prima facie* plausible. On the reasonable assumption that the costs of possessing this disposition (if indeed there are some) aren't too significant, then the difference in fitness between an individual who lacks this disposition and one who possesses it cannot be too significant either. To provide a point of contrast: this fitness difference is bound to be much smaller than that between an individual who, say, lacks the disposition to share food (and yet is not punished/ostracized by group members) and one who is disposed to share food. Hence, in the case of the disposition to share social information, it may not be that difficult for selection operating at the level of the group to offset the process of subversion from within, despite the slower pace of the former process.

On the other hand, one might think that the only fair-minded thing to say in this context is that the evolution of this disposition is actually a theoretical difficulty for *both* individual-selection and group-selection accounts (though for different reasons). (This is the view suggested to me by the above-mentioned referee.) On this view, Bowles and Gintis' argument would still be seen as raising an important evolutionary question (i.e., how could this disposition evolve?), just not one which group-selection

Matters are even clearer in the case of indirect reciprocity. For as Bowles and Gintis point out, reputational effects will only work to promote cooperation if group members in fact possess copious amounts of true social information about one another. Surely, some of this information can be gathered on the basis of one's own perceptual experience (including one's experiences of observing others interact), but not all of it. This is especially clear in cases where the interaction takes place in private (e.g., two individuals go off hunting together). The rest of this social information, which is a sizable package, to be sure, must be supplied by others. But again, Bowles and Gintis will want to know why we should expect group members to share social information with one another. Until individual-selection proponents have offered an explanation for how this disposition can evolve via purely individual-selection means, they will not have offered an adequate account of human cooperation.

In sum: Bowles and Gintis take the free-rider identification problem to pose a serious stumbling block for individual-selection accounts of human cooperation. In this section, I have suggested that the real crux of their argument comes down to the following: the problem of free-rider identification can only be solved provided that group members are disposed to share social information. Bowles and Gintis' view is that group-selection accounts have a ready explanation for this disposition, one which human beings clearly do possess[43], whereas individual-selection accounts do not.

---

accounts have an obviously easier time answering. I'm inclined to think that this view is a bit too strong, but I will not argue for that claim here.

[43] See, for example, Emler (1992, 1994), Dunbar et al. (1997), and Wiessner (2005). Emler sampled from a wide-range of conversational contexts (his subjects were adult Westerners) and found that about 70% of time spent conversing involves the transmission of social information. Following this study up, Dunbar et al. examined the content of conversations taking place between students in a university refectory. They found that about 60% of conversational time is spent on social information. Finally, based on 308 multi-hour recordings of conversations between Ju/'hoansi !Kung individuals, Wiesnner reports that 56% of conversations included criticism of group members, and (only!) 7% praise of group members.

5.2 Sterelny's criticism of Bowles and Gintis

Sterelny thinks that the problem of free-rider identification has been much exaggerated by researchers like Bowles and Gintis. For Sterelny, the only real challenge for individual-selection accounts of human cooperation is why our Pleistocene ancestors would have been disposed to punish free riders (thereby incurring at least a short term cost) once free riders were identified. Because Sterelny thinks that a disposition to punish can be explained wholly within an individual-selection framework, he is unconvinced that we need to appeal to group-selection to explain how human cooperation evolved.[44]

Here is the key passage from Sterelny on Bowles and Gintis' argument:

> … Bowles and Gintis think that the problem of scale is intractable for individual-selection accounts of cooperation. … Cooperation is even unstable in modestly sized groups, unless free riders can be specifically targeted. That presupposes that such free riders can be identified and, once identified, punished. I do not see identification as a serious problem: I have already argued that in small, repeatedly interacting band-sized groups, agents will become well-informed about one another just by direct observation and memory. This effect is magnified by gossip, once such communicative capacities evolve. Of course, gossip is not perfectly honest and reliable, but as I have argued … multisender, multireceiver networks are well insulated against deception, and participating in those networks brings benefits of reciprocation and information pooling. No social environment is perfectly transparent, but in the social environments in which strong-reciprocity psychologies evolved in, agents were awash in information about their peers. (p. 182-3)[45]

Sterelny puts several important ideas on the table in this passage. First, Sterelny claims that direct observation and memory go considerably further towards identifying free riders than researchers (in particular, modelers) such as Bowles and Gintis generally realize. Second, he claims that the "shape" of human communication networks can serve to significantly damp down deceptive exploitation. Third, he claims that participating in

[44] It should be noted that Sterelny is not in general opposed to the idea that group selection has played a role in human evolution (see p. 178 [2012], for example); he simply rejects the idea that we need to appeal to group selection to explain how the disposition to cooperate evolved in humans.
[45] All page numbers following block quotes in this section refer to Sterelny (2012).

such networks produces real benefits for the individual. Let us consider each of these ideas, respectively.

