Connexions

An online journal of cognitive science

ISSN 1368-3233

EDITORS: GAVIN BOYCE, TOM DICKINS, KEITH FRANKISH, AND ADAM HUFFMAN

Issue 3 January 1998

Contents

Sponsored by the Hang Seng Centre for Cognitive Studies

Copyright © 1998 Jill Boucher, Gavin Boyce, Seth Bullock, Peter Carruthers, Andy Clark, T. E. Dickins, Adam Morton

Extract from "Twisted Tales: Causal Complexity and Cognitive Scientific Explanation" (forthcoming)

Andy Clark
Dept. of Philosophy
Washington University at St Louis

Q/ Should we think of gen es as coding for specific developmental outcomes?

Relatedly (we'll see why later), Should we think of inner neural structures as programming for specific behaviours?

There is of late a flood of scepticism about such notions. See e.g. Thelen and Smith (1994), Elman et al (1996), Oyama (1985).

And there are indeed many reasons to be doubtful about such notions, most of which I won't touch on here. But here is one consideration that seems to be adding fuel to the fire, but which (I shall argue) should not move us:

ARGUMENT FROM COMPLEXITY (to be rejected)

Many outcomes (both behavioural and developmental) depend on a complex and extended interplay of factors, both internal and external. The argument from complexity trades on two myths.

Myth 1: 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 2: 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 that Y. Causal symmetry, by contrast, implies explanatory symmetry.

To see the sort of thing that I think is going wrong here, lets focus on myth number 1. In this vein, Elman et. al. argue that genes should be seen as catalysts not codes or programs because:

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).

But this claim, as far as I can see, is simply false. A 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 (a b c)) then add new items using operators such as cons (concatenate). The input (cons d (a b c)) adds d to the head of the list, yielding (d a b c). You can also use functions such as (first) and (rest) to remove items from lists.

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 functions 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 those effects. Instead it constitutes just one factor which, in the special context of a computing device set up to compile and interpret such programs, will reliably lead the overall system to discover a solution to the target problem.

Ordinary computer programs are thus not informationally 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 involved (genes or motor programs) may be regarded as coding for a specific outcome *on the assumption* that some ecologically normal backdrop prevails.

Dennett (1995), p. 196-197 makes essentially this point about the genome-organism relation. Like Dennett, I want to flag the issue of informational self-containment. I suspect it is a major red herring in our debates, since real information is *never* thus self-contained. I'm trying to work this all up into something bigger, so any feedback at this point would be especially useful.

References

Dennett, D (1985): *Darwin's Dangerous Idea* (Simon and Schuster).

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K (1996): *Rethinking Innateness* (MIT Press)

Oyama, S (1985): *The Ontogeny of Information* (C.U.P.)

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

In reply to the extract from Andy Clark's "Twisted Tales: Causal Complexity and Cognitive Scientific Explanation"

On Metaphorising and Care less Talk

Jill Boucher
Dept. of Human Communication Sciences
University of Sheffield

Clark raises the question as to whether or not it is appropriate to 'think about' genes as 'coding', 'programming', 'prescribing' etc specific developmental outcomes. He also raises the related question concerning whether or not it is appropriate to think of inner neural structures as programming for specific behaviours. He suggests that although there may be many reasons to be doubtful about thinking of genes as coding, etc., the argument from complexity is not one of them. He then goes on suggest that the argument from complexity trades on two myths: the myth of the 'Self-Contained Code' and the myth of 'Explanatory Symmetry'.

According to the myth of the 'self-contained code', if X is a program for outcome Y, then X must constitute a detailed description of Y even when X is considered independently of its normal ecological backdrop. In other words, there is a requirement that the program fully specifies the outcome. Clark argues against this criterion for what can properly be said to constitute a code, or a program, first by pointing out that a standard computer program written in LISP 'is far from being a self-contained repository of all the information necessary to solve a problem.' Similarly, he argues that the motor program involved in the infant stepping response is only one factor underlying the response. In other words, he accepts causal complexity but does not see that as inconsistent with talking about genes programming developmental outcomes.

With regard to the myth of 'explanatory symmetry', Clark appears to be making a similar argument: namely that a behavioural outcome may be multifactorially determined, and the role of genetic factors may be equalled by that of other factors, but that this is not an argument against proposing that genes code for behavioural outcomes.

I am perfectly happy with the logic of Clark's arguments. However, I notice that the arguments involve some slight shifting of ground. Most signficantly, in the case of stepping Clark no longer claims that the genes code for the behavioural outcome itself, but only for the 'neural events' or 'motor programs' which contribute to the behaviour. However, he does still want to claim that a motor program can code for a specific outcome given a normal environmental context. Similarly, and less importantly, he is presumably happy to speak of genetic programs as a (contributory) cause of certain developmental outcomes, rather than as the (sole, or only significant) cause. So Clark's logic leads him to claim that in the case of steppinggenes can code for neural events which constitute motor programs which are a contributory cause of a specific behavioural outcome. This quite cautious and complex claim is, however, far from the claim that it is appropriate to 'talk of genes as programming for certain traits, behaviours, or outcomes', a claim which Clark actually reiterates in the penultimate paragraph of his piece. So I think he misrepresents the proper end point of his argument: in arguing against one type of complexity he walks into another, but doesn't appear to notice it.

The complexity which actually undermines talk of genes coding for behavioural outcomes (as opposed to other types of developmental outcomes, such as eye colour, onset of puberty, or normal aging), stems, in my view, not so much from the fact that behavioural outcomes have complex causes which necessarily include environmental factors, but rather from the indirect links between genes and behavioural outcomes. Genes, together with environmental factors, contribute to the formation of brain and body stuff over time. That is the sole function of genes: they help make the hardware. However, the brain and body stuff has certain properties and propensities to function in certain ways, in response to endogenous or exogenous stimuli. To the extent that genes contribute to the properties and propensities of the physical body and brain, the genes may be said to contribute indirectly to behavioural outcomes.

In the case of a behavioural trait, for example 'aggressiveness', the route is relatively direct, since aggressiveness is known to correlate with certain characteristics of brain chemistry, and one may assume that the genes play some role in determining levels of particular substances in the brain. In the case of even a simple reflex such as infant stepping, however, the route is less direct, as Clark himself makes clear. The genes presumably play a critical role (since the stepping response is universal) in determining that a quite rigidly prescribed pattern of neural events occurs in response to a certain range of stimuli experienced in certain contexts (what the ethologists would call a fixed action pattern). The neural events produce a co-ordinated set of motor behaviours which we describe as 'stepping'. So genes have a relatively direct effect on neural events and, via them, contribute to stepping. In the case of more complex behaviours, such as, for example, mate selection, it seems likely that what we, as observers, perceive and describe as a single adaptive function is achieved via a multiplicity of contributory behaviours (Hendriks-Jansen, 1996) some of which are primarily biologically determined (eg. level of sex drive, sensitivity to certain sensory cues) and some of which are learned either through experience or through language. Genes would seem likely to play a more direct causal role in the former set of contributory behaviours than in the latter. The route of causal links between genes and the learned behaviours involved in mate selection is likely to be exceedingly complex and essentially impossible to trace. This does not mean to say that genes will not influence learned aspects of mate selection, but rather to underline the fact that to say the learned aspects of mate selection behaviour are partly genetically determined would involve an act of faith rather than a reasoned argument.

In addition to suspecting that there are only a very few, rather primitive behaviours (like stepping; or an aggressive trait) for which a route between genes and behaviour can be confidently identified, I have a more fundamental objection to Clark's approach. However, I am aware that Clark accepts that 'there are many reasons to be doubtful about (thinking of genes/neural structures as coding for specific developmental outcomes)', and he may well share my second set of reservations. So I will mention them briefly.

In brief, I don't accept the assumption that using the language of computers to talk about the human brain, mind and behaviour is always useful. In fact I think it is counterproductive if we don't constantly remind ourselves that the human brain is not a computer, and that computers are merely illuminating metaphors, and sometimes tools, for exploring our understanding of minds. In the present case, pursuing the metaphor as if it had literal meaning leads to a forced and ridiculous analogy between genes and ...what?: the factory processes and physical substances involved in making the actual computer? Curiously enough, this is quite illuminating, for surely we would never fall into the trap of suggesting that such processes and substances code, program, or prescribe the problem-solving or learning outcomes achieved by computers?

A second example of pushing the metaphor too far comes from Clark's introduction of the words 'program' and 'code' along the route from genes to behaviour in the case of stepping. If, in the case of stepping, genes contribute to the formation of the physical body and its properties and propensities, and thereby to the neural events which may be triggered in response to certain stimuli, and thereby to certain motor actions, it may be useful to refer to the set of neural events as a 'motor program', perhaps to emphasise the fact that the events form a reliably related set. However, there is no 'program' for real: just a set of neural events and a corresponding set of motor behaviours. Similarly, there is no 'code' in the usual meaning of the word, any more (or less) than certain neural events in my left frontal cortex 'code' the muscle movements in my right hand. The terms 'program' and 'code' are therefore redundant in any explanation of stepping, and they have only crept in because of the power of the computing metaphor.

