Dickins, T.E. (2003) What can evolutionary psychology tell us about cognitive architecture? *History and Philosophy of Psychology*, *5* (1), 1-16

What can evolutionary psychology tell us about cognitive architecture?[1]

Thomas E. Dickins

University of East London

Abstract

Certain evolutionary psychologists have argued that a massively modular cognitive architecture is the necessary outcome of natural selection. This argument appears to be predicated upon three core and questionable assumptions - 1) a Classicist Computational Theory of Mind; 2) Marr's (1982) tripartite explanatory project for the cognitive sciences; and 3) the view that adaptive behaviour must come under fairly direct cognitive control. This paper argues that, under these core assumptions, there is in fact no direct support for a modular architecture from an adaptationist analysis. This is because such analyses are always at the behavioural level and there is nothing in the behavioural data to help decide between possible architectures. Behavioural data can only be used to provide an extensional functional analysis and architectural claims are to some extent intensional. This leaves only an argument from theoretical cognitive science and this too is weak. The paper closes with a discussion about whether or not an evolutionary perspective can inform models of cognitive architecture at all and the conclusion is drawn that it can. Such an approach favours ground–up modelling of functions and thereby imposes a form of parsimony on algorithmic and architectural speculations.

Keywords: Cognitive architecture, Modularity, Parsimony

Introduction

As the title suggests, this paper has a clear focus upon evolutionary explanations of cognitive architecture. I will provide a brief outline of what I term the Santa Barbara school of evolutionary psychology

[1] A version of this paper was given at the Annual Conference of the Human Behaviour and Evolution Society, Rutgers University, USA, June 2002.

(henceforth SBEP). After this I shall make some comments on the general project of cognitive science and use them to highlight some concerns I have about SBEP and its architectural claims about modular organisation. Finally I shall speculate upon the precise relationship between functional evolutionary analyses and models of cognition.

Before I embark upon any of the above I intend first to give a summary of the main argument in order to guide the reader through the paper. This summary will seem a little abstract at this juncture but as the paper progresses I hope that it will seem less so.

As with Sociobiology, SBEP applies evolutionary explanations to behaviour and these explanations are grounded within conceptions of "selfish genes" (Dawkins, 1976). However, SBEP sees a missing level of explanation within Sociobiology, that of the proximate psychological mechanisms that mediate between selfish genes and adaptive behavioural phenotypes. The clearest statements about this form of evolutionary theorising have been made by Leda Cosmides and John Tooby (both at the University of California, Santa Barbara; Cosmides, Tooby and Barkow, 1992; Cosmides and Tooby, 1994) who have argued that adaptive problems in our ancestors' past will have led to the selection of cognitive abilities that were able to solve these problems effectively.

The argument from SBEP can be made starker – natural selection led to the emergence of adaptive behaviours within *Homo sapiens*. Adaptive behaviours, by their very nature, are domain-specific as they solve a particular adaptive problem and no other (disregarding arguments from exaptation, and subsequent secondary adaptations for now, which are not relevant to our central concern – see Andrews, Gangestad and Matthews (forthcoming) for a discussion about exaptation and evolutionary psychology). Experimental work in SBEP supports the notion of highly specialised behavioural abilities that meet adaptive problems entailing a psychology that enables the regular production of these adaptive behaviours. SBEP further assumes that the cognitive architecture underpinning the behaviour is organised along similar, almost unit-similar, domain-specific lines, such that specific cognitive mechanisms deliver specific adaptive behaviours.

In this paper I shall maintain the following assumptions:

1. If cognition is organised in a domain-specific fashion then the only way that organisation could have been instantiated is through natural selection;
2. If natural selection is true then a predominantly domain-specific cognitive architecture is not the only possible outcome.

SBEP shares assumption 1 with me but not 2. Indeed, advocates of SBEP believe in the converse assumption:

3. If natural selection is true then a predominantly domain-specific architecture is the only outcome.

I maintain that descriptions of cognitive architecture are by their very nature intensional hypotheses, as they are about properties of the mind. Behavioural data is purely extensional and cannot be used to support a particular intensional supposition. In other words, there is nothing in the behavioural data that can confirm or deny the architectural hypothesis of SBEP: it is possible that another form of architecture could produce the same behaviour. Later on I shall indicate some alternatives that rely on a more pressing role for recurrent environmental structure. This leaves us looking for another reason to favour modular or domain-specific architectures that would allow us to make better practical use of my first assumption. Such reasons exist within the theoretical cognitive science literature and they will be discussed. Unfortunately they are not conclusive and therefore do not allow us to immediately opt for assumption 1.