I agree with Sterelny that modelers of human cooperation often make entirely unrealistic assumptions about the cognitive abilities of members of real social groups. Bowles and Gintis are no exception. A good example of this tendency, one Sterelny actually singles out, can be found in the first passage from Bowles and Gintis cited above. Bowles and Gintis tell us that onlookers "may think a retaliatory defection is a free-riding defection, inviting further retaliatory defections" (p. 64, 2011). Regarding this assumption, Sterelny writes that:

> As the models represent [cooperative] interactions, onlookers see no intrinsic difference between one agent defecting against a second, and that agent punishing the second. In real social environments, these are very different. For example, communication among those interacting will be very different. The defector will be attempting to persuade his victim that he is in fact cooperating; in the case of punishment, the agent will be denouncing his target. Denunciation will not tell us whether the punishment is just. But onlookers will nonetheless see two very different interactions. (p. 184)

I am inclined to side with Sterelny here. On the other hand, his concession that observation alone will not suffice to tell us whether the one individual is justified in punishing the other is an important one. (If an onlooker is unable to tell whether the punishment is just or unjust, he or she will not know which attitude it is appropriate to adopt towards the interactant who is doing the punishing.) More generally, though, I contend that even if Bowles and Gintis do underplay the role of direct observation and memory in identifying free riders, the core of their argument remains intact. If the free-rider identification problem is to be solved in all but the most intimate social settings, it is crucial that individuals are disposed to share social information, and to abstain from deceptive manipulation. Surely, this will at least be true when many cooperative

interactions only involve a subset of group members, and take place in private. My guess is that Sterelny would concur, as he himself draws attention to the role played by gossip in solving the free-rider identification problem.

Sterelny's claim that the shape of a communication network can serve as an honesty-enforcing mechanism is, in my view, spot on.[46] In particular, Sterelny points out that when the receiver of a message can check the message's truth by consulting several or more group members, it will often be in the sender's interest to communicate truthfully, if he communicates at all. Deception in such a context is not only unlikely to be effective; it can indeed serve to give one a bad reputation. In addition, Sterelny points out that as the number of receivers that a sender communicates with goes up, so too does the chance that his deception will be detected.[47] This is because it becomes less likely that all of the receivers involved will be ignorant in a way that makes them vulnerable to the deception. For example, if a sender falsely relates that a food patch has been exhausted, he may well persuade those individuals who have not visited the patch of late. But he will *not* succeed in persuading those who have just visited the patch themselves. So, if the sender's deception is broadcast widely, there is a good chance that it will be received not only be individuals of the former sort, but also by those of the latter. Together, Sterelny

[46] Sterelny does not claim that this is the only honesty-enforcing mechanism, of course. In particular, he approvingly cites the work of Sperber (e.g., Sperber 2000, 2001) on the role of folk epistemology in helping us to evaluate the reliability of various sources. What Sterelny disagrees with Sperber (and many others) on is that epistemic vigil against deception is the primary means of keeping deception in check. Sterelny's point is that the shape of human communication networks already does much of this work for us.

[47] The communal meetings which many hunter-gathers hold at the end of each day provides us with a vivid illustration of a communicative situation involving very many receivers. For example, Hames and Vickers (1982) report that, "When [Siona-Secoy hunters, an indigenous Amazonian people] return from the forest, they gather to discuss their various successes and failures, the signs of game encountered, the location of fruiting plants favored by particular species, and the difficulties of tracking, stalking, and pursuit" (p. 368). Other examples can be found in Ray (1963), Marshall (1976), and Tanner (1978).

claims that these two features of human communication networks will work to undercut the temptation of self-regarding individuals to deceive their group mates.

