"Twisted Tales: Causal Complexity and Cognitive Scientific Explanation" MINDS AND MACHINES 8:1998: p.79-99 Reprinted in F. Keil and R.A.Wilson (eds) EXPLANATION AND COGNITION (MIT Press, 2000) p. 145-166

TWISTED TALES: CAUSAL COMPLEXITY AND COGNITIVE SCIENTIFIC EXPLANATION

Andy Clark

andy”indiana.edu

0. Abstract

Recent work in biology and cognitive science depicts a variety of target phenomena as the products of a tangled web of causal influences. Such influences may include both internal and external factors, as well as complex patterns of reciprocal causal interaction. Such twisted ales are sometimes seen as a threat to explanatory strategies that invoke notions such as "inner programs", "genes for" and sometimes even "internal representations". But the threat, I shall argue, is more apparent than real. Complex causal influence, in and of itself, provides no good reason to reject these familiar explanatory notions. To believe otherwise, I suggest, is generally to commit (at least) one of two seductive errors. The first error is tot think that the general notion of a state x coding for an outcome y involves the state's constituting a full description of y. This is what I call the "myth of self-contained code". The second error is to think that the practice of treating certain factors as special (e.g., seeing genes as coding for outcomes in a way environmental factors do not) depends on the (often mistaken) belief that the singles out factor is somehow doing the most real work. Where the work load is evenly spread, it is assumed there can be no reason to treat one factor in a special way. This is what I term the "Myth of Explanatory Symmetry." Avoiding these errors involves reminding ourselves of 1) the context-dependence of even standard, unproblematic uses if the notions of code, program and information-content, and 2) the difference between explaining why an event occurred and displaying the full workings of a complex causal system.

1. Introduction: Complexity and Explanation

Recent work in biology[1], cognitive science[2] (as well as economics[3], cognitive anthropology[4], and philosophy[5]) displays an increasing sensitivity to what might be termed the problem of *complex causation*. Complex causation obtains when some phenomenon n of interest looks to depend on a much wider and more tangled web of causal influences than we might have hoped or imagined. Thus although all causation is arguably complex, it

is certainly the case that we can discover that, in a given instance (say, the explanation of mature form in certain biological organisms) the relevant causal web is much broader and more multi-factored than it once appeared. Such a web (we shall see examples in Section 2) may actively involve both internal factors (such as genetic influences), external factors (such as environmental influences and basic laws of form), and extended processes of reciprocal interaction in which some factors xxx and are modified by the action of the others[6].

Complex causal webs have, of late, figured in a number of arguments designed to put pressure on familiar explanatory constructs. Target constructs include the notions of "inner programs", "genes for" and "internal representations". Thus Thelen & Smith (1994) argue that action and cognition are not to be explained by reference to genetic blueprints or programs, because they are emergent out of the interactions of multiple forces spanning brain, body and world. As a result "there is order, direction, and structure ... but there is no design [nor] program in the genes" (op.cit., p.xix). Elman et al. (1996) argue, for similar reasons that "it is more useful to view genes as catalysts rather than codes or programs" (op.cit., p. 351), and go on to promote a multi-factor, highly interactionist view as an alternative to the widespread idea that we are born with innate knowledge concerning grammar, physics, theory of mind, and so on (op.cit., p. 357-396). Perhaps most ambitiously of all, the multiple and complex interactive relations that characterize real-world, real-time activity have been seen by some as threatening the spread to internal programs and computations in the cognitive scientific explanation of action. The leading idea here is that "the relation between nervous system and environment is one of influence of dynamics rather than specification of state" (Wheeler (1994)) may afflict the inner organization itself to the extent that it becomes fruitless or impossible to try to see "the causal interactions between...modules as representation-passing communications" (op.cit.,

p. 40). Traditional explanatory models, according to these arguments, underrate the extent to which action is structures by the continuous interplay of multiple (inner and outer) forces. Hence the appeal, to many theorists, of the dynamical systems perspective which depict "complexes of parts or aspects..all evolving in a continuous, simultaneous and mutually determining fashion" (van Gelder & Port (1996), p. 13).

There is much that is true and important in all these claims and arguments. Some of the detailed disputes I comment upon elsewhere -- see, e.g., Clark (1997b) and Clark (to appear). The present focus, however, is much more narrow. I shall examine just one aspect of the arguments viz the (putative) tension between explanations that speak of "genes for," "programs for," "codes for," and so on, and the fact (assuming it is a fact) that specific outcomes depend on a multitude of subtly interacting internal and external factors and forces. The appearance of tension, I shall argue, is largely illusory and is fostered by the (explicit or tactic) acceptance of one or both of the following myths:

Myth One: The Self-Contained Code

If some x is to be properly said to code for, program for, describe or even prescribe some outcome y, then x must constitute a detailed description of y, even when x is considered *independently of its normal ecological backdrop*.

Myth Two: Explanatory Symmetry

If the overall causal web is complex yet x is to be cited as the cause of y, then x must be the factor that does the *most actual work* in bringing it about than y. Causal symmetry, by contrast, implies explanatory symmetry.

Both myths are untrue (that=s why they are myths). But they are pernicious and exert a clear (often explicit -- see sections 4 and 5) force on our thought and argument. The remedy for the myth is, I think, some reflection on 1) the nature of causal explanation -- in particular, on the role of what I shall call the Alocus of plasticity” on our intuition about

causal pathways. And 2) the notion of the *practical information-content* of a message, code or inscription.

The structure of the paper is as follows. In the next section (section 2), I display two examples of the kind of causal complexity at issue. Section 3 then canvasses some of the more challenging responses to the discovery of such complexity. Sections 4 and 5 questions these responses by undermining the twin myths described above. Section 6 is a brief conclusion.

2. Ways of Web-Making

Here are two examples of the kinds of complex causal webs that might confound the unwary explanation-giver. The first is borrowed from Elman et al. (1996) and displays the range of factors and interactions involved in a newly hatched chick=s well-known capacity rapidly to "lock on" to, or "imprint upon" a mother hen. The second is from Thelen & Smith (1994) and involves the development of walking skills in human infants.