A third example of pushing the metaphor too far comes from Clark's use of the example of programs written in LISP to demonstrate that (computer) programs are not informationally self-contained. The argument is fine, but is it relevant to brains, minds and behaviour? I think not, unless there are one-to-one correspondences between computer programs and programs in the mind, and between computer languages and language for 'writing' programs in the mind: which I don't think there are.

So I conclude that it is alright to think about genes as programming, coding and prescribing behavioural outcomes, but not to seriously argue that genes actually do this, in any literal sense. I wonder whether Elman et al.'s suggestion that genes should be seen as catalysts may not also be, at least in part, a metaphor. If so, we might have the silly situation of opposing camps arguing at cross purposes because they are each assuming the literalness of their preferred metaphor.

While sitting at my computer (which is not at all like a brain) thinking about Clark's notes, I wrote the following to express my frustration at the dominance of the computer metaphor. I hope it may mildly amuse readers of Connexions, and even strike a chord in some, without giving offence to any:

Chorus of philosophers (antiphonally):

Cognitivist Phil.1.: Genes code;

C.P. 2: Genes describe;

C.P. 3: Genes prescribe;

C.P. 4: Genes program for traits, behaviours or outcomes.

Chorus: The brain is a computer, a connectionist network, a neural net.

The mind is the software.

We are getting there.

We have travelled far.

We are no longer ideas men only, the old men of Philosophicus:

We have models which work:

A whole new vocabulary

Provided by oracles in Computer Science departments ably

assisted by certain psycholinguists and psychologists of

a cognitivist persuasion.

Messenger from Neurox in Ethologica: Woe, woe!

My message is doom!

I have seen King Fodorpus and Queen Chomkasta

Criticised by their offspring, (Andygone amongst others),

Chided by the Furies Elman, Bates, Karmiloff-Smith, Parisi, Plunkett et no doubt al..

Flee! Flee to a shadier spot!
Take refuge in the groves of Neuroscience,
Learn the lingo (central executive and that lot);
But more: learn about brains,
And try some other way to bridge the gap
Between the frontal lobes and not picking your nose in public.

Chorus of philosophers

Cognitivist Phil.1.: Neural structures code;

C.P. 2: Neural structures describe;

C.P. 3: Neural structures prescribe;

C.P. 4: Neural structures program for traits, behaviours or outcomes.

Chorus: The brain is a computer, a dynamical systems network.
We have travelled even further:
We have new models which work:
A whole new vocabulary
Provided by oracles in forward-looking Computer Science departments, ably assisted by neuropsychologists of a cognitivist persuasion.

(To be continued, when 'artificial life' and situated robotics have taken hold...)

In reply to the extract from Andy Clark's "Twisted Tales: Causal Complexity and Cognitive Scientific Explanation"

The emptiness of the self-contained coder

Seth Bullock

The rings on a tree stump can be used to tell how old the tree was when it was felled. In a certain sense the number of rings codes for the age of the tree. However, this sense is a subtle one. When whichever clever lumberjack realised that ``No. Rings = Age(Tree)'' (or whatever the exact relationship is) one imagines his fellow lumberjacks being pretty impressed (perhaps). ``Hey,'' one might call over to an approaching workmate, ``Bill's found out how to tell the age of a tree just by lookin' at the trunk rings''.

However, what one might find harder to imagine is the same lumberjack claiming that Bill had `discovered a code' within the tree. Perhaps Bill could be better described as having `invented a code', but even this construal seems to be stretching the term. Why doesn't the number of trunk rings *code* for the age of a tree?

The genetic code is obviously a much more complicated beast than the trunk-ring-relationship, but are they none-the-less members of the same species? Are they both codes, or are neither codes, or do they differ, and if so why?

Crick and Watson

Like Bill the lumberjack, Crick and Watson stumbled across a systematic relationship within nature. The presence of certain genes rather than others consistently resulted in the presence of certain phenotypic traits rather than others. Or, more accurately (proximally?), certain genes made certain proteins.

Crick and Watson are typically credited with the *discovery* of the genetic code, but if the code existed prior to their discovery, for which interpreter were its coded messages intended? Surely not Crick and Watson themselves?

``Genes code for proteins'' is an oft uttered nugget of biological wisdom. But who needs the code to produce the proteins? Surely the proteins just happen? Although biologists liberally sprinkle their accounts of genetic processes with terms such as translation, interpretation, etc., these are merely glosses on chemical/mechanical processes. Certain genes are merely the kind of chemical structure which result (under suitable conditions) in certain proteins. Just as tree rings are the result (under certain conditions) of the tree's growth processes.

However, there is a difference, and in order to point it out it will be helpful to take a brief look at the work of Ruth Millikan (1984,1992).

Millikan

Millikan concerns herself with constructing a naturalist account of representation (amongst other things). During this undertaking Millikan finds it necessary to establish the set of conditions which must be met for some naturally occurring (i.e., biological) device to mean something. These conditions are quite complex, but the essence of her position is easy to outline.

The meaning of a biological device (a signal for instance) must be determined through examining both the business of the system which produces it, and the business of a second system which consumes it. Typically such systems share a coincidence of business; in some sense they enjoy interdependent agendas.

Crudely, a biological system's business is established over evolutionary time by natural selection (or, sometimes, over ontogenetic time through some learning process analogous to natural selection). Its agenda involves the carrying out of those activities which, through being carried out by its ancestors, enabled them to survive and reproduce, and thereby to bring it into existence. It is this process of reproduction and selection which ensures that it is natural to speak of the heart's function as being to pump blood, since it was through pumping blood (rather than making a thumping noise, or being coloured red) that the heart's ancestors came to survive long enough to allow the organism to which they belonged to reproduce.

By contrast, it is the absence of reproduction and selection which means that a diamond has no such function. (The story for clubs and spades is more complex and must wait for another day.)

Code Consumers

So, Millikan offers us some criteria with which to assess the biological function of genes. Whether these criteria are adequate or even suited to this task is debatable. Millikan's scheme was developed to account for the function of phenotypic traits, not genetic traits, nor the mapping from genes to phenotypic traits. However, at first blush the criteria seem useful in that they allow a distinction to be made between the tree-ring-relationship, and other biological phenomena which are uncontroversially code-like.

The `bee language', which foraging bees, upon returning to the hive, use to inform other bees of the location and quality of distant resources (typically nectar, but sometimes water, or alternative hive locations) is almost paradigmatically iconic. It oozes symbolism. The dance takes the form of a series of movements which vary along two informative dimensions. The angle of the dance indicates the bearing, from the hive, of the resource, whilst the vigour of the dance indicates the quality of the resource. In 1973, after von Frisch (1967) had elucidated the details of this `dance language', he was rightly awarded a Nobel prize for *discovering* those details (more specifically for ``discoveries concerning organisation and ellicitation of individual and social behaviour patterns''). There was no question of his having *invented* the code with which the bees elicited each others behaviour patterns.

The difference between the tree-ring-relationship and the bee dance is the presence of readily identifiable systems acting as consumers in the case of the bee dance (i.e., the watching bees), and the absence of such consumers in the case of the tree-ring-relationship. This difference is all that is needed to rule out the tree-ring-relationship as a natural code. Before Bill the lumberjack stumbled across it, the relationship between tree-rings and tree age was no one's business. It was a coincidental correlation buried within every tree. (If I understood Wittgenstein's private language argument, I might be able to make it do some work hereabouts.)

Coding for Proteins

So, we can now turn our attention back to the (so called) genetic code. Can we identify a consumer for this code? If not, it will have to be consigned to the class of self-contained codes which includes the tree-ring-relationship and countless other correlations, coincidences, and causal regularities, which, although useful to observers interested in predicting natural phenomena, were empty of any meaning prior to their use by such observers.

At first glance no such consumer appears to be present. Genes simply arrange the material that surrounds them (via RNA) into proteins.

However, the material that surrounds genes is not arbitrary, it itself is subject to natural selection. Presumably, over evolutionary time, the environment in which genes operate has become supremely adapted to the role of `medium capable of supporting morphogenesis'.

From this perspective, a natural account of genes casts them as encoders, sending messages in the form of messenger RNA to a consumer system constituted by the protein factory within each cell.

In this light genes are not encoded phenotypic traits, rather their products, RNA molecules, are the codes, written in a language that cellular protein factories have long understood, a language that Crick and Watson can be said to have gone some way towards *decoding*.

Coding for Traits

However, the shorthand description of the function of cellular machinery given above (a `medium capable of supporting morphogenesis') could be expanded considerably, since it was not just through enabling any old morphogenesis that the medium historically enabled the reproduction of the system to which it belonged.

In much the same manner as one might legitimately expand upon the function of the heart, drawing upon its role in oxygenating the blood via pumping it through the lungs, or feeding oxygenated blood to the brain, or any number of functions more complex than merely `pumping blood', one might similarly describe the role of the cell as governing a *specific* morphogenetic regime. This specific regime results in specific phenotypic traits including gross body morphology (Bauplan), eye colour, hairiness, etc. Thus, a less proximal account of the function of the cell machinery responsible for building proteins at the behest of genetic messages would implicate these genetic messages in a complex web of functional processes resulting in specific phenotypic traits, both morphological and behavioural.

The specific phenotypic traits which result from such a morphogenetic regime (one which completes without error under normal conditions) are more than the causal consequence of a complex of biogenetic processes, they are the *normative* outcome of such processes.