Let us assume that ancestral human groups were in fact characterized by the sort of communication networks Sterelny has in mind here (i.e., many-to-many networks). If so, then Sterelny does have an explanation for why the individuals belonging to such groups would have been disposed to abstain from the spreading of false information, an explanation that is at the very least *prima facie* plausible. This is key, as widespread deception concerning others' behavior certainly seems like it would render reciprocal altruism and indirect reciprocity incapable of supporting cooperation. But, one might wonder, is the many-to-many nature of such communication networks really something which individual-selection accounts of cooperation are entitled to help themselves to? I take it that part of what makes it the case that a given communication network is many-to-many is that the nodes belonging to the network are disposed to share information with one another. Would such a disposition have served to confer a net expected benefit upon the individual? And more to the present point, would a disposition for sharing *social information* in particular have served to confer a net expected benefit on the individual? Bowles and Gintis would claim that a disposition to share social information would *not*, on the whole, have been individually advantageous. Sterelny thinks otherwise. He writes:

> It is true that sharing information via gossip is indeed an *N*-player cooperation problem. But the costs and benefits of informational cooperation do not mirror those of ecological cooperation. Costs are lower; benefits are higher. The benefits of informational sharing increase as *N* increases, for Condorcet-like effects increase the reliability of consensus. And the more gossip is multisourced and multitargeted, the less likely it is to be deceptive manipulation. Moreover, in contrast to sharing material resources, the costs of sharing information do not increase with the number of agents aided. But the number of reciprocation sources does rise with informational targets. You can help more people—and hence legitimately expect help back from more

> people—for a fixed cost. Finally, Nowak and Sigmund [2005] point out that experimental evidence shows that agents are sensitive to others' reputation and expect others to be sensitive to their own reputation. (p. 183)

In my view, this passage cuts to the very heart of the debate at hand. Sterelny suggests that sharing social information is correctly viewed as posing a kind of cooperation problem in itself. On this point, Bowles and Gintis would surely agree. However, Sterelny *also* thinks that individual-selection accounts have a ready explanation for why individuals would be disposed to share social information. I think matters are a good deal more complicated than Sterelny realizes.

First, Sterelny again draws attention in this passage to how the many-to-many shape of communication networks can serve to damp down deception. I am happy to go along with the idea that the communication networks via which social information flowed in ancestral human groups were in fact of this sort. However, the question we are presently interested in is *why* the individuals belonging to these groups were disposed to share social information in the first place. Why were they willing to share social information as opposed to simply keeping information of this sort to themselves? Appealing to network shape in order to explain the existence of such a disposition not only seems to get the order of things backwards (as I just mentioned, a network for transmitting social information *counts* as many-to-many in virtue of the fact that the network's nodes are disposed to share social information); it also strikes me as a bit of a red herring. The *truthfulness* of the information that in fact flows through some communication network is one thing; why there is a network there *at all* is another. (Of course, the two are not *entirely* unrelated. In particular, should deception become rampant, receivers will presumably stop listening.)

Second, Sterelny says that Condorcet-like effects help to explain why individuals would have been disposed to share information. Here he is alluding to the Condorcet Jury Theorem which tells us that if each "juror" has a more than 50% chance of being correct, then the probability of the majority view being correct goes up with the number of jurors. This suggestion is on the right track in the sense that it is the right kind of reason for explaining why individuals would have been disposed to share information. Sterelny's idea is that it would have been in individuals' own self-interest to pool their information so as to produce Condorcet-like effects. In this way, each individual would have regularly gained access to more accurate judgments, which is something that surely would pay at the individual level. I agree with Sterelny that the logic behind the Condorcet theorem provides a compelling explanation for why ancestral humans would have been disposed to share *some* kinds of information; what I am skeptical of is that this logic applies in the case of sharing social information. If, for example, a group must decide whether they have a better chance of surviving if they move to a new location rather than staying put (suppose they are faced with a severe drought), then the Condorcet theorem tells us that it is in the self-interest of each individual to speak up. What is far from clear to me that this sort of case is a good general model for thinking about the sharing of social information. To give a concrete example: suppose that A discovers that B is hoarding food. Can a disposition for A to inform members of his group about anti-social behavior of this kind really be explained, even in some ultimate sense, by A's gaining access to a more accurate judgment about B's personality type? At the very least, we are surely owed another layer of detail here, one focused specifically on the sharing of social information, not information in general.