Example I: Imprinting in Chicks.

Newly hatched chickens rapidly become attached to the first mobile object they see. This attachment manifests itself as a tendency to follow and attend to the >imprinted= object in preference to all others. In the wild, this process of imprinting leads the chick to attach itself to a mother hen. But in the laboratory, the process can be manipulated so that the chick imprints upon some other mobile object such as a moving ball or cylinder (Johnson (1997), or review in Johnson & Bolhuis (1991)). But how, exactly, does this process work? Is it simply that the chick is "pre-wired" so as to fixate upon the first conspicuous object it sees? That, to be sure, is a fair description of the outcome. But the process itself turns out to involved "interactions at a number of different levels: organism-environment, brain systems, cellular and molecular" (Elman et al. (1996), p. 324).

To begin with, the imprinting process seems to involve two quite independent neural systems. The first (called *Conspec* by Johnson & Morton (1991)) disposes the chick to

prefer stimuli that fir the head and neck configuration of a similar-sized bird and mammals (Johnson & Horn (1988)). The second (a learning systems called IMHV, for "Intermediate and Medial Hyperstriatum Ventrale" and located in that specific region of chick forebrain) develops a representation of the highly-attended object that allows the chick to recognize the object despite variations in spatial location and orientation (Johnson (1997), Ch.4, Elman et al. (1996), Ch.6). The presence of these two distinct contributory neural systems is indicated by, for example, lesion studies that show that damage to IMHV impairs preferences acquired by learning (as when a chick imprints upon a rotating red box) yet does not affect the general predisposition to prefer broadly hen-like stimuli (Johnson & Horn (1986), Johnson (1997), p. 110).

Given a normal ecological backdrop, the Conspec and IMHV systems collectively yield a powerful and robust attachment to a single attended hen. But how do they actually interact? One simple possibility is that the Conspec system acts as a kind of internal filter that selects the training data *seen* by IMHV. Further investigations, however, have shown that such internal filtering is probably not occurring (Johnson & Bolhuis (1991)). Instead, the two systems look to be internally non-communicating. The interaction between them looks to go via a loop that involves the real-world behavior of the whole chick. Conspec, operating in a normal ecological setting, causes the whole organism (the chick) to expose itself to a heavy dose of training inputs targeted in a mother hen. IMHV has, in addition, certain restrictions on the kinds of thing it can learn about. It requires a mobile stimulus of a certain size before it "kicks in”. The combination of Conspec and IMHV, operating against a natural ecological backdrop, thus leads the chick (via a loop out into the attending behavior of the whole organism) to rapidly and robustly develop a translation-invariant representation of a mother hen.

The learning restrictions of IMHV have been speculatively explored using a connectionist model (O'Reilly & Johnson (1994)) in which simple architectural biases yield

the focus on mobile stimuli and explain the fluent acquisition of a translation-invariant representation (one able to pick out an object despite variations in viewing angle and spatial operations). The details need not detain us (they are nicely laid out in Elman et al.(1996), p. 327-333, and Johnson (1997), p. 105-107), but involve the combination of internal positive feedback loops and Hebbian learning. The positive feedback loops caused some units active for an object in position P1 to remain active as the object moved into P2, P3, etc. The associative learning then allows the development of top-level units that respond to the object (the co-occurring set of features) in whatever spatial location it appears. The system thus develops location-invariant object detectors. In real chicks, this learning process is linked to the expression of a particular gene (c-fos -- see McCabe & Horn (1994)), and is thus revealed as itself dependent upon a variety of molecular level interactions.

As a final not if complexity, the Conspec system that forms the other half of the take is not, in fact, active at birth. Instead, it depends on details of early motor activity. To become active, the Conspec system requires the chick to run about freely at least a small period of time between the ages of 12 and 36 hours. Deprived of such motor activity, Conspec lies dormant and the learning system operates alone, without the benefit of the behavior-based input selection mechanism[7].

Summary: Chick imprinting involves the subtle interplay of such diverse factors as: the statistical regularities in the chicks visual experience; the presence of motor activity triggering Conspec; the organism-level behavioral effects of Conspec in operation: the genetic bases of IMHV and Conspec; and the nature of the ecologically normally hatching environment (see Johnson (1997), p. 116, Elman et al. (1996), p. 332). The *simple* phenomenon of filial imprinting in chicks thus turns on a twisted tale in which "multiple sources of constraints, both from within levels and from other levels (molecular, organism-

environment, etc.) Ensure a particular outcome: a spatially invariant representation of the mother hen" (op.cit., p. 332).

Example II: Learning to Walk.

(Here I shall be brief, as I address this case in greater detail elsewhere[8]). Consider the process of learning to walk. This process, in human infants, now looks to involve a complex series of interactions between neural states, the spring-like properties of leg muscles and local environmental factors. This vision (of >soft assembly= -- Thelen & Smith (1994), p. 60) is contrasted with the image of learning to walk as the temporally staged expression of a prior set of instructions encoded in e.g., a genetically specified central pattern generator or neural control system (Thelen & Smith (1994)), p. 8-20, 263-266). In place of a single, privileged, inner-or-genetic cause, Thelen & Smith display a multi-dimensional interaction process in which "the organic components and the context are equally causal and privileged" (op.cit., p 17).

Evidence of the multi-factor view comes from a variety of striking experiments in which (to give just a few examples):

- stepping motions are induced in Anon-stepping" infants by holding the baby upright in warm water
- non-stepping seven month olds help upright upon a motorized treadmill perform coordinated alternating stepping motions (even compensating for twin belts driving each leg at different speeds!)