The normative character of the systems concerned with generating the phenotype (i.e., that such systems have functions which they may fail to perform) ensures that there is a fact of the matter concerning what phenotype *should* develop. This `normal' phenotype, is simply that phenotype, the development of which enabled the systems' ancestors to be reproduced.

This implies that phenotypes resulting from morphogenetic sequences involving developmental trajectories which differs from the norm are *malformed* rather than merely different. Such trajectories may be the result of non-normal conditions (e.g., egg temperature may fall outside of some normal range), or simply errors in some morphogenetic process (e.g., gene transcription). In addition, phenotypes faithfully developed from genetic mutants must also be classed as aberrations under such a reading, since the systems with which the genes interact were not adapted by evolution to generate the mutant phenotype.

Complexity, Causality, and Self -contained Codes

Such an analysis places little weight on any ``complex extended interplay of factors'', either ``internal'' or ``external''. The mechanisms which instantiate any coding are secondary to the evolutionary functions of the systems involved since it is these functions which are the determinants of the norms by which the systems must be judged, determinants of any content that the system might enjoy, and indeed determine whether the system is contentful at all.

In addition, such an analysis clearly has no place for a ``self-contained code''; such a concept is a contradiction in terms. Codes mediate between the businesses of interdependent systems. The informational content of such codes is necessarily predicated upon the character of both the producing and consuming systems' agendas.

Thirdly, as explanations must be couched in functional rather than causal terms, symmetry of explanation is not implied by causal symmetry. The system that ``does the *most actual work*'' (i.e., the *causal* system most strongly implicated in the `coding for' process) is not necessarily the system that was selected for during evolution, and thus the system with the *function* of bringing about a certain systematic relationship between symbol and world (i.e., the system for whom the `coding for' process is its business).

Similarly, if I choose to use the eruption of Mt. Etna to indicate that it is time for lunch, the complex web of causal processes responsible for the eruption can clearly be glossed over for the purposes of explaining the coding-for relationship which I have established.

Conclusion

So, to conclude, I agree with Andy Clark when he states that he suspects that ``the issue of informational self-containment'' is ``a major red herring''. I would add that norms of behaviour brought about by evolutionary (or analogous learning) processes are the backdrop against which any explanation of what biological systems are for should be cast.

References

Millikan, R. (1984): Language, Thought and Other Biological Categories (Cambridge: Cambridge University Press)

Millikan, R. (1993): White Queen Psychology and Other Essays for Alice (Cambridge: MIT Press).

von Frisch, K. (1967): The Dance, Language and Orientation of Bees (Cambridge: Harvard University Press)

Reviews

R.D.Ellis, *Questioning Consciousness: The Interplay of Imagery, Cognition and Emotion in the Human Brain.*
John Benjamins Publishing Company, Amsterdam/Philadelph ia

Gavin Boyce
Dept. of Philosophy
University of Sheffield

1. The Organicist Hypothesis

One very good reason for being interested in Artificial Intelligence (AI) is the desire to know more about ourselves. The idea being that in addressing the problems confronting such a project we may discover how it is that humans overcome those same problems. If this is your reason for being interested in AI then you are most probably also concerned about the problem of consciousness. From the perspective of AI the problem of consciousness can be seem as the worry that we could create a system that was behaviourally indistinguishable from ourselves and yet for which there was nothing it was like to be that system (a philosopher's zombie). It is thought uncontroversial that it is like something to be a human being, it is likewise almost universally held that it is not like anything to be a plant. In between these extremes there is probably very little consensus, indeed perhaps the only conclusion everyone would be prepared to reject would be that cows, for example, were not conscious whilst, say, ants were. So a theory of consciousness ought to track complexity in some manner.

One unhappy possibility for artificial intelligence, aside from supernatural intervention stories, is what might be called the "organicist" hypothesis. The organicist stance has it that only those creatures with a biological basis can be the subjects of conscious experience. That is to say that a biological basis is necessary for consciousness. It would be odd if the claim were that biology is sufficient for consciousness. This would have it that ants, amoeba and algae were conscious, not something that many would agree with. So in order for something to be conscious it has to biological and x, where x will probably refer to some sort of complexity threshold, whether it be to do with behaviour or information processing. As it stands this stance shows no more than a certain chauvinism. As William G. Lycan says:

"I see no obvious way in which... a creature's... subneural chemical composition should matter to its psychological processes or any aspect of its mentality." (1987: 126)

There is, of course, one obvious way in which the subneuroanatomical chemical composition might matter and that is if it is the only way in which the requisite information processing can be carried out. This is, however, an empirical claim that fails to support an a priori exclusion of non-biological creatures. So, either the organicists are guilty of chauvinism, or what they are really concerned with is the information processing requirements of a conscious system.

Now there does seem to be a strong pre-theoretical intuition that the organicists are right, however, no-one wants to be a chauvinist, so one would expect that there would be an abundance of theories of consciousness that tell us why you can only get the sort of information processing needed for a conscious system from an organic machine. Interestingly enough the majority of theories of consciousness couched in terms of information processing would seem to be ambivalent, to say the least, as to the substrata required to run the sort of information processing that conscious system would require. This may be a result of our pre-occupation with the information processing framework driven by the advance in computers. That information processing framework has a number of basic characteristics:

1. The mind is a symbol processing system.
2. Symbols are acted on by various processes which manipulate and transform them into other symbols which ultimately relate to things in the external world.
3. The aim of psychological research is to specify the symbolic processes and representations which underlie the performance of all cognitive tasks.
4. This symbol system depends on a neurological substrate, but is not wholly constrained by it.

Given such a framework it would be surprising indeed if any of those theories working within it had anything to say out the substrata of conscious systems.

I don't see that the onus of the argument is either one way or t'other here. How you feel about the organicity claim should be driven by your favourite theory of consciousness. Note though, that given the pre-theoretical intuition (whether or not that itself is driven by chauvinism) and the predominance of substratum-ambivalent theories, there is a niche here waiting to be filled; there is probably a demand (albeit a silent one) for an organicist theory of consciousness. If you are looking for some such then I have some good news and some bad news. The good news is that there has been a recent attempt to fill that niche, the bad news is that I'm not so sure it has been successful.

2. Questioning Questioning Consciousness

Ralph Ellis's "Questioning Consciousness" is an attempt at understanding how and why cognitive functions are associated with brains of living organisms. There are perhaps two constraints on acceptability that any theory of consciousness should satisfy. First, there should, where possible, be a fairly clear definition of the target notion. Second, there should be an attempt to account for the empirical data from within that theory. The former is necessary as there is often very little agreement as to what it is that we are trying to explain, and this often leads to theorists talking across each other. Ellis takes his lead here from Merleau-Ponty's reflections on phenomenology and constructs an account which he then applies to much neurophysiological data. How satisfying the reader finds this account will most likely depend upon their theoretical prejudices and to this extent Ellis's proposal fairs no worse on the above two constraints than any other. There is however a further constraint which I feel Ellis's story in particular should satisfy if it is to gain general acceptance and this stems from the niche which it is trying to occupy. If this theory is to sell itself I think it has to give us some reason for thinking that consciousness has to be associated with organic systems. This is the interesting claim in Ellis's book and the one that needs to be backed up with argument if it is to prove convincing.

If you think that consciousness is a process that requires some sort of substratum (organic or not) then you are half-way there, albeit the easy half. What is needed is some description of the relationship between the process and the substratum that allows us to see that that relationship can only hold with an organic substratum. Note that it would be too much to require that the relationship be a conceptually necessary one, when dealing with substantive theorising mere natural necessity should be sufficient. >From an AI perspective all we really want to know is how consciousness in fact arises in humans and why it is that we are thwarted in any attempt to create an inorganic conscious system. So the first question is why does a conscious system have to be an organic system? For Ellis a process can only be conscious if the system in which that process exists:

"... is capable of reproducing elements of its own substrata in order to maintain, continue, or expand its particular pattern of 'cognitive' activity when the availability of an adequate substratum becomes a problem for it; and it must do this because of an internal motivation resulting from its own desires, not the result of an externally imposed process." (1995: 182)

Now the research has shown that the organic substratum that subserves our own consciousness is capable of reorganisation and this plasticity seems to be the feature which, in discussion, most organicists take to be the important distinction between ourselves and inorganic automata, as far as consciousness is concerned. Now I shall not pursue the question of why it is that this plasticity is so important or indeed whether we could create a conscious system that lacked that plasticity, for most AI researchers will admit that plasticity is very important and yet wish to maintain that in the long run such plasticity will be achievable on inorganic machines. No, what is of interest here is that Ellis take himself to be ruling out the case of self-reorganising inorganic systems:

> ...[I]f a neuroscientist engineers a robot in such a way as to systematically replace its own parts as ey become dysfunctional, this does not mean that the robot is an organism or is capable of consciousness, because it is not the robot's 'own' desires (or even its own 'desires') that motivate the replacement, but rather the neuroscientist's desires. (1995: 182)

Now Ellis is very careful to mark the distinction between desires and 'desires' where the latter are merely metaphorical. So I take it that the above is not a denial that a robot can have 'desires'. The point would seem to be that the robot is not conscious because it has a designer. But now there would appear to be a conflict with the organicist intuition that we could create an organic system that could achieve consciousness, an intuition that Ellis shares:

> Probably in the near future, scientists will be able to mix together certain inorganic substances under the right circumstances to generate tissues that are not only inorganic, but also capable of growing to reproduce the substratum for their own states f consciousness as governed by their own desires, and these organisms may well then possess some degree of consciousness. (1995:182)

So it would seem that merely having a designer is insufficient to rule out a particular system from having consciousness. But if this is the case what is the difference between the organic artificial system and the inorganic artificial system?