Sterelny offers two further considerations in support of his claim that a disposition to share social information can be explained via individual-selection alone. The first of these concerns *reciprocal* information sharing. Put roughly, the idea here is that if A shares information with B at time *t*, then B is more likely to share information with A at a future time *t′* than B would have been otherwise. The real importance of this point for Sterelny in the present context is that (according to him) the expected costs of sharing some piece of information remain fixed as the number of group members with whom this information is shared increases. Consequently, it will be quite easy for the expected benefits of sharing this information to outweigh the expected costs. I will look at this line of reasoning in more detail in the next section.

The other consideration Sterelny raises here concerns individuals' reputations. Sterelny cites the experimental work of Martin Nowak and Karl Sigmund (2005) showing that individuals condition their behavior (cooperate vs. defect) on the reputations of those with whom they interact, and expect others to do the same. There is a narrow and broad way of thinking about results like Nowak and Sigmund's. Looked at narrowly, what these results tell us is that when group members have access to accurate information about how one has behaved in past cooperative settings, it will often be in one's self-interest to behave cooperatively, even when the chance of a repeat interaction is low (or possibly even nil). Looked at broadly, these results simply drive home something which is clear to each of us from everyday life, namely: humans typically reflect upon what they know about individuals with whom they are interacting, or with whom they are contemplating interacting, in deciding on a specific course of action. On this view, the importance of reputation extends far beyond prisoner's dilemma-style interactions; in particular,

reputation is a key factor influencing *partner choice*.[48] In choosing a partner, we not only attend to how some candidate partner has behaved towards ourselves and others in the past; we may also factor in general personality traits (e.g., is brave, is hot-tempered, etc.) as well as expertise (e.g., is a plant-expert, is mechanically-inclined, etc.).

It is very probable that reputational effects (construed broadly) will figure centrally into any plausible individual-selection account of informational sharing.[49] In sharing information with one's group members, one can often indicate to others that one possesses certain otherwise hidden attributes that make one a desirable partner for future activities, activities that will produce real benefits for the individuals involved.

To see how this might go, consider the following case. Consider an individual A who is disposed to share information with his group mates which he has gathered while out hunting alone. On this particular day, A observes tracks and other signs left by a prized game animal. On the basis of these signs, A is able to work out the species, size, health, and likely whereabouts of the animal. Upon returning to camp, A, as usual, shares this newly-acquired information with the his fellow hunters.

From an individual-selection perspective, A's disposition can indeed appear quite puzzling. Reason being, such acts of informational sharing on A's part would seem to produce real costs for A.[50] For example, when A informs others that he has observed signs of a high-quality game animal, and furthermore tells them where he believes the

---

[48] For state-of-the-art work on partner choice in the human domain, see Baumard et al. (2013).

[49] I am indebted to an anonymous referee for this insight. Without his/her detailed comments, I would not have appreciated the importance of partner choice in this context. The material presented in the remainder of this section draws heavily upon those comments. The main way in which my discussion goes beyond that suggested to me by the referee is in showing how the precise *content* of the information that is shared can signal the possession of specific personality traits (e.g., is very knowledgeable about animal behavior).

[50] To be clear: I here assume that the hunter is perfectly capable of bagging the prey without enlisting the help of others. If he wasn't, then his sharing of such information, at least with some of his group members, wouldn't be at all surprising. I thank an anonymous referee for this clarification.

animal to be located, A lowers the chance that he will be the first to lay eyes upon the animal and make the kill. This is because other hunters now strike off in pursuit of the prey.[51]

Wouldn't A be better off *not* having such a disposition? It is surely tempting to think so. This impression is made all the stronger, I think, by the fact that were A to frequently withhold certain pieces of valuable information from the group instead, others would generally not be in a position to directly perceive such acts of defection on A's part. In this regard, refusing to share information is importantly different from refusing to share material resources like food or to provide physical assistance to others (when one is visibly strong).