Such results (see Thelen & Smith (1994), Ch.1 and 4) show that stepping is not under the control of a simple inner variable. Bodily parameters (such as leg weight, which is effectively manipulated by partial immersion in water) and environment factors (such as the presence of the treadmill) are also playing a role. In the case of the treadmill, further experiments revealed that the crucial factor was the orientation of leg and foot to the treadmill. Infants that made flat-foot belt contact exhibited treadmill stepping, whereas

those that made only toe-contact failed to step. Thelen & Smith (op.cit., p. 111-112) hypothesize that the infant leg, when stretched out is acting like a spring. At full back stretch, the spring up coils and swings the leg forward. Flat-foot belt contact may precociously ensure this full back-stretch and hence initiate stepping. Relative flexor or extensor tendencies in the legs thus contribute heavily to the emergence of coordinated stepping (op.cit., p. 113).

Summary: Infant stepping behavior depends upon the precise balance of interplay of a variety of factors including: the weight of the legs; the "relative flexor (very tight) or extensor (more loose) tendencies of the legs" (op.cit., p. 113); and whatever central neural structures are implicated in the motor control process itself. Stepping behavior thus "emerges only when the central elements cooperate with the effectors -- the muscles, joints, tendons -- in the appropriate physical context" (op.cit., p. 113).

3. Webware (Seeds, Catalysts, Modifiers, and Control Parameters)

What kinds of explanatory story should we tell to make best sense of cases involving complex and heterogeneous causal webs? One widespread negative response goes like this: whatever stories we tell, they must not involve the isolation of "privileged elements," or give Aontological priority" to any particular strands in the web (see e.g., Thelen & Smith (1994), p. 17, 580; Elman et al. (1996), Ch.6, van Gelder & Port (1995), p. 13. For the same claims made in a more purely genetic context, see also Kelso (1995), p. 183, Goodwin (1995), p. 119, and Oyama (1985)(1992)).

The belief that no element in the causal web is in any sense privileged rapidly leads to scepticism concerning these types of understanding or model that depict certain elements (be they in the genes or in the actual neural circuitry) as inner programs for the production of certain behaviors. In extreme cases this translates into scepticism concerning the very idea of internal representations[9]. More obviously, it translates into wariness concerning the idea of (usually inner) elements acting as codes, recipes, blueprints, prescriptions,

descriptions, sets of instruments, etc., etc. (See e.g., Thelen & Smith (1994), p. Xix, 9, 33, 83, 112; Elman et al. (1996), p. 350-352).

Such negativity accrues an obligation. How else are we to comprehend the natural genesis of form and behavior? The general tendency, at this point, is to favor accounts that invoke multiple interaction biases and that depict form and behavior as emergent properties of the overall causal mesh. The case of genetic determination provides a nice example. For the image of the gene (or genes) as directly coding for specific morphological or behavioral outcomes is, it is universally accepted, a simplification (at best). Genes (as we will see in more detail in section 5) bring about their effects via an extended sequence of interactive processes. These may include local chemical interactions, basic physical laws governing the emergence of form (see, e.g., Goodwin=s (1995) work in "morphogenetic fields" and the complex interplay between development and environmental factors (e.g., the use of ambient temperature to determine the sex of Mississippi alligators -- Goodwin (1995), p. 38, or the more complex and extended example of the chick imprinting mechanism). In such cases, the relation between the genes and the final product is mediated by multiple types and levels of interaction (see especially Elman et al. (1996), Ch.6). Such mediation, it is argued, works against the notion of the genes as codes, programs, algorithms, descriptions or prescriptions. Instead, we should think of genes as being more like "catalysts" (op.cit., p. 351), "seeds" (Goodwin (1995), p. 16) or "modifiers" (op.cit., p. 144). In support of e.g., the "genes-as-catalysts-not-programs" view it is argued that:

> Programs are (more or less) informationally self-contained. Catalysts, on the other hand, are embedded in an environment of natural laws and processes
>
> Elman et al. (1996), p. 351.

A catalyst, the authors note, is individually inert. Alone, it does nothing. But place it in a certain context (e.g., a vat of chemicals) and it can ensure an outcome that would otherwise not occur. Thus the presence of a gene may produce an enzyme that speeds up a

reaction. The gene does not "define the conditions for reaction" -- that is left to the laws of biochemistry. Instead the genes "harness those laws by ensuring that critical components are present at the right time and then nudging the reaction forward" (both quotes: op.cit., p. 351-352).

Why not count the gene as, if not a full blueprint, at least a program: an algorithm for bringing about a certain effect? The reason given that programs (according to Elman et al.) are informationally self-contained:

> One can examine a program and -- looking only at the code -- make a reasonable guess about what it will do. This is not possible with genetic material. The relationship between DNA nase triples and amino acids may be direct; but the assembly of amino acids into proteins, the timing of when specific genes are expressed and the effect of a gene's products are highly context-sensitive
>
> Elman et al. (1996), p. 351.

Genes, it is argued, are not informationally self-contained. Taken alone, their information content (like that of a catalyst) is zero (op.cit., p. 351). But taken in context, the information content explodes: it becomes "potentially enormous, embracing whatever *information* there is in the environment" (op.cit., p. 351).

This discussion of the *information content* of some part of an extended causal process (in this case, the part is a gene: but that is not essential) is both problematic and revealing. For I believe it displays an important widespread confusion centered on the unreachable grail of *informational self-containment*. Unraveling this confusion is the task of section 4. For the present, however, notice how easily this kind of vision carries over to the more developmental cases rehearsed in the previous section. The neural system Conspec cannot, on its own, account for chick imprinting. But placed in the rich context of the effects of Conspec on whole organism behavior, the learning profile of IMHV and the natural, mother-hen-rich hatching environment, Conspec effectively catalyzed the learning process.

By extension, the genetic bases of Conspec function (via a further series of interactions) to ensure that this bias is present and hence -- via an extended sequence of environment-exploiting interactions -- act to ensure successful imprinting (keeping all the other factors fixed). I show the "dependence on interactions is repeated at successively higher levels of organization" (op.cit., p. 351).