3. How Not To Sell Organic Produce

Ellis has it that:

> ... the difference between 'organic' and 'purely mechanistic' systems is this: An organism is structured so that at first a process has certain material elements as its substratum; then, because the organism either 'wants' or wants to maintain a similar pattern in the future, it finds other material elements to serve as future substratum for the continuing process. (1995: 188)

This just raises the original question of why it is that we cannot design an inorganic system structured in such a way that the process has certain material elements as its substratum then, because the system either wants or 'wants' to maintain a similar pattern in the future, it can find other material elements to serve the future substratum for the continuing process. The obvious response to this is that for an organic system this process is achieved through the assimilation of nutrients whereas, presumably, an inorganic system would have to add on components in some way. Ellis again:

> ... the computer does not assimilate these alien elements and expand its own operation to use them as additional substrata; even if it did, it would not do this in the service of its own desire for internal conflict resolution, but rather because of one particular lower-level 'desire' (for electrical neutrality in a specific circuit) that a technician has inserted into it in partes extra partes fashion. (1995:178)

So we can take it that the real stumbling block is not that an inorganic system cannot replace its own substratum or even that it replaces its substratum in the wrong fashion (for what difference could the means make?) but rather the problem is that the inorganic system cannot have desires (although it can have 'desires').

Now I believe we are in a position to assess the importance of the central claim of Ellis's theory that:

> The process of questioning is where desire overlaps with representation and it is just in this overlap that 'desire' becomes conscious desire and 'representation' becomes conscious representation. (1995: 30)

The main thrust of Ellis's theory is that it is through a process of self-questioning that organic systems become conscious. So for Ellis the "biological and x" is cashed out as a "biological and self-questioning", so that a system that is biological and goes in for self-interrogation in the fashion outlined by Ellis will be a conscious system. Now there are many merits to that part of Ellis's theory that deals with the extra complexity required, over and above organicity, in order for a system to be conscious, however, none of this appears to address my central worry here which is that Ellis's theory does not seem to provide support for the premise that a conscious system needs to be an organic system. Why couldn't the 'desires' of an inorganic system become desires through a process of self-interrogation?

I have spent most of my time here focusing upon the organicity issue for the reason that it is amongst those organicist intuitions that Ellis's theory is likely to find most sympathisers. But this being the case there is more work needed from within the framework provided by Ellis if the AI fraternity is to be dissuaded from working with inorganic materials, or indeed if the organicist stances is to gain more than just intuitive acceptance.

Reference

Lycan, W.G. (1987) *Consciousness*. Cambridge, Mass.: MIT Press.

Horst Hendriks-Jansen, *Catching Ourselves in the Act*

1996, Cambridge, Mass.: MIT, Press ISBN 0 -262-08246-2 pp.367 Price: £24.95 (hbk)

Jill Boucher
Dept. of Human Communication Sciences
University of Sheffield

This is a scholarly book about models of the mind and of behaviour, and about theories concerning the evolutionary and ontogenetic development of the human mind and behaviour. The book discusses work from the philosophy of science, the philosophy of mind, computer modelling, ethology, cognitive psychology, developmental psychology, and other disciplines. This makes it quite a difficult read - unless you are conversant with work in all of these fields, which I for one am not. However, the book is clearly written and includes plenty of opening and closing sections within individual chapters which signal the main arguments to come, or summarise the main points which have been made. In addition, Chapter 1 includes a guide to different routes through the book, highlighting those chapters which are essential to the main arguments as opposed to those which may be of interest mainly to specific groups of readers. In the middle of the book there is a short chapter explicitly linking the first part of the text which is largely critical of existing theories and models of the mind, through to the second part of the book which introduces and elaborates an alternative model.

The main themes of the book are best stated by Hendriks-Jansen himself. In the opening paragraph of Chapter 1 he writes:

> The new disciplines of artificial life and situated robotics have called into question some of the basic principles of cognitive science and traditional AI. Decomposition by function, the conceptualisation of an autonomous agent as an input-output model, the separation of central processes from peripheral processes, and the principle of formal task definition have all come under attack. If these criticisms are taken seriously, as I believe they should be, they call for a fundamental reappraisal of the computational paradigm.

And on p. 2 he writes:

> I shall argue that cognitive science, conceived as a legitimisation of folk psychology and the intentional stance through computer-based models, has so far failed to provide a satisfactory scientific explanation of human and animal behaviour. Explanations of this type (irrespective of whether they use classical or connectionist architectures) depend on internal representations derived top down from an analysis of full-blown symbolic behaviour. This strategy has not succeeded in identifying defensible 'natural kinds', and it might be based on a misconception about the role that explanatory models have traditionally played in science.

From the quotes above it will be clear that Hendriks-Jansen has several substantial targets in his sights. At the outset, he is interested in the logic of explanations of human behaviour, basing much of his discussion on the arguments of Nagel (1961). Nagel's work provides him with a framework for considering whether or not computational models might be able to explain human behaviour. He recognises "the explanation-imitation dispute that has haunted AI", and presents the case made by Simon and his colleagues over the years in favour of AI models as explanations of behaviour. However, he comes down against Simon et al., for a number of reasons. These revolve round the fact that computational models do not select as entities to be explained those 'natural kinds' of behaviour which may be identified by painstaking observation and analysis of behaviour as it occurs in natural environments. Instead, the method of functional decomposition is used, whereby the terms of folk psychology are presumed to identify the natural kinds of behaviour which any model of the mind or of behaviour must explain.

Hendriks-Jansen illustrates the fallacious nature of this approach using Konrad Lorenz's observation that what we refer to, using a process of functional decomposition, as 'parenting behaviour' in a mother duck, turns out to consist of a number of different activities triggered by diverse stimuli emanating from a duckling. So 'parenting behaviour', though a useful descriptive phrase, is not a natural kind in terms of the duck's behaviour, or in terms of the evolution of duckling-focused behaviours. Nor does the parenting behaviour we observe in a duck imply that the duck has an internal representation of the duckling on which her behaviour focuses. Explanations of parenting behaviour in ducks and of the evolution of this set of behaviours should not, therefore, include either 'parenting behaviour' or internal representations of ducklings as constituents.

Hendriks-Jansen goes on to consider Millikan's (1984) suggestion that the natural kinds to be explained by theories of behaviour are the proper functions of behavioural or morphological traits, 'proper functions' being defined as those to which a trait owes its survival through natural selection. Millikan argued that proper functions can only be identified by studying the evolutionary history of use of particular traits, and Hendriks-Jansen agrees with this, concluding that "natural selection constitutes our only hope of a principled explanation of behaviour". He is also enthusiastic about Millikan's argument that proper functions must be identified by observing individuals' behaviour in normal environments, since the meaning of any particular behavioural act can only be understood in terms of the context in which it occurs. He writes: "The great advantage of Millikan's radical proposal is that it makes a clean break with causal and informational theories of content and locates meaning squarely outside the head....Natural selection, for Millikan, confers true intentionality. Proper function under normal conditions just is intentionality."

However, Hendriks-Jansen parts company with Millikan at the point at which she argues that proper functions become allied to goals or objects of behaviour which are internally represented - (the 'duckling in the head' which the mother duck has as the object of her behaviour in Lorenz's example). Hendriks-Jansen argues that prior to language acquisition, behaviour cannot be directed towards an internally represented goal, since, in absence of shared language, the concept grounding/symbol grounding problem is insuperable. A substantial portion of one chapter is devoted to explaining why he takes this view, including a discussion of the work of Marr, Harnad and others.

In place of Millikan's 'proper functions' as the natural kinds to be explained in any theory of human behaviour, Hendriks-Jansen therefore suggests that primitive reflexive actions and situated activity patterns should constitute the natural kinds to be explained. These, he argues, are the basic building blocks from which behaviour develops, and it is the interaction between these primitive behaviours and the environments in which they occur which must be invoked in any explanatory theory - (more of this below).

Hendriks-Jansen makes several other arguments against classical AI models as explanations. For example, he points out that such models have failed to deliver the unifying calculus looked for in computational psychology. Rather, there has been a proliferation of programs, each designed to model behaviour within some narrowly defined task. Moreover, he points out that classical AI sets itself tasks which are amenable to a symbol processing approach, for example, playing chess. Where a task is not so obviously symbol-based, the task is nevertheless formally described in terms of symbols: thus many of the tasks which computers are given to solve do not have the properties of behaviour, but rather the formal properties of computer models.