The problem with this way of thinking is that it does not take into account the benefits our individual A would be likely to see as a result of having such a disposition. In regularly sharing valuable information which he has gathered while out hunting, A not only builds and maintains a reputation as a helpful and generous member of the group; in addition, *he cultivates the perception in others that he is keen-minded and a strong natural historian* (not just anyone can recognize animal tracks, infer the quality of the animal, infer its direction of movement, etc.).[52] Of course, these are just the sorts of (otherwise hidden) attributes that make one an excellent partner for a wide-range of hunter-gather activities. Thus, while A's disposition may well have real individual costs, these costs (one might think) are ultimately outweighed by the reputational gains that

---

[51] Of course, even if meat is typically shared amongst the group, it is not hard for us to imagine reasons why each hunter prefers to be the one who actually bags the prey. For example, he who bags the prey may enjoy an uncontested claim to some especially highly-valued portion of the meat.

[52] Marlowe (2010), for example, relates how Hadza people attempt to signal that they have "good eyesight, coordination, strength, knowledge, endurance, [and] bravery" (p. 231). Marlowe's discussion concerns the sharing of particular foods (it may take great bravery, say, to secure a certain food type), but shared information (if true) could serve essentially the same purpose.

flow from A's informational sharing. A is more likely to be chosen for future profitable activities by members of his group than he would have been had he not shared information in this way.

To be sure, more than a few questions arise at this point regarding the plausibility of this sort of explanation for informational sharing. Unfortunately, it is beyond the scope of this article to go fully down that path. What I plan to do instead is consider the applicability of this sort of explanation to the case of sharing social information.

5.3 Sharing social information

In this section, I turn to the expected costs and benefits of having a disposition to share social information.

It is helpful to begin by stating what is common ground between Bowles and Gintis, and Sterelny. First, both agree that humans are disposed to report on the behavior of others. (It can hardly be doubted that humans relish the opportunity to engage in some "good" gossip.) Second, both (presumably) agree that there is a widely-obeyed norm of truth-telling as concerns the transmission of social information. Finally, both would agree that these two facts form (at least part of) the foundation for an explanation of how the free-rider identification problem is solved in all but the most intimate social settings. Were humans *not* so disposed, then free riders would too often go unidentified and so unpunished. Consequently, cooperation within the group would begin to break down and eventually disappear altogether.

On what do the two parties *disagree*, then? As I see matters, what they ultimately disagree on is whether an individual-selection explanation for these facts about human

communication can be provided. Bowles and Gintis think that no such explanation is forthcoming. Sterelny, by contrast, thinks that proponents of individual-selection have a ready explanation.

Below, I set aside issues about truthfulness. I agree with Sterelny that the opportunities individuals encounter to engage in successful deception have been much overestimated by researchers. It *is* true, however, that those features of human communication networks which serve to damp down deception (the fact that information is multi-targeted and multi-sourced, as Sterelny says) should not be taken for granted by proponents of individual-selection, but I will not consider this issue further here. Instead, in the remainder of this article, my focus will simply be on why individuals are disposed to share social information at all. In determining whether an individual-selection explanation can ultimately be provided for this disposition, we need to inquire into the expected costs and expected benefits of having such a disposition. If it is plausible that the expected benefits would have outweighed the expected costs, given the kind of social groups our Pleistocene ancestors lived in, then I take it that a disposition to share social information can indeed be explained along individual-selection lines. But if it is plausible that the expected costs would have *failed* to outweigh the expected benefits, then I take it that some other kind of explanation must be provided (i.e., either a group-selection explanation or else some kind of non-adaptive explanation[53]).

[53] The fondness which we show for sharing social information certainly stands out in comparison with our feelings towards sharing other kinds of information. One might think that this constitutes strong evidence that our disposition to share such information has been shaped by one or another selective force. Put differently, one might think that it is very unlikely that we would show such a fondness for sharing social information by chance alone or as a by-product of some other psychological trait. If this is correct, then we should obviously be quite skeptical of any non-adaptive explanation for our disposition to share social information. This is really a topic deserving of an in-depth treatment of its own, however. I take no definite stance on the relative plausibility of adaptive vs. non-adaptive explanations for this disposition in the present work.