Such dependence on interactions is also at the root of Thelen & Smith=s insistence that (in the stepping example described in section 2) there is

> no essence of locomotion either in the motor cortex on the spinal cord. Indeed, it would be equally credible to assign the essence of walking to the treadmill than to a neural structure...
>
> Thelen & Smith (1994), p. 17.

Much in the spirit of Elman et al.'s notion of the gene as catalyst, Thelen & Smith argue that flexor tone (the relative tightness or *give* in the infant's legs) is acting as a >control parameter= that cuts so as to "engender the shift into stable alternate stepping" (op.cit., p. 112). But importantly,

> as a control parameter, flexor tone constrained the interacting elements but did not prescribe the outcome in a privileged way
>
> Thelen & Smith (1994), p. 112.

Finally, consider Elman et al.'s (1996) argument against the idea that innate knowledge underpins our capacities to rapidly learn about grammar, physics, other minds and so on. In briefest outline[10], the argument is that nature looks to rely not on detailed pre-specifications of "fine-grained patterns of cortical connectivity" (op.cit., p. 360) but on the provision of a variety of simpler biases involving architecture (neuron types, numbers of layers, connectivity between whole brain regions) and timing (waves of synaptic growth and loss, relative development of sensory systems, etc.). (See table 1.3, op.cit., p. 35). These biases lead, in environmental and developmental context to the organisms exhibiting

specific skills, forms and behaviors, including, for example, the robust acquisition of grammatical knowledge. We are thus innately predisposed to learn a grammar, but in a way that falls short (it is arguable of requiring the innate prespecification of actual grammatical knowledge. Instead, constraints at the levels of timing and architecture, in collaboration with environmental survival, inexorably nudge the system towards the target knowledge. In such a case, we are told, "the knowledge itself ... would not be innate and would require appropriate interactions to develop" (op.cit., p. 364).

All these arguments and assertions demand attention in their own right. They all share in virtue of drawing our attention in the sheer complexity and heterogeneity of the causal webs that underlie various phenomena of scientific interest. And in specific claims and conclusions displayed all innate consideration of a host of pertinent issues, both for and against. In the context of the present project, however, I want to focus attention on just one common thread: the tendency to cite causal complexity and the important role of repeated interactions as a reason to eschew talk of specific states or items as prescribing, programming or coding for specific outcomes. Call this the ;inference to egalitarianism=. I believe this inference to be false, and for two fairly deep reasons. The reasons center first (section 4 following) on the problematic notions of self-containment and information-content, and second (section 5 following) on the difference between invoking a cause and unpacking the workings of a complex system.

4. The Myth of the Self-Contained Code

The first reason to be wary (of the inference to egalitarianism) concerns the putative contrast between genuine programs (codes, recipes, etc.) And factors that bring about effects only in the context of a rich backdrop of the contributory processes and interactions. The contrast is explicit in Elman et al/ (1996, p. 351) characterization of programs as being A(more or less) informationally self-contained."This claim, as far as I can see, is simply false. Program, in any ordinary sense of the word, is far from being a self-contained

repository of all the information necessary to solve a problem. Think, for example, of a standard program written in a language such as LISP. LISP, as we all know, is a List Processing Language. That means you can do things such as store a list (say (abc)) then add new items using operators such as cons (concatenate). The input (cons d (abc)) adds d to the head of the list yielding (dabc). You can also use functions such as (first) and (rest) to remove items from lists[12].

The point to notice is just that the operation of these functions -- upon which the success of just about any LISP program depends -- is by no stretch of the imagination even >more or less= given as part of any actual program written in LISP. Instead, like the operating system firmware -- the function work due to the *ecologically normal* backdrop against which a LISP program brings about its effects. The program -- at least as we commonly use the term -- does not itself specify exactly how to bring about these effects. Instead, (to put it in the kind of terminology used for the cases examined earlier) it constitutes just one factor which -- in the special context of a computing device set up to compile or interpret such programs -- will reliably lead the overall system to discover a solution to the target problem.

Ordinary computer programs are thus not informatically self-contained. So the fact that the genes (for example) do not contain all the information needed to describe a biological organism cannot (in and of itself) constitute a reason to reject talk of genes as programming for certain traits, behaviors or outcomes. Likewise, the fact that neural events are just one factor amongst many whose combined activity yields stepping behavior cannot (in and of itself) constitute a reason for rejecting the idea of motor programs. In each case, the factor invoked (genes or motor programs) may be regarded as coding for a specific outcome *on the assumption* that such ecologically normal backdrop prevails.

This point is forcefully made by Dennett (1995) in a discussion of the complexities of the genome-organism relation. Dennett notes that DNA constitutes a most indirect manner

of *instructing* the process of building a phenotypic body. For much of the necessary information is not given in the DNA itself but only in the combination of DNA and a set of environmental conditions. But, Dennett argues, even in the case of a library (universally accepted as being a *storehouse of information*) it is "really only libraries-plus-readers that preserve and store information" (op.cit., p. 197). Likewise DNA codes for organismic features only in the context of an environment capable of >reading= the DNA. The code can do its work only against a certain backdrop. To take a homely example:

> every time you make sure that your dishrag gets properly dry in between uses, you break the chain of environmental continuity (e.g., loss of moisture) that is part of the informational background presupposed by the DNA of the bacteria in the dishrag whose demise you seek
>
> Dennett (1995), p. 197.

The DNA codes for specific outcomes only in a context that includes both reliable local chemical reactions and wider environmental contingencies (such as moisture). Without this extended *reading system* DNA sequences, Dennett notes, "don't specify anything at all." Yet this rampant presumptiveness should not, he argues, prohibit us from speaking of e.g., genes for x. For the gene (or genes) may be "for x" in the simple sense that it is a feature whose presence or absence is a difference that makes a systematic (usually population level) difference to the presence (or absence) of x[13]. We will return to this point in section 5 below.