Hendriks-Jansen extends his criticisms of classical AI to connectionist models, and there is a chapter in the book entitled "Connectionism: Its Promise and Limitations as Currently Conceived". While allowing that "the methodology and mathematics of network theory may give reasons to suppose that a new explanatory level has been achieved (statistics or dynamical systems theory instead of symbol processing), ... ultimately the paradigm still rests on an input-output model, and therefore needs to incorporate a theory about the nature of its inputs and outputs, as well as the internal representations in between". Hendriks-Jansen argues that connectionist models have failed to provide such a theory, and that all that connectionism has been able to come up with to date "are redescriptions of the symbolic level it purports to replace".

The relatively new disciplines of artificial life (Alife) and situated robotics mentioned in the opening paragraph of the book are, however, looked on more favourably, and substantial portions of the book are devoted to discussing key theories and models within these disciplines. Hendriks-Jansen sees Alife and situated robotics as being potentially capable of providing explanatory models of behaviour (though not of either the brain or the mind). He also considers that 'situated activity', i.e. the activity of autonomous agents interacting with their natural environments, may be capable of grounding the intentional behaviour and contentful thoughts of preconceptual/presymbolic minds. The example he gives comes from the wall-following robot developed by Mataric (1992). This robot can learn to negotiate its way around a walled area, eventually behaving as if it had representational knowledge of walls. The robot does not in fact have any representational knowledge of walls, but rather a set of behaviours relating to wall-negotiation which come to be reliably triggered as the robot experiences different landmarks in the course of its situated activity. In a similar way, Hendriks-Jansen argues, a young child's innately determined patterns of activity interact with the environment (which includes, pre-eminently, a responsive mother) to produce apparently purposeful, apparently representation-guided behaviour. It is not, however, until the child learns to associate words with elements of situated activities that the uniquely human conceptual/symbolic mind begins to form.

In place of the top-down, language-guided, symbol mediated approach of AI and connectionism, therefore, Hendriks-Jansen is advocating that a bottom-up, observation-guided, situated-activity focused approach should be used first to identify the natural kinds which should form key constituents of any explanatory theory, and then to build the theory using insights from situated robotics and Alife models. He advocates that the methods of ethology should be used to identify the innately determined action patterns which constitute the natural kinds to be explained in a new theory of human behaviour. These, he claims, will be present in human infants, and include those which have been identified by developmentalists such as Kaye (1979) and Trevarthen (1977), using ethological methods. The critical aspect of these action patterns is that they elicit certain predictable reactions from caregivers, who by treating the infant's behaviour as intentional and meaningful, gradually shape it to be so, just as the behaviour of Mataric's robot became intentional via interactions with the environment, even in the absence of internal representations. Language is learnt as a form of situated behaviour, and only then are 'true concepts' introduced into the mind (see below).

It should now be clear, I think, why I described this book as a difficult read: it is rich with interlocking arguments borrowing from, and moving off from, important and sometimes controversial topics from a number of disciplines. For this reason, I have almost certainly undersold Hendriks-Jansen in my attempts to communicate some of his arguments and may even have misrepresented his detailed views at specific points. For these inadequacies I apologise, but hope that I have given readers of Connexions at least a flavour of the content of the book.

What do I actually think of the book? Well, I think that anyone interested in the philosophy of mind, and in what has come to be called 'cognitive science' should read it. This is not because I think that all - or even many - of the arguments in the book can withstand counter-attack unscathed, but because I think that Hendriks-Jansen is attempting to force through a paradigm shift, and I think that it is a shift in a needed direction. In fact his efforts are not isolated: Hendriks-Jansen himself recognises that authors such as Dennett (1991), Sinha (1988), Clark (1993) and Bruner (1990) are all in a variety of ways trying to get out from under what Hendriks-Jansen describes as "the dead hand of computationalism". In my own reading as a developmental psychologist the book 'Rethinking Innateness', which I recently reviewed for Connexions, also stands out as an important part of the reaction against what may be shorthanded as the 'standard cognitive science model of the mind' (following Pinker).

To me, the main contribution of 'Catching Oneself in the Act' is the case made against the assumption that the symbols which we use to talk about human behaviour and the human mind are the appropriate entities, or natural kinds, to figure in explanatory theories and models of the evolutionary and ontogenetic development of behaviour and mind. Related to this is the assumption inherent in explanatory models based on classical AI, that the human mind is essentially and pervasively representational, and that a one-to-one correspondence will eventually be found between the symbols utilised in models, representations in the mind, and patterns of brain activity. Hendriks-Jansen questions both of these assumptions, illustrating the fallacious nature of both assumptions within the example taken from Lorenz.If Hendriks-Jansen is correct in arguing against both these assumptions, then a massive amount of psychological, philosophical and other theorising about behaviour and the mind is seriously wrong.

I think that Hendriks-Jansen's central arguments are correct, and that a great deal of recent theorising about human behaviour is seriously wrong in certain respects. With regard to the way in which the AI computational approach has had a negative influence on theories of child development, one may take the example of the recent literature on 'theory of mind'. As is well-known, observations of the responses of young children and of children with autism carrying out false belief tasks led researchers to coin the phrase 'theory of mind' as an inferred explanatory construct. This in turn generated a body of work discussing whether so-called theory of mind actually constitutes a theory, as properly understood in science. In addition, many cognitive scientists now discuss a so-called 'theory of mind module' or quasi-module of the Fodorian type. Other observations of the behaviour of children with autism led to the inference that there are an 'eye direction detection mechanism' , a 'shared attention mechanism' and a 'theory of mind mechanism' complete with a 'selection processor' within the 'theory of mind module', and to the further inference that some abnormality affecting one or other of these innate subsidiary mechanisms causes autism. Hendriks-Jansen's work underlines the fact that it is a fallacy, or category mistake, to assume that inferred explanatory entities such as 'theory of mind' or 'shared attention mechanism' have any parallel existence in the mind, let alone within the brain. He has in fact argued this point in detail, and with reference to explanations of autism, in a paper to be published in the near future.

With regard to the negative effect which the AI computational approach has had on theories in evolutionary psychology, one only has to look at the proliferation of traits (usually said to constitute 'modules') supposedly selected for in the process of evolution and somehow 'written into' the genes of the human species. These include 'parenting behaviour', 'mate selection', 'cheater detection', as well as 'theory of mind'. Equally, I agree that it is wrong to claim that natural selection has ensured that infants are born with innate knowledge of, for example, the boundedness of objects, or linguistic universals. Natural selection acts through the genes, and the only thing which the genes can influence is brain and body stuff, i.e. the hardware and its propensities to function in particular ways, either for endogenous reasons or in response to environmental stimuli. These propensities may cause an individual to behave in ways which can be appropriately conceptualised and labelled as parenting behaviour, cheater detection, or following a syntactic rule, etc. But it is illegitimate to move from the linguistic description of a certain set of behaviours to a claim that 'parenting behaviour', 'cheater detection', or 'knowledge of syntax' have been selected for and are innate (cf. what Elman et al., 1996, say about innate representational knowledge).

While thoroughly endorsing what Hendriks-Jansen has to say about the pre-representational mind and its development during evolution and ontogeny, I find his chapters on the development of the conceptual/symbolic mind much less well-worked out. In making the case for the intentionality of the pre-conceptual, pre-symbolic mind, Hendriks-Jansen adopts Cussins' (1990) notion of an S-domain, defined as "a substrate corresponding to the system's perceptuomotor abilities.....not represented by the system itself". The perceptuo-motor abilities within the S-domain are viewpoint dependent, and cannot be manipulated, detached from their situated origin, or used in different combinations. The 'abilities' in the S-domain thus have clear similarities to the knowledge in Karmiloff-Smith's (1992) Level 1 'procedures', which, like S-domain abilities, are not available to the system as data. However, Hendriks-Jansen distances himself from Karmiloff-Smith's argument that an endogeneously driven process of representational redescription is responsible for breaking down procedures into component parts which can be manipulated independently of each other and which, over successive processes of redescription, become increasingly explicit. Instead, Hendriks-Jansen follows Clark (1993) in suggesting that "What cognitive scientists call 'systematicity'...is not the result of internalised atoms of meaning combined and transformed by syntactic rules; (but rather) the ability to act appropriately in a variety of contexts involving the object, property, or relation which the public concept picks out'. In other words, Hendriks-Jansen follows Clark's lead in emphasising the role of the environment in providing data from which systematicity of thought and language are abstracted.

As far as concept development is concerned, this suggestion is not new, since Nelson (1974) long ago proposed at early concept development grew out of the child's experience of routine activities and the event schemas and 'scripts' based on these experiences. However, Hendriks-Jansen. is not prepared to follow Nelson - or, perhaps, Clark - in allowing that situated activity is sufficient to ground 'true concepts' capable of being manipulated, detached from their situated origin, or used in different combinations. He writes: "It is a long way from the grounding that has just been described to proficient conceptual thinking. That is where language comes in." I think that what he is denying is the possibility that repeated instances can give rise to anything other than a grouping, or category (eg. of dogs, of things coloured red, of hopping things), whereas Karmiloff-Smith is proposing that the human mind/brain will spontaneously generate a representation which is more abstract than the instances which form the category. For Hendriks-Jansen, only the language of others can enable the child to make that leap, so that it is only with the onset of language that the child's mind becomes truly representational.