Let us consider the expected costs first. To be clear, what we are looking for are expected costs that *typically* accompany acts of sharing social information, not ones that arise from the details of highly-specific social interactions. (For example, we *don't* want to factor into our analysis at this stage cases in which one would be violating the expectations of one's communicative partners, were one to share some piece of social information with them.) I think there is at least one major source of expected costs that individuals incur by sharing social information, though this expected costs is limited only to the sharing of *negative* social information, that is, information about socially disapproved of behaviors. (I think negative information is pretty clearly the more important of the two types of social information—the other being *positive* social information, that is, information about socially approved of behaviors—when it comes to sustaining cooperation, though I won't argue for that claim here.) The expected costs stem from the relation that the sharer bears to the *target* of the information being shared, whereby "target", I mean the individual whose behavior the information is about.

Suppose A shares some negative and therefore potentially damaging information about individual B with group members C, D, E, and so on. In so doing, A incurs a risk. The risk is this: should B come to know that A has shared this information with members of the group, B may well retaliate against A. Of course, one way B may learn of this fact is that one of the individuals A informs (or one of the individuals that is informed by an individual that A informs, etc.) simply tells B. That is, some individual may alert B (perhaps in an attempt to win favor with B) that A is sharing information about B's anti-social behavior. Alternatively, if the information in question is something *only* A would know, as would be the case when the information concerns B's behavior in the context of

some private interaction between A and B, then the fact that this information has spread beyond A will, by that very fact, alert B that A (and possibly others) have shared this information. Now, B's retaliation against A may assume a variety of forms. One thing B might do is simply break off a preexisting social or utilitarian arrangement with A, an arrangement which has brought A real benefits in the past. (For example, perhaps B has been a useful hunting partner to A in the past, and now B refuses to hunt with A; or perhaps B is an expert weaponry-maker, and now B will no longer trade with A.) However, B might not be so cool-headed. In the extreme case, B's retaliation against A may take the form of full-fledged physical violence against A. Whether retaliation of the latter sort would ultimately be more costly for A than retaliation of the former will of course depend on a range of variables. At any rate, the important thing to note is just that, in sharing negative social information with one's group mates, an individual thereby incurs the possibility of real target-induced costs. Moreover, it would appear that the expected costs will in general *rise* with the number of individuals that one informs, assuming that sharing this information with a larger number of individuals increases the chance of the target learning that one has shared negative information about him.

What about the expected benefits that typically accompany the sharing social information? In the last section, we mentioned two possible sources of expected benefit, namely, reciprocal informational sharing and reputational gains. Let us take these, respectively. Suppose that A shares information with C about B's (let's say, anti-social) behavior. In so doing, A may be contributing a real benefit to C; C may have been considering B as a partner for a future activity, one requiring trust and honesty, and now C knows she should avoid B. C recognizes this good deed on A's part and is inclined to

pay it back. More specifically, C now provides A with some valuable social information (or perhaps some other kind of information[54]) at a future point in time. The contrast case to consider here is one in which A instead withholds this information from C, and C, as a result, withholds valuable social information from A in the future. Assuming this kind of pattern generalizes, it is easy to see that A gains real benefits from a disposition to share social information. A can reliably expect to receive social information back from all, or at least most, of those individuals whom she has aided with social information. Moreover, at least in the case of positive social information, we would expect (apparently incorrectly, as it turns out—see below) for the expected costs of sharing such information to remain fixed as the number of individuals one informs goes up.

Moving on to reputational gains: it has been suggested that, by sharing social information, one can send a signal to others that one enjoys a place of social power or influence within one's group (e.g., Kurland and Pelled 2000; Foster and Rosnow 2006; Guerin and Miyazaki 2006; Dunbar 2004). The idea here is that only those individuals who enjoy a place of social power within the group will have much social information to share; hence, one should indeed share this sort of information, and share it widely, if one has it. This explanation makes use of the same kind of logic which we used to explain a disposition to share ecological information (or more precisely, information acquired while out hunting) in our example from the last section. That is, in both cases, the acts of informational sharing which flow from the disposition work to cultivate the perception in

[54] The possibility that acts of sharing social information might be "paid back" by acts of sharing other kinds of information (e.g., ecological information), or just information more generally, is an important one. There is not room to consider this possibility here, however, and so I set it aside in what follows. What I want to focus on is the possibility that, in sharing social information, an individual gains access to a specific domain of information (social information) that she would lack access to if she did not herself share social information.

others that the sharer has certain attributes that make him or her an attractive partner for some range of activities (in the case at hand, the revealed attribute is a high degree of social power.)