What then, of the notion of informational self-containment itself? We are, I think, quite properly pulled in two directions. On the one hand, we might like to say that by keeping a certain ecological backdrop constant, we can legitimately speak of information about biological form being given in the DNA. This, after all, is no worse than supposing that the books in the library (keeping the human reader constant) contain information about architecture, plumbing, etc. Nor is it worse than saying that a certain LISP program

contains the information needed to solve a given problem. On the other hand, we should not thereby be blinded to the large extent to which the finished product depends on a wider variety of other factors and forces. It is in this sense that, for example, the quantity of information encoded in the genome falls spectacularly short of what would be needed to describe the organism. These is, in short, a conflict between the simple, quantitative measures of information used in information theory and the effective information content that can be carried by a force or structure able to piggyback upon (or assume) a certain reading system or a certain context of effect. The apparent mismatch between quantitative information-theory and semantics is, of course, well-known. What is emerging here is the extent to which those mismatch may be rooted in the way some bearers of content (such as messages) trade on assumptions concerning contexts and readers.

Cohen & Stewart (1994, p. 353) drive this home using a simple thought experiment. Suppose you are told that "If I don't phone you tonight, Aunt Gertie will be arriving on the 4:10 train from Chattanooga. Take her home." That evening, you receive no phone call. The null event (of your not receiving a call) "conveys a sizable quantity of information with a zero-bit message." Maybe, the authors note, we really have a one-bit message here (one on/off choice). But the xxx is unaffected: a complex set of events is reliably set in motion by a sparse signal -- a signal that nonetheless effectively conveys a rich content. By contrast, a bare television screen caption that reads "call 1-800-666-7777" conveys an effective content comprising just 36 bits of information (11 decimal digits). Yet the information-theoretic measure of the television signal us very much higher, as such a signal must specify the activity of 100 lines each involving 1000 phosphate dots and capable of exhibiting these different colors. The signal this constitutes (from this perspective) an 800,000 bit message -- see Cohen & Stewart (1994), p. 353.

We are then lead to a contrast between (what I am calling) the "effective content" of a message and its information-theoretic measure. Effective content (as in the case of the null

telephone message) is revealed as a thoroughly context-dependent phenomenon and one that depends on somehow "triggering" the access of information from a specific range of possibilities" (op.cit., p.353). Information about this range of possibilities lies not in the triggering signal but in the receiver, reader or the environment in which it has its effects. This we have now seen, is true not just of DNA[14] and neural structures but of words in library books, standard LISP programs -- in fact, just about every case where we would standardly talk of one set of items as *coding for* something else. Cohen & Stewart sum it up well:

> the meaning in a language does not reside in the code...[it] stems from the existence of a shared context. For language, the context is the culture shared by those who speak that language. For the DNA message, the context is biological development...all messages in the real world that really are messages happen within a context. That context may be evolutionary, chemical, biological, neurological, linguistic or technological, but it transforms the question of information-content beyond measure...
>
> Cohen & Stewart (1994), p. 354-5.

The observation that chemical factors and rich environmental interactions (etc.) Play a crucial role in bringing about certain effects thus cannot (in and of itself) constitute a good reason to reject the image of genes or inner neural structure as coding for, prescribing, or programming those effects. For rich context-dependence is always the rule, even in mundane and unproblematic uses of the notions of program, code and message. The putative contrast with a fully context-independent way of embodying meaning is misguided: the self-contained code is a myth.

5. The Myth of Explanatory Symmetry

The inference to egalitarianism has, however, a second string to its bow. For in designating some factor x as coding for, or programming, an outcome y, we are treating x

as somehow special. For we want to say that x codes for y whereas the ecological backdrop provides the *reading environment* relative to which x bears its effective content. But whence this asymmetry? Could we not equally well depict the environmental factors as coding for y and the other factor (be it genetic or neural) as the backdrop against which these bear the effective contents they do? At which point the whole value of treating one type of factor as coding for or programming the outcome looks to be called into question. Why not just admit, in that case, that we confront a complex causal web whose combined activity yields the outcome, seek to understand as much as we can of the web itself and leave it at that? Such I think, is the thrust of Thelen & Smith's injunctions against "privileged elements" and of Elman et al.'s suggestion that we focus attention not on components but on the "complex web of interactions" (op.cit., p. 321). It is also the explicit moral of Oyama's influential (1985) work on the explanation of biological form, which claims that we must give up the practice of assigning priority to either internal or external forces and instead focus on the interactions themselves as primary objects of study.

I must tread gently here, for I believe that there is something overwhelmingly *right* about these ideas and structures. If we want to understand how the outcome comes about, the proper explanatory strategy is indeed to confront the complex interactive process as a whole. In the course of such a confrontation we may sometimes discover that in terms of actual work done (measured as the degree of control exerted over the final product) the factors that I have been lumping together as the "ecological backdrop" in fact carry the bulk of the explanatory burden. This might be the case if, for example, the production of a certain biological form is heavily determined by basic laws of physics and chemistry and the genetic material simple "seeds" the process (see, e.g., Goodwin (1994) on morphogenesis, or Kauffman (1993)).

But our explanatory attention is not always limited to the project of understanding how the effects come about. Sometimes, at least, we seek to understand why they come

about. And it is here that we may begin to break the apparent causal symmetry that would depict all factors on an essentially even footing.

Thus consider a paradigmatic case of genetic disease: phenylketonuria[15]. This disease (known as PKU disease) causes mental retardation, shortness of stature and lack of pigment (see Gifford (1990), p. 333). Here is how it works:

> the normal gene at the PKU locus produces the liver enzyme phenylalanine hydroxylase, which is requires for the metabolism of the amino acid phenylalanine into tyrosine. Individuals homozygous for the PKU gene cannot produce this enzyme. If one=s diet contains the normal amount of phenylalanine, the serum level of phenylalanine rises dramatically. This interferes with the production of myelin, the protective sheath of nerve cells in the brain. *But these effects are avoided if a diet low in phenylalanine is proved and this is what is done for therapy*
>
> Gifford (1990), p. 333 (my emphasis).