This position does not, however, tally with what is known about the development of signs and symbols in prelinguistic children, and I think that Hendriks-Jansen is unduly pessimistic about the possibility of grounding concepts and symbols in situated activity without the aid of language. Hendriks-Jansen's position also makes it hard to imagine how language could have evolved, if there are no concepts until another human being provides you with them: here again, situated activity and interaction with others is usually thought to provide the grounding for protolanguage, at least.

Clark's suggestion, apparently approved of by Hendriks-Jansen, that the rules of syntax may be inferred from the experience of acting appropriately in varied contexts, is more novel and contentious. Hendriks-Jansen spells out his own view of language acquisition as follows:

> The alternative (to Chomsky's view) I have sketched proposes that language competence does have a partial explanation in natural selection, in that we have certain species-typical activity patterns adapted for language acquisition but that the actual development of the competence can come about only through performance and requires the prior existence of a public language that can serve as scaffolding.

Hendriks-Jansen recognises that this gives him the problem of explaining how public language come about in the first place, and appeals to an argument proposed by Freyd (1983) that "the structures evident in language ...might be a property of language itself, and their explanation may lie in historical processes of transmission and elaboration". Hendriks-Jansen is also aware of the arguments for innate knowledge of universal grammar, and writes:

> I have tried to show that the interactive context provided by the parents, combined with certain species-typical perceptual abilities that seem to be adapted to speech comprehension, as well as extensive practice with speech acts before language acquisition begins, make this - (i.e. the problem of grammar acquisition) - less of a stumbling block than has generally been assumed.

I feel quite certain that the main arguments for innate knowledge of universal grammar remain virtually untouched by anything which Hendriks-Jansen's has said.

In sum, this book is ambitious and important. The range of knowledge which it displays is impressive, not least because the author is, according to the cover blurb, a mathematician by training, and a commercial systems designer by trade, only recently turned academic. I am fairly certain that the book will be vigorously attacked and, unfortunately, I suspect that there are a number of weak points amongst Hendriks-Jansen's many arguments (other than those I have indicated). However, I think that his central arguments about the nature of explanation, about natural kinds and functional decomposition, about the methodology involved in modelling the mind, and about evolution and intentionality in the preconceptual mind will hold, and help to lift 'the dead hand of computationalism'. Personally, I can't wait. However, I do also sincerely trust that the role of classical AI in modelling the fully formed, adult symbolic mind will not be thrown out along with the role of classical AI in modelling the development of mind in young children and during evolution.

References

Bruner, J. (1990): *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Clark, A. (1993): *Associative Engines: Connectionism, Concepts and Representational Change*. Cambridge, MA: The MIT Press.

Cussins, A. (1990): 'The connectionist construction of concepts'. In M.Boden (Ed.). *The Philosophy of Artificial Intelligence.* London: Oxford University Press.

Dennett, D. (1991): *Consciousness Explained*. London: Allen Lane, Penguin Press.

Elman, J. et al. (1996): *Rethinking Innateness*. Cambridge, MA: The MIT Press.

Freyd, J.J. (1983): 'Shareability: The social psychology of epistemology'. *Cognitive Science*, 7, 191-210.

Kaye, K. (1979): 'Thickening thin data: The maternal role in developing communication and language.' In M. Bullowa (Ed.). *Before Speech: The Beginning of Interpersonal Communication*. Cambridge: CUP.

Karmiloff-Smith, A. (1992): *Beyond Modularity: A Developmental Perspective on Cognitive Change.* London: The MIT Press.
Mataric, M.J. (1992): 'Integration of representation into goal-driven behaviour-based robots.' *IEEE Transactions on Robotics and Automation*, 8, 304-312.
Millikan, R. (1984). *Language, Thought, and other Biological Categories: New Foundations for Realism*. Cambridge, MA: The MIT Press.
Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientifi Explanation*. London: Routledge, and Kegan Paul.
Nelson, K. (1974). 'Concept, word and sentence: Interrelations in acquisition and development'. *Psychological Review*, 81, 267-285.
Sinha, C. (1988). *Language and Representation*. Hemel-Hempstead: Harvester-Wheatsheaf.
Trevarthen, C. (1977). 'Descriptive analyses of infant communicative behaviour.' In H.R. Schaffer (Ed.) *Studies in Mother-Infant Interaction: Proceedings of Loch Lomond Symposium*. New York: Academic Press.

Rocco J. Gennaro, *Consciousness and Self-Consciousness*
1996, John Benjamin Publishing Co., Philadelphia/Amsterdam.

Peter Carruthers
Dept. of Philosophy
University of Sheffield

Gennaro defends a form of higher-order thought (HOT) theory of mental state consciousness, following but emending Rosenthal (1986, 1993). Although I am sympathetic to Gennaro's overall position (see my 1996, chs. 5-7), the book is poorly written and not very well structured; and some of the argumentation is weak. Here I shall focus on just two points of interest - Gennaro's disagreement with Rosenthal, and his attempts to argue that state-consciousness (and consequently HOTs) will be quite widespread in the animal kingdom, requiring little in the way of conceptual sophistication.

According to Rosenthal, a mental state M of mine counts as conscious if and only if I entertain at the same time a HOT about M, provided that the HOT is non-inferentially caused by the presence of M. Gennaro objects that this makes consciousness an extrinsic, rather than an intrinsic, property of a state. That is, one and the same state which is now conscious could have existed without being conscious, and is only conscious in virtue of its extrinsic relation to another state (the relevant HOT). In order to preserve intrinsicality, Gennaro proposes that consciousness should be seen as a complex state, comprising both the first-order state M and also the HOT which renders the complex state conscious. That state (the state consisting of M, the HOT about M, and the non-inferential causal relationship between them) could not have been non-conscious; and its status as conscious is due to its intrinsic properties (namely that it consists of a HOT which is appropriately targeted on another mental state), not its relations with anything else.

It is an unfortunate consequence of this view that no first-order mental states, and no simple mental states, are themselves conscious. (This is apparently not noticed by Gennaro, who continues to speak, e.g., of 'conscious pains'.) Rather, such states are at best components of complex states which are conscious. So my current perception of a glass on a desk is not itself a conscious state; rather, it is a part of a complex state which is conscious, made up of the percept itself and a HOT about that percept. Indeed, there will be no such things as conscious perceptions, on Gennaro's account; nor will there be any conscious pains, or any conscious first-order thoughts. What there are, rather, are conscious states which have percepts, pains, and first-order thoughts as parts. This is highly counter-intuitive. When one is subject to a flash of bright light, say, one wants very much to say that it is the experience of the flash itself which is conscious; not just that the percept in question is a part of something which is conscious.

What reason is there to believe that consciousness is an intrinsic property of a conscious state? Gennaro provides none, beyond claiming that this seems manifest to common-sense intuition. He says that from a first-person point of view, it seems as if consciousness is an intrinsic property, and is quite unlike such properties as being the cousin of or being to the left of. But this intuition is easily explicable, on Rosenthal's account. For the latter's view is that the HOT which renders a given state M conscious will characteristically not, itself, be a conscious one. Small wonder, then - if we lack any awareness of the HOT which renders M conscious - that we should lack any intuitive sense that the conscious status of M is extrinsic, dependent upon its relation to a HOT. If the full extent of our awareness, when we undergo a conscious state M, is generally the presence of M itself, then it is hardly surprising that we should be inclined to think that the conscious status of M is intrinsic to it.

I turn now to the question of how widely distributed consciousness will be, on a HOT account. Gennaro acknowledges that if possession of a conscious mental state M requires a creature to conceptualise (and entertain a HOT about) M as M, then probably very few creatures besides human beings will count as having conscious states. Let us focus on the case where M is a percept of green, in particular. If a conscious perception of a surface as green required a creature to entertain the HOT [that I am perceiving a green surface], then probably few other creatures, if any, would qualify as subjects of such a state. There is intense debate about whether even chimpanzees have a conception of perceptual states as such (see, e.g., Povinelli, 1996); in which case it seems very unlikely that any non-apes will have one. So the upshot might be that consciousness is restricted to apes, if not exclusively to human beings.

This is a consequence Gennaro is keen to resist. He tries to argue that much less conceptual sophistication than the above is required. In order for M to count as conscious one does not have to be capable of entertaining a thought about M qua M. It might be enough, he thinks, if one were capable of thinking of M as distinct from some other state N. Perhaps the relevant HOT takes the form, this is distinct from that. This certainly appears to be a good deal less sophisticated. But appearances can be deceptive - and in this case I believe that they are.

What would be required in order for a creature to think, of an experience of green, that it is distinct from a concurrent experience of red? More than is required for the creature to think of green that it is distinct from red, plainly - this would not be a HOT at all, but rather a first-order thought about the distinctness of two perceptually-presented colours. So if the subject thinks, this is distinct from that, and thinks something higher-order thereby, something must make it the case that the relevant this and that are colour experiences as opposed to just colours. What could this be?

There would seem to be just two possibilities. Either, on the one hand, the this and that are picked out as experiences by virtue of the subject deploying - at least covertly - some concept of experience, or a near equivalent (such as a concept of seeming, or sensation, or some narrower versions thereof such as seeming colour or sensation of red). This would be like the first-order case where I entertain the thought, that is dangerous, in fact thinking about a particular perceptually-presented cat, by virtue of a covert employment of the concept cat, or animal, or living thing. But this first option just returns us to the view that HOTs (and so mental-state consciousness) require possession of concepts which it would be implausible to ascribe to most species of animal.