I claim that we cannot hope to settle the issues at play by simple reflection on the possible costs and benefits of sharing social information in the sorts of groups our Pleistocene ancestors lived in. We can imagine reasons why having such a disposition might work to the individual's benefit, but we can also imagine reasons why it could hurt the individual. Unfortunately, what little experimental work exists in this area hardly gives us a clear verdict. At most, what this work entitles us to conclude is that being disposed to share social information at high frequencies (to report on others behavior often, and to a wide audience) is likely detrimental to the individual on the whole (Turner et al. 2003; Farley 2011). (Surprisingly, this appears to be true even when much of the social information one shares is of the *positive* variety.) In a word, those who share social information at high frequencies are apparently perceived as being less trustworthy, less likeable, and less socially powerful than those who share social information at low frequencies. Of course, it is not hard to see how such perceptions on the part of others might translate into real costs for the individual (and this is to say nothing of the target-induced costs mentioned above). Most obviously: having a reputation for being a "big mouth" may well deter others from sharing social information with you, and may even cause you to be excluded altogether from social interactions where sensitive information is likely to be divulged.[55] What this work does *not* tell us is how a disposition to share social information at a more moderate level might bear upon individual fitness (a point

[55] The reason I did not include these expected costs earlier in the discussion in this section is that they flow from a rather extreme form of the disposition to share social information.

Farley herself makes in discussing the shortcomings of her study), although it does seem to rule out simple-minded versions of the idea that sharing social information can be explained by status enhancement.

More theoretical and (in particular) empirical work[56] is definitely needed in this area before we can hope to definitively answer whether the human disposition to share social information can be explained from a purely individual-selection perspective. To begin with, we will presumably need a much more precise description of the patterns of informational sharing which individuals actually engage in. (Since there are bound to be individual as well as cultural differences as regards such sharing, we will eventually want to replace talk of "the disposition to share social information" with more detailed language. I bracket this complication below, as I have done so far in this article.) More specifically, we will need to know, among other things:

- How is individual decision-making about whether to share social information sensitive to the relationship between a potential *sharer* and the *target* of that information?
- How is such decision-making sensitive to the relationship between a potential sharer and his or her *communicative partner(s)*?[57] and
- How is such decision-making sensitive to the *specific content*[58] of information that is a candidate for sharing?

---

[56] I construe "empirical" broadly, so as to include, for example, looking to preexisting ethnographic accounts to gain a better understanding of how social information is exchanged in small-scale societies.

[57] The experimental work of Turner et al. cited in the foregoing paragraph did look at how perceptions of the sharer were affected by relation to the sharer (i.e., whether the sharer was a friend or a stranger). Surprisingly, they found that sharing social information, regardless of whether it was positive or negative, tended to adversely affect perceptions of the sharer for both types of relationships. What I am asking about here is different, though. I am asking how individuals factor in their relationship to their communicative partner(s) in deciding what social information, if any, to share with the latter. A more general question here is how such decision-making is sensitive to the personality traits one's communicative partner(s).

In short, we will need a fuller picture of the factors which enter into individuals decisions about which social information to share and with whom to share it. But in addition to these questions focused on the sharer side of things, we will also need to know how acts of sharing tend to affect the behavior of those with whom social information is shared. In particular:

- If A shares social information with B, is it *actually* the case that B is more likely to share social information with A in the future than if A hadn't shared such information with B?
- If A regularly (though still selectively in some sense) shares social information with his or her group mates, is A *actually* more likely to be chosen as a partner for future activities that payoff for A?

With even partial answers to these questions, we will be in a much better position to evaluate the claim that the disposition to share social information can be explained via individual-selection alone. That is, we will be better able to assess the claim that, for an individual belonging to a social group of the sort our Pleistocene ancestors lived in, the expected benefits to the individual of having such a disposition would have outweighed the expected costs.