Gifford notes the interesting consequence: this is a disease which can be avoided or cured by a simple environmental manipulation. The disease is a joint effect of the abnormal gene and the diet. But PKU disease s classed as a paradigmatic case of a genetic problem. Why? Gifford's suggestion (one endorsed in various forms by both Dawkins (1982, p. 23) and Dennett (1995, p. 116)) is that we are thereby drawing attention to the fact that the diet is a common factor in the base population, whereas the PKU gene is not. Relative to the base population, it is the gene that *makes the difference* (Dennett (1995), p. 116), even though the workload (the causal etiology of the disease) is spread between genetic and environmental factors, and even though the outcome is this fully manipulable by non-genetic means. Gifford this proposes that:

> (DF) A trait is genetic (with respect to population P) if it is genetic factors which :make the difference" between those individuals with the trait and the rest of population P[16].

The answer to the "why" question (why did that person develop PKU disease?) Thus isolates the genetic factors as especially relevant. But the answer to the "how" question (how does PKU disease arise) implicates genetic and environmental factors pretty well on even footing.

The cost of this maneuver is clear enough. Change the normal environmental conditions and what was once a genetic disease becomes an environmentally induced problem. This is because the "why" question is always framed against a fixed background. Gifford thus noted (following Burian (1981)) that in the hypothetical case of a population whose normal diet (unlike our own) is low in phenylalanine, the very same causal story would be classed as a case of environmentally induced disease. For the locus of relevant plasticity (as I shall say) here lies not in the genes but in the diet: it would be those (rare) individuals who are both homozygous for the PKU gene and consume high amounts of phenylalanine that fall ill, whilst the genetic factors alone (being homozygous for the PKU gene) would not normally lead -- in that population -- to the development of the disease. What counts as genetic this depends "not only on the causal processes in the individual, but also on a fact external to this: the causal factors shared in the population" (Gifford (1994), p. 334). Such relativity to a contextual baseline is, however, exactly what we should expect given our earlier discussion of the close relation between effective content and an assumed ecological backdrop. The context-relativity in no way impurges the correctness (relative to the actual population and environment) of singling out the PKU gene as especially relevant in the production of the disease. What we must not do, of course, is allow this fact to blind us to the full causal picture and hence to the full causal picture and hence to the possibility of an environmental cure for the disease itself.

Explanatory priority (in a given context) is thus turns not on what factor (if any) does the greatest amount of actual work but in where we should look for the *differences that make the difference* between the cases where the outcome obtains and those where it does

not. This is the natural explanatory concomitant of the idea (section 4 above) of detailed effective contents being conveyed by simple (but context-exploiting) physical transactions. In the genetic case, we can take this a step further by noticing that genetic material is naturally *designed* to function as a primary locus of plasticity -- it is the *natural* function of the genetic material to be the kind of difference that (relative to an assumed ecological backdrop) makes a specific organism-level difference. In this vein Sterelny (1995) argues that the genome *represents* developmental outcomes because it is its evolved function to bring about those outcomes. The fact that this bringing about involves multiple gene-environment interacting does not undermine the description of the genome as a representation because "representation depends not on correlation but function" (op.cit., p. 165). The correlations may be messy and indirect. But the function shines through, and is the source of the explanatory asymmetry between genome and environment. Both factors correlate equally with developmental outcomes, but they play asymmetric roles. For example;

> snow guns have a different growth pattern in environments in which they are exposed to wind and snow. Both the triggering environmental and the snow gum genome are necessary for the guns response to climatic adversity. But one element of the developmental matrix -- the genome -- exists only because of its role in the production of the plant phenotype. That is why it has the function of producing that phenotype and hence why it represents that phenotype. So an informational idea of a replicate can preserved
>
> Sterelny (1995), p. 165.

The extension of the line on explanatory priority to the case of neural codes and programs is immediate. Here too we should say that a neural structure or process x codes for a behavioral outcome y, if against a normal ecological backdrop, it is the difference that makes the difference with respect to the obtaining of y. A neural event may thus code for a

behavior (say, reaching out an arm) even if the outcome depends equally upon a variety of bodily and environmental factors such as the force of gravity and the spring-like qualities of arm muscles. For such factors are the ecologically normal backdrop against which the neural state was selected to bring about its effects[17].

Notice, finally that this criterion does not simply beg the question in favor of inner or genetic states. Instead, it invites us to keep constant the stabilities and features of the normal ecological backdrop and to focus attention (for the purposes of answering the why question) on the locus of plasticity: the place to which differences in the outcome (in the normal context) are best referred. As bare biological brains increasingly parasitize and exploit the environment as a kind of extended information-processing resource[18], the location of this primary explanatory locus may sometimes tend to shift outwards. Such complexities, however, are best left for another occasion[19].

The observation that the real workload involved in bringing about some effect may be evenly spread between allegedly "privileged" factors (such as genes and neural events) and other influence (environmental, chemical, bodily) cannot, I conclude, in and of itself, constitute a good reason to reject the practice of treating certain factors as special: as coding for, programming, or prescribing the outcome in question. It cannot do so because the relevant asymmetry lies *not in the causal chain itself* but in the extent to which difference in respect of that outcome *within a baseline population and ecological setting* may be traced to difference in the privileged item. If our goal is to explain those observed differences, we may properly single out a few threads in the complex causal weave. If our project is to understand exactly how the outcome is produced, we may attend instead to the full intricacies of the woven whole[20].

6. Conclusion: Living in Complexity

Life is terrifyingly complex. Things interrelate in deep and often desperately confusing ways. Yet adrift in this dizzying whirlpool of causal flow, we heroically succeed

in *making things happen*. When we do so, it is not because we are the self-contained repository of the desired outcome. Nor is it (usually) because we command a detailed description of how to manipulate all the causal chains that link us to our goal. Instead, it is because our strategies have been learned and tuned against a backdrop of culture and physical and social laws and practices. Our strategies take this complex backdrop for granted and manipulate the flow of events by piggybacking upon these unremarked currents in the causal nexus.