On the other hand, the subject's indexical thought about their experience might be grounded in a non-conceptual discrimination of that experience as such. We might model this on the sort of first-order case where someone - perhaps a young child - thinks, that is interesting, of what is in fact a coloured marble - but without possessing the concepts marble, sphere, or even physical object - by virtue of their experience presenting them with a non-conceptual array of surfaces and shapes in space, in which the marble is picked out as one region-of-filled-space amongst others. Taking this second option would move us, in effect, to what Dretske (1995) calls a Higher Order Experience (HOE) account of consciousness. Just such a view has been recently defended by Lycan (1996), following Armstrong (1968, 1984). (Gennaro himself alleges - surely wrongly - that there is no real distinction between HOE and HOT accounts.)

How plausible is it that animals might be capable of HOEs? Lycan (1996) faces this question, arguing that HOEs might be widespread in the animal kingdom, perhaps serving to integrate the animal's experiences for purposes of more efficient behaviour control. But a number of things go wrong here. One is that Lycan seriously underestimates the computational complexity required by such internal monitors. In order to perceive an experience, the organism would have to have the mechanisms to generate a set of internal representations with the content (albeit non-conceptual) of that experience. For remember that both HOT and HOE accounts are in the business of explaining how it is that one aspect of someone's experiences (e.g. of movement) can be conscious while another aspect (e.g. of colour) can be non-conscious. So in each case a HOE would have to be constructed which represents just those aspects, in all of their richness and detail. But when one reflects on the immense computational resources which are devoted to perceptual processing in most organisms, it becomes very implausible that such complexity should be replicated, to any significant degree, in generating HOEs. (It is one of the great advantages of HOT theories over HOE theories that they require no such complexity. They just require the subject to possess concepts of experience, to be deployed in response to the very same states which represent the world to us and which ground our first-order thoughts about that world. See my 1996, chs. 6 and 7.)

Lycan also goes wrong, surely, in his characterisation of what HOEs are for (and so, implicitly, in his account of what would have led them to evolve). For there is no reason to think that perceptual integration - that is, first-order integration of different representations of one's environment - either requires, or could be effected by, second-order processing. So far as I am aware, no cognitive scientist working on the so-called 'binding problem' (the problem of explaining how representations of objects and representations of colour, say, get bound together into a representation of an object-possessing-a-colour) believes that second-order processing plays any part in the process.

Notice that it is certainly not enough, for a representation to count as a HOE, that it should occur down-stream of, and be differentially caused by, a first-order experience. So the mere existence of different stages and levels of perceptual processing is not enough to establish the presence of HOEs. Rather, those later representations would have to have an appropriate cognitive role - figuring in inferences or grounding judgements in a manner distinctive of second-order representations. What could this cognitive role possibly be? It is very hard to see any other alternative than that the representations in question would need to be able to ground judgements of appearance, or of seeming, helping the organism to negotiate the distinction between appearance and reality. But that then returns us to the idea that any organism capable of conscious mental states would need to possess concepts of experience. (See my 1996, ch. 5.)

I conclude that Gennaro has failed to motivate his amendment to Rosenthal's HOT account; nor has he succeeded in showing that HOTs, and hence conscious mental states, are possessed by any creatures besides ourselves and perhaps the other great apes. (In fact, the real competition for Rosenthal's HOT theory comes, I believe, from a dispositionalist alternative, according to which conscious states are not necessarily those which actually cause a HOT about themselves, but those which figure briefly in a special-purpose short-term memory store whose function is, inter alia, to make its contents available to HOTs. See my 1996, chs. 6 and 7.)

Acknowledgement
I am grateful to Tom Dickins and Susan Granger for their comments on an earlier draft

References

Armstrong, D. 1968. *A Materialist Theory of the Mind*. London: Routledge.

Armstrong, D. 1984. 'Consciousness and causality.' In D. Armstrong and N. Malcolm, *Consciousness and Causality*. Oxford: Blackwell.

Carruthers, P. 1996. *Language, Thought and Consciousness: an essay in philosophical psychology.* Cambridge: Cambridge University Press.

Dretske, F. 1995. *Naturalizing the Mind.* Cambridge, MA: MIT Press.

Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.

Povinelli, D. 1996. 'Chimpanzee theory of mind?' In P. Carruthers and P. K. Smith (eds.), *Theories of Theories of Mind*, 293-329. Cambridge: Cambridge University Press.

Rosenthal, D. 1986. 'Two concepts of consciousness.' *Philosophical Studies*, 49, 329-359.

Rosenthal, D. 1993. 'Thinking that one thinks.' In M. Davies and G. Humphreys (eds.), *Consciousness*, 197-223. Oxford: Blackwell.

Conference Report on Richard Byrne's paper "Origins of mind: Using the primate order to trace the early evolution of cognition".

Thomas E. Dickins
Dept. of Human Communication Sciences
University of Sheffield

Richard Byrne, of the University of St. Andrews, gave the British Psychological Society Book Award Lecture in London on the 16 December 1997. Byrne won the award for his book "The Thinking Ape: Evolutionary Origins of Intelligence" (1995, Oxford University Press). Like the book the talk was succinctly organised, cogent and very well delivered. Here is the core of Byrne's thesis:

Byrne opened with two questions "why study primates as a psychologist?" and "how did intelligence arise?" He argued that the latter question was often thought only to lead to speculative answers. However, when such enquiry is combined with comparative work on primates, speculation becomes honest theory.

As the opening questions indicate there are two distinct stages to understanding the origins of mind. First, we need to trace actual ancestors, and second we use behavioural data from related species. When we look at data from the second stage we are bound to notice four

1. Monkeys and Great Apes show more complex behaviour than most mammals.
2. This appears to be caused by a relatively enlarged neocortex that possibly allows rapid learning.
3. Great Apes show understanding of mechanism.
4. This is potentially caused by the possession of representational abilities.

Byrne discussed the social alliances found in monkeys and apes, and how they are maintained through mechanisms such as social grooming. Given that there is complex social structure a certain amount of social knowledge might be predicted and work on tactical deception, amongst other things seems to provide evidence of such knowledge, and its manipulation. What is more, tactical deception appears to be a species wide trait within monkeys and apes and they are quicker to learn socially relevant information than other species.

There is a good example of tactical deception in Byrne's book (1995:124). He describes how a young baboon, on coming across an adult baboon who had just exhumed a root to eat, let out a scream. This scream alerted the young baboon's mother, who outranked the adult. The mother rushed over, took in the scene of the adult, the root and the seemingly distressed young animal and chased the adult off. Meanwhile the young baboon ate the root. It seems the young baboon knew what the scene he had created would look like to his mother and how it would be interpreted. What is more, he was aware enough of social ranking in his group to use it to his advantage. All of this points to a brain specifically "designed" by evolutionary mechanisms for group living. As indicated above the area of most interest in terms of brain design is the neocortex.

At this point Robin Dunbar's (1) recent work showing a positive correlation between neocortical ratio (the ratio of neocortical volume to whole brain volume) and social group size was cited. Byrne argued that group size is an indirect measure of social complexity. Furthermore, he has found a good positive correlation between neocortical ratio and the frequency of occurrence of tactical deception. So increase in neocortical ratio might allow for these social abilities. Byrne ruled the possibility of trial and error learning of tactical deception out because it is unlikely that there would ever be enough trials in "real world scenarios".* - see comment at end

Social abilities aside Byrne also described differences between the two broad primate groupings. Great Apes appear to understand mechanisms and tools better than Monkeys. Monkeys do seem to learn these skills by trial and error for each new instance of a tool that they encounter - even if the tool remains constant over many tasks. Chimps, for instance, appear to understand the utility of a tool and generalise it across instances. What is more they appear to plan. Gorillas can do this too and also perform complex and directed series of actions in order to deal with various plant defences when preparing their food. Indeed, Byrne showed a flow-diagram that detailed the stages of a Gorilla stripping down a plant covered in thorns whereby the Gorilla held the plant in specific ways in order not to encounter the thorns at their functional angle. This diagram had all the features one would expect of a systematic human behaviour. In short Chimps and Gorillas are able to build up hierarchical structured routines using one program as an embedded subroutine in another. In humans we call this order of behaviour planned. Byrne's big claim is that this sort of behaviour must be based upon mental representation.

The representational abilities proposed for planning-type behaviours are not of the same order as social knowledge, and may not be caused by increased neocortical ratio as some Great Apes have a smaller neocortical ratio than some monkeys. (This raises two points: first, what is the nature of mental representation that Byrne is proposing? Social knowledge has to be a form of mental representation, yet Byrne is keen to see it as cognitively different from representation used in more "abstracted" off-line processing. Second, there is an indication here of domain specific cognitive organisation which is very much in line with the evolutionary speculations of Leda Cosmides and John Tooby (2). Such architecture would imply specific computational differences between different systems, not merely qualitative differences between actual representations.) None the less, all Great Apes have brains absolutely larger than any monkeys and this may be the key to this order of representation. If this is so Byrne tantalisingly suggested that we could assume such abilities in, for instance, whales.