Where have we gotten to, then? We began by clarifying Bowles and Gintis' argument from the free-rider identification problem to a group-selection account of the evolution of human cooperation. I claimed that the real crux of their argument is not that this problem is exceedingly difficult to solve in even modestly-sized social groups, but rather that its solution in such groups requires that individuals are disposed to share social

---

[58] I have more in mind here than simply whether the information is positive vs. negative.

information with one another. This is the explanatory challenge which I think Sterelny underestimates. While Sterelny *does* realize that proponents of individual-selection accounts cannot simply take such a disposition for granted, the brief explanation he gives leaves too many important questions unanswered. Crucially, however, this is *not* to say that we should be convinced by Bowles and Gintis' argument; they are unduly skeptical, I think, of the possibility that an adequate individual-selection explanation for the disposition to share social information can be provided. I leave the examination of this possibility for another day.

References

Alexander R (1987) Biology and Human Affairs. University of Washington Press, Seattle
Axelrod R, Hamilton W (1981) The Evolution of Cooperation. Science 211:1390-1396
Baumard N, André J, Sperber D (2013) A mutualistic approach to morality: The evolution of fairness by partner choice. Behav Brain Sci 36:59-78
Bowles S, Gintis H (2011) A Cooperative Species: Human Reciprocity and Its Evolution. Princeton University Press, Princeton
Dunbar R (2004) Gossip in evolutionary perspective. Rev Gen Psychol 8:100-110
Dunbar R, Marriott A, Duncan N (1997) Human Conversational Behavior. Hum Nature-Int Bios 8(3):231-246
Dawkins R (1976) The Selfish Gene. Oxford University Press, Oxford
Emler N (1992) The Truth about Gossip. Soc Psychol Sect Newsl 1:171-193
Emler N (1994) Gossip, Reputation and Social Adaptation. In Goodman R, Ben-Ze'ev A (eds) Good Gossip, University of Kansas Press, Lawrence, pp 119-140
Farley S (2011) Is gossip power? The inverse relationships between gossip, power, and likeability. Eur J Soc Psychol 41:574-579
Foster E, Rosnow R (2006) Gossip and network relationships. In Kirkpatrick D, Duck S, Foley M (eds) Relating difficulty: the processes of constructing and managing difficult interaction. Erlbaum, Mahway
Guerin B, Miyazaki Y (2006) Analyzing rumors, gossip, and urban legends through their conversational properties. Pscyhol Rec 56:23-34
Hames R, Vickers W (1982) Optimal Diet Breadth Theory as a Model to Explain Variability in Amazonian Hunting. Am Ethnol 9(2):358-78
Kurland N, Pelled L (2000) Passing the word: Toward a model of gossip and power in the workplace. Acad Manage Rev 25(2):428-438
Marlowe F (2010) The Hadza Hunter-Gatherers of Tanzania. University of California Press, Berkeley
Marshall L (1976) The !Kung of Nyae Nyae. Harvard University Press, Cambridge

Nowak M, Sigmund K (1998) Evolution of Indirect Reciprocity by Image Scoring. Nature 393:573-577.
Nowak M, Sigmund K (2005) Evolution of indirect reciprocity Nature 437 (7063):1291-1298
Ray V (1963) Primitive Pragmatists: The Modoc Indians of Northern California. University of Washington Press, Seattle
Sperber D (2000) Metarepresentation in an evolutionary perspective. In Sperber D (ed) Metarepresentations: A Multidisciplinary Perspective. Oxford University Press, Oxford, pp 117-137
Sperber D (2001) An evolutionary perspective on testimony and argumentation. Philos Topics 29:401-413
Sterelny K (2012) The Evolved Apprentice: How Evolution Made Humans Unique. MIT Press, Cambridge
Sugden R (1986) The Economics of Rights, Co-operation and Welfare. Basil Blackwell, Oxford
Tanner A (1978) Divinations and Decisions: Multiple Explanations for Algonkion Scapulimancy. In Schwimmer E (ed) The Yearbook of Symbolic Anthropology. McGill-Queens University Press, Montreal, pp 89-101
Trivers R (1971) The Evolution of Reciprocal Altruism. Q Rev Biol 46:35-57
Turner M, Mazur M, Wendel N, Winslow R (2003) Relational Ruin or Social Glue? The Joint Effect of Relationship type and Gossip Valence on liking, Trust, and Expertise. Commun Monogr 70(2):129-141
Wiessner P (2005) Norm Enforcement among the Ju/'hoansi Bushmen: A Case of Strong Reciprocity? Hum Nature-Int Bios 16(2):115-145