In this one respect, at least, life, words, programs and genes are all fellow travelers. They all bring about their effects by working within a complex and extended causal fabric. It is the distinctive virtue of much recent work in biology, anthropology, and cognitive science[21] to begin to recognize the extent and impact of this causal complexity and heterogeneity. Such recognition, however, should not be seen as a threat to explanatory strategies that invoke notions such as coding for, programming, or prescribing specific behavioral or morphological outcomes. The illusion of such a threat is, I have argued, linked to the explicit or tactic endorsement of two (interrelated) myths. The first s the myth of the self-contained code: the belief that to really code for (or program, or prescribe) an outcome an entity must contain, within itself, a detailed description of the outcome. This myth is flatly incompatible with any normal use of the notions of program, code and message. The second is the myth of explanatory symmetry: the belief that the practice of treating certain causal threads using the special terms of codes, programs, and contents cannot be justified if the actual workload is evenly spread between a wide variety of factors and forces. This belief fails, however, to allow for the fact that our explanation-giving practice often involve not the simple measurement of causal work but the (context-and-backdrop relative) assessment of the locus of differentiation. We judge, that is, that observed differences are best explained by keeping a certain background fixed[22] and asking what differences then make the difference among the ordinary population. Causal equality

at one level (the level of *work done*) may thus co-exist with genuine asymmetry at another level (the level of greatest relevant plasticity). Teleological approaches (such as Sterelny's story about the snow gum) add a further consideration viz, that nay privileged locus play the special functional role of existing so s to bring about its characteristics effects. The point, in both cases, is that causal equality need not imply explanatory symmetry.

Puncturing the twin myths blocks any direct[23] inference from facts about causal complexity to the rejection of notions such as inner codes, programs, instruction or prescriptions. It also casts doubt on arguments against innate knowledge[24] that depend on contrasting highly interaction-dependent phenomena with self-contained storehouses of domain-specific information. For it suggests that the basic notion of a state=s bearing a specific effective content is fully compatible with the need to place the state in a rich ecologically context: a context that acts as the assumed backdrop of the original encoding. The same point, substituting :internal representation" for "innate knowledge," can be made against recent attempts to stress organism-environment interactions in (apparent) opposition to reliance on internal represetnations[25].

Moving even further a field, the present treatment may perhaps suggest a somewhat conciliatory angle on the interalism/externalism debate in the philosophy of mind[26]. For a purely inner state may be said to bear a certain effective content, even thought the actual causal chain which determines what that content is now extends far outside the agent=s head. The content is thus referred to the inner state, but its true physical vehicle involves a wide range of additional environmental structures and circumstances. The question of where to locate the "supervenient base" for the content thus admits no straightforward answer. The correct diagnosis is just that the inner state itself bears the effective content but in a way that cannot help but assume an extended ecological backdrop[27].

There is much that remains unclear and problematic in all these debates and I do not claim to have done much more than scratch the surface here. The cash value of the

enterprise is perhaps this: that it underlines just how badly we still understand the apparently foundational notion of the information-content of a physical state and how very hard it is to take ecological context as seriously as we surely must. Yet it is in the balance of these slippery factors that mind finds its place in the natural world. Like cheap detectives, we follow gingerly in its wake.

Notes

1. See Oyama (1985), Johnston (1988), Gifford (1990), Bray (1992), and Goodwin (1995).

2. Some examples include Thelen & Smith=s (1994) treatment of child development, Elman et al.=s (1996) account of innateness, work in dynamical systems theory (Kelso (1994), essays in Port & van Gelder (1995), and the large explosion of work on emergent phenomena and artificial life (this is a massive field, but some good starting points include Steels (1994), Maes (1994), Resnick (1994), and essays in Boden (1996)).

3. Especially Arthur (1990) -- see also Clark (1997a).

4. See Hucthins (1995).

5. Useful discussions with a philosophical slant include Varela, Thompson & Rosch (1991), essays in Griffiths (1992), van Gelder (1995), van Gelder & Port (1995), Sterelny (1995), Dennett (1995), Ch. 5 & 8, Godfrey-Smith (1996) and Clark (1997b).

6. This notion of Acircular causation" is described in Varela et al. (1991), van Gelder & Port (1995) and Clark (1997b).

7. It is speculated that the motor activity increases testosterone levels that I turn activate Conspec -- see Horn (1985), Elman et al. (1996), p. 326.

8. Clark (to appear).

9. I shall not dwell on such cases here. For discussion, see Clark & Toribio (1994), Clark (1997a) (to appear).

10. For a fuller treatment, see Clark (in progress).

11. One worry at this point is that there is a danger of confusing the (clearly correct) observation that the nature grammatical knowledge was not fully specified in advance of learning and the (more contentious) claim that the innate endowment involves no grammatical knowledge (properly so called) whatsoever. It is not part of the present project to engage the argument at that level. But see Clark (in progress).

12. See e.g., Franklin (1995), p. 151, or any LISP textbook.

13. See Dennett (1995), p. 116-117, Dawkins (1982), p. 23.

14. For some genetic examples, that parallel the case of the null telephone message. See Cohen & Stewart (op.cit.), p. 354.

15. The following text leans heavily in the account of PKU disease development in Gifford (1990), p. 332-335.

16. Gifford does not rest solely with the DF criterion, but also adds a criterion of Aproper individuation." These complexities and xxx are treated in depth in Gifford (1990).

17. I pursue this case in detail in Clark (to appear) using the idea of a partial program: a notion that aims to do justice both to the intuition that effective content trades heavily on assumed context and that the work directly specified by the neural command may often be substantially less than we had imagined.