Byrne ended his talk with some putative timescales for the onset of representational abilities. The common ancestor of all monkeys and apes was extant circa 40 million years ago (mya). The first primates were to be found circa 65 mya. So primate brains obviously began to enlarge between 65 and 40 mya due to the probable selective pressure of increases in group size which in turn might be due to various ecological. By 16 mya mental representation is likely to have entered the equation, as this is when the common ancestor for all Great Apes was to be found. But what could be the possible selective pressure for representation in apes? Byrne speculated that competition with Monkey groups, who are better able to digest unripe fruit and can thus clean up on many food sources, might lead to certain cognitive abilities being selected for if they help in making available sources of nourishment that Monkeys are not privy to.

Obviously such a brief overview of a talk cannot do it justice but I hope that I have given you the sense of Byrne's approach. The good biological thinking that motivates not only this lecture but also Byrne's book is a pleasure to absorb. More than this it is good to see evolutionary approaches to both functional and causal explanation in the behavioural sciences being given serious thought.

Richard Byrne writes in response:

This is in fact a misunderstanding, and the exact reverse is the case. The vast majority of records of tactical deception are most likely to have been learnt by trial and error. This reflects the ability of large-brained monkeys and apes to learn rapidly, especially in social situations. The idea, that a tactic that functions by confusing or misleading a conspecific must have been intended to confuse or mislead, is a common misunderstanding. The fact that learning-without-understanding is a likely mechanism of acquisition has been emphasised repeatedly in my publications since 1985, so it may be that humans are in some way pre-adapted to misconstrue these situations!

References

Dunbar, R.I.M. (1996) *Grooming, Gossip and the Evolution of Language*. Faber and Faber.
Barkow, J., Cosmides, L. and Tooby. J. (1992) *The Adapted Mind*. Oxford University Press.

Acknowledgement

The author would like to thank David C. Giles and Lucy C. Reid for their comments on an earlier version of this review.

Never ignore co-evolution: reflections on Paul Bloom's "Some issues in the evolution of language and thought."

Adam Morton
Dept. of Philosophy
University of Bristol

The stated purpose of Bloom's article (to appear in *Evolution of the Mind* edited by Denise Cummins and Colin Allen, MIT Press, also presented at the Hang Seng workshop in Sheffield 28-29 June 1997) is to show that evolutionary considerations can throw light on how we came to think and speak. So one might expect either claims about pathways and selective pressures that could have led to particular aspects of thought and language or claims about how the whole complex capacity for conceptual thought and verbal communication could arise out of more basic primate behaviors. But the aim is not really either of these. It is rather to keep the reader's faith in the power of evolutionary theory alive, while expressing a number of warnings about simplistic thinking and hasty conclusions. The central conclusion is that "each aspect of the human mind will have its own history." Don't look for a single magic factor.

The detailed argument of the paper reinforces this main conclusion by urging us to individuate aspects quite finely. In particular, not to think of conceptual thought as an aspect of language. So considerable evidence is collected which suggests that language draws on non-linguistic categorisations. Semantic space is largely structured by the patterns of thought we have developed for other purposes. (This idea connects well with two papers Pascal Boyer has presented to the workshop: or perhaps it is one paper presented twice.) At this point in Bloom's exposition he begins to emphasise the importance of the need for communication in shaping language. The argument seems to be: language didn't evolve to give conceptual thoughts, so it must have evolved for communication. But remember here Bloom's own warnings. Language is a rich system with many very distinctive aspects which more-or-less pull together when we think and talk. So looking for *the* factors driving the evolution of language is likely to be as futile as looking for unique factors driving the evolution of conceptual thought. In particular, a communicative capacity can ratchet up thought. One way in which this can happen is that communication opens up social possibilities that make demands on thought. Another is that the syntax of language may squeeze the natural shape of some pre- linguistic concepts. Another is that language can provide ways in which one conceptual module can influence another, influencing then long-term development of each. (By speaking to others we are speaking within ourselves.) An obvious instance of this third possibility is that linguistic communication stretches and pressures the concepts implicit in theory of mind, and in the related conceptualisations underlying moral life.

So never ignore co-evolution. In particular consider the co- evolution of syntax in language and thought. An image that haunts me in this connexion is Sperber and Wilson's elephant. In their book *Relevance* (Blackwell 1986) Dan Sperber and Deirdre Wilson imagine asking an elephant what the purpose of noses is. "To pick things up, of course" answers the elephant. But the elephant is wrong; noses are for filtering and warming the air an animal breathes, and elephants have developed a particular specialisation of the nose for picking things up. So, suggest Sperber and Wilson, if you ask a human what the purpose of syntax is they are likely to answer "for communicating, of course." But this is wrong; syntax is for thinking, and humans have evolved a particular specialisation of it for communication. Notice the similarities to Bloom's point of view. Communication builds on pre-existing capacities, which evolved to serve individual cognitive rather than social purposes. Note also the similar internal tension: syntax is taken for the sake of this point as a single homogeneous capacity, while the logic of the argument suggests at least as a serious possibility that non-linguistic syntax itself may be a patchwork of separate adaptations stitched together serendipity and chance.

Before returning to this last point reflect for a moment on non-linguistic syntax. Philosophers often have trouble with the thought that thoughts could have semantical and syntactical properties independently of the words that express them. One reason for the philosophers' worries is that they realise that if the ascription of conceptual content is to do any real work it must be accompanied by some suggestions about the individuation of content: when the thought of a person at one time is the same as the thought of another person (or the same person) at another time (or the same time). One test case is the thought that "o is P and o is not Q", where P and Q apply to exactly the same things. Being in a state with this as content must not be the same as being in a state with content "o is P and o is not P" or with content "o is P and o is R" where R is not the same concept as Q. The easiest way to satisfy this is to tie the possession of the thought to the disposition to express it with particular words with particular meanings. (So the thought in question is different from those other thoughts because of the particular meanings of the *words* for P and Q.) The philosophy of language, or linguistics, then becomes the mystery element that will make the whole mixture work.

Cognitive scientists don't waste a lot of time worrying about the individuation of concepts. The reason their insouciance does not bring disaster is the implicit presence of syntax. When a state with a content is ascribed to an individual it is, if things are done right, ascribed in the presence of a determinate system of symbols with a determinate set of rules for structuring their combinations. The thought that "o is P and o is not Q" is thus distinct from the thought that "o is P and o is not P" when the inferences and transformations that the rules allow treat the two differently. (I don't mean to say that this solves all the problems. The philosophers' worries are real. But it is enough to let the enterprise get moving in a coherent way.)

Now we can return to the point about the probable variety of syntax. A syntax is a data-structure, a way of holding items of information so that they can be used in specific ways. As any computer scientist will tell you, there are as many data structures as you care to invent, each with its suitability for some specific purpose. (Any mathematician will tell you that they are all just special cases of a few set theoretical combinations. But mathematicians do not understand about bounds on computation.) There is no reason to believe that the data structures found in (for example) visual perception, the categorisation of animals, social thinking, and the processing of a speech signal, will be the same. (So we should be very cautious about going from "cognition is computation" to "there is a single data structure for all cognition.")

One example of this, relevant to issues of language and thought, is linearity. Constraints of the medium make spoken speech linear: it comes in strings. But not all data structures are linear. Logicians, for example, sometimes use branching structures such as

```
(for all x) (there is a z)
                                R holds between x, y, z, and w
(for all y) (there is a w)
```

This cannot be expressed in a linear form, such as "(for all x) (for all) (there is a z) (there is a w) R holds between x,y,z, and w", because in the branching formula the choice of z depends only on the choice of x (for all x there is a z, which one depending on which x), and similarly for y and w, while in the linear formula the choice of z depends on the choice of x and of y. If you consider enough examples of sentences like this you find yourself able to think thoughts that are very difficult to think in terms of a linear syntax. (This is actually an oversimplification: you can say the branching thing in linear form if you expand your vocabulary. There may be a systematic racheting between syntax and lexicon here.)

Several ideas come together at this point. Suppose that human psychology is modular, consisting of relatively independent capacities which have evolved to serve related but differentiable purposes. Suppose that communication through speech is just one module (or complex of modules). Suppose that each module possesses its own distinctive data structures. Suppose that many of them are seriously distinct from those underlying spoken speech. And last of all suppose that speech draws on the output of many of these related and distinctive modules. Then it follows that there will be mismatches, tensions, pressures. To fit speech some kinds of thinking will have to change or develop speech-oriented interfaces; to bring some kinds of thinking into its range speech will have to modify its syntax or develop new kinds of vocabulary. The syntactical mismatch between speech and thought is likely to be a continuing source of co-evolutionary pressure on both.

(Afternote: at the Hang Seng workshop at which Paul Bloom presented this paper a dispute arose between Bloom and Derek Bickerton about the likely nature of primitive language. Bickerton claimed that it would have had a restricted syntax, and Bloom that is might have had a restricted subject matter. The point of view of this paper suggests that these may both be aspects of the same phenomenon. In order to extend the range of modules with which a speech module can interact it is likely to need a more flexible syntax. It is in fact likely to acquire a set of peculiar and quirky devices, which from their origins as ways of handling the structures characteristic of some particular modules develop to become universal, though arbitrary, features of language.)