18. See Hutchins 91995), Dennett (1995), Ch. 12 & 13, Kirsh & Maglio (1994), Clark (1997b), Ch. 9 & 10.

19. For a few stabs, see Clark (1997b), Ch. 10, Clark & Chalmers (submitted).

20. Elman et al. (1996) are pretty clearly engaged in precisely this latter project. It is not so clear, however, that the ambitions of those who postulate certain forms of innate knowledge are the same. It is for this reason, I believe, that some of the stringent criticisms leveled by Elman et al. may tend to miss their mark.

21. E.g., Goodwin (1994), Hutchins (1996), Elman et al. (1996), Thelen & Smith (1994), Clark (1997b).

22. This is, of course, related to the old idea of a contrast-class underlying causal explanations. For a useful discussion, with a special focus on >why-questions= see Van Fraassen (1980), Ch.5.

23. I add this caveat because I believe axxx other arguments, outwit the scope of the present treatment, that do indeed cause trouble for our familiar explanatory styles. See Clark (1997b) (to appear).

24. For a detailed treatment of this case, see Clark (to appear).

25. See Clark (to appear).

26. See e.g., Putnam (1975), Burge (1979).

27. Dennett ((1995), p. 409-412) develops an account that looks similar to this. For more on the notion of an extended supervenient base, see Clark & Chalmers (submitted).

References

Arthur, B. (1990). Positive Feedbacks in the Economy. *Scientific American, 92*(February 1990), 99.

Boden, M. (Ed.). (19xx). *The Philosophy of Artificial Life* . Oxford: Oxford University Press.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy, 4,* 73-122.

Burian, R. (1981). Sociobiology and Genetic Determinism. *Philosophical Forum, 13,* 43-66.

Clark, A. (1996). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (1997a). Economic Reason: The Interplay of Individual Learning and External Structure. In J. Drobak (Ed.), *The Frontiers of the New Institutional Economics* . New Jersey: Academic Press.

Clark, A. (in progress). Re-Thinking Innateness: A Critical Study. .

Clark, A. (to appear). The Dynamical Challenge. *Cognitive Science*.

Clark, A., & Chalmers, D. (submitted). *The Extended Mind* (PNP Research Report No.). Washington University.

Clark, A., & Toribio, J. (1994). Doing Without Representing? *Synthese*.

Cohen, J., & Stewart, I. (1994). *The Collapse of Chaos*. London: Penguin.

Dawkins, R. (1982). *The Extended Phenotype*. Oxford: Oxford University Press.

Dennett, D. (1995). *Darwin's Dangerous Idea*. New York: Simon & Schuster.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (Eds.). (1996). .

Franklin, S. (1995). *Artificial Minds*. Cambridge, MA: MIT Press.

Gifford, F. (1990). Genetic Traits. *Biology and Philosophy, 5,* 327-347.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Goodwin, B. (1995). *How the Leopard Changed its Spots*. London: Phoenix.

Gray, R. (1992). The Death of the Gene. In P. Griffith (Ed.), *Trees of Life: Essays in the Philosophy of Biology* (pp. 165-210). Dordrecht: Kluwer.

Griffiths, P. (Ed.). (1992). *Trees of Life: Essays in the Philosophy of Biology* . Dordrecht: Kluwer.

Horn, G. (1985). *Memory, Imprinting and the Brain*. Oxford: Clarendon Press.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Johnson, M. (1997). *Developmental Cognitive Neuroscience*. Oxford: Blackwell.

Johnson, M., & Bolhuis, J. (199x). Imprinting Predispositions and Filial Preference in the Chick. In R. Andrew (Ed.), *Neural and Behavioral Plasticity* (pp. 133-156). Oxford: Oxford University Press.

Johnson, M., & Horn, G. (1986). Dissociation of Recognition Memory and Associative Learning by a Restricted Lesion of the Chick Forebrain. *Neuropsychology, 24*, 329-340.

Johnson, M., & Horn, G. (1988). The Development of Filial preferences in the Dark-Reared Child. *Animal Behavior, 36*, 675-83.

Johnson, M., & Morton, J. (1991). *Biology and Cognitive Development: The Case of Face Recognition*. Oxford: Blackwell.

Johnston, T. (1988). Developmental Explanation and the Ontogeny of Birdsong: Nature/Nurture Redux. *Behavioral & Brain Sciences, 11,* 617-663.

Kauffman, S. (1993). *The Origins of Order*. Oxford: Oxford University Press.

Kelso (19xx). .

Kirsh, D., & Maglio, P. (1995). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science, 18*, 513-549.

Maes, P. (1994). Modeling Adaptive Autonomous Agents. *Artificial Life*, *1*(1/2), 135-162.

McCabe, B., & Horn, G. (1994). Learning-related Changes in Fos-like Immunoreactivity in the Chick Forebrain After Imprinting. *Proceedings of the National Academy of Sciences of the USA*, *91*, 11417-11421.

O'Reilly, R., & Johnson, M. (1994). Object Recognition and Sensitive Periods: a Computational Analysis of Visual Imprinting. *Neural Computation*, *6*, 357-390.

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

Port, R., & Gelder, T. V. (Eds.). (1995). *Mind as Motion: Dynamics, Behavior, and Cognition* . Cambridge, MA: MIT Press.

Putnam, H. (1975). The meaning of "meaning". In H. Putnam (Ed.), *Mind, Language and Reality* (pp. 215-275). Cambridge: Cambridge University Press.

Resnick, M. (1994). *Turtles, Termites and Traffic Jams: Explorations in Massively Parallel Microworlds*. Cambridge, MA: MIT Press.

Steels, L. (1994). The Artificial Life Roots of Artificial Intelligence. *Artificial Life*, *1*(1/2), 75-110.

Sterelny, K. (1995). Understanding Life: Recent Work in Philosophy of Biology. *British Journal for the Philosophy of Science*, *46*(2), 155-183.

Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Clarendon Press.

van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, *XCII*(7), 345-381.

van Gelder, T., & Port, R. (1995). It's About Time: An Overview of the Dynamical Approach to Cognition. In R. Port & T. v. Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 1-44). Cambridge, MA: MIT Press.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.