This paper is not to be read as an anti-modularity paper but rather as a caution against a wholesale top-down approach to cognition from evolutionary psychology. As assumption 1 will indicate, I take natural selection very seriously and I think it is more than reasonable to assert its truth, but as with all good things it should be used sparingly in order to build theory in other areas. I shall reflect on such matters in the closing sections of this paper.

The Empirical Case from Santa Barbara

Cosmides and Tooby have provided empirical support for their claims through work on the Wason Selection Task (WST; Cosmides 1989; Cosmides and Tooby, 1992). The WST in its original form asks participants to check an abstract conditional rule of the form $P \rightarrow Q$, where four cards represent P, Q, not-P and not-Q. The standard finding is that participants generally attempt verification and check the P and Q cards. However, when the task was set in a narrative, that represented the potential violation of a social contract, with the cards representing people and potential actions, performance was significantly improved with participants generally attempting falsification, which is the logically correct procedure.

This finding was used to support a number of claims. First, it was argued that humans would be sensitive to cheats. Social contracts are the stuff of group cohesion and any individual cheating threatens to destabilise the finely balanced economy of the group, which could be disastrous for the individuals that comprise that group. However, individuals will always try to maximise their fitness and cheating can facilitate this. Such Machiavellian pressures are thought to have led to highly adapted social cognitions within many primate groups.

The second claim is that detecting cheats is a domain-specific adaptive problem and this will have led to the selection of a contentful, domain-specific cognitive faculty or set of faculties to solve it. Within the psychological literature the term "modular" has been used to denote domain-specificity of function for any given proximate psychological faculty. As such, the subsequent claim from SBEP, that much of the mind has been organised in this way through natural selection, has been dubbed the Massive Modularity hypothesis. "Modularity" is not a term explicitly used by Cosmides or Tooby but all of their architectural descriptions conform to modularity theories.

The term "massive" has been attached to this modularity claim to emphasise an historical transition in thinking about modules. In the recent psychological literature modularity dates back to Fodor (1983), who introduced the idea of modular organisation for peripheral, input functions. Fodor characterised these modules as having some of the following properties:

Encapsulated – it is impossible to interfere with the inner workings of a module.

Unconscious – it is difficult or impossible to reflect on the operations of module.

Speedy – modules are very fast.

Shallow outputs – modules provide limited output, without information about the intervening steps that led to that output.

Obligatory firing – modules operate reflexively, providing predetermined outputs for predetermined inputs regardless of the context.

Ontogenetic universals – modules develop in a characteristic sequence.

Localisation – modules are mediated by dedicated neural systems.

Pathological universals – modules break down in characteristic fashion following insult to the system.

Domain specificity – modules subserve a specific function, they operate in one domain.
(Adapted from Elman et al., 1996)

This form of modularity was proposed in order to explain empirical evidence about the functional isolation of input devices, such as those of the visual system. Fodor went on to argue that higher-order cognition could not be modular because of its holistic properties that allow for such phenomena as abductive inference and analogical thinking. As such this was a purely computational argument about mechanisms – modules consisted of computations that processed domain-specific inputs in a certain way; higher-order thinking processes were still computational but also domain-general or global (taking many inputs). Fodor despaired of ever understanding the computations that determine higher-order cognition. The assumption from the evolutionary perspective discussed here is that at least a significant part of higher-order cognition is also organised into domain-specific modules and, as such, this represents a *massive* proliferation of modules compared with Fodor's early speculations.

There are two possible forms of the Massive Modularity hypothesis – a weak form and a strong form (Samuels, 2000). Strong massive modularity would argue that the whole of the mind is constructed of modules. As Samuels makes clear this position is not explicitly defended in the literature and it would be hard to defend, as it would rule out some of the flexibility we see in human cognition – such as reasoning by analogy. Thus there is a case for some domain generality in the mind, but also an argument from adaptationism for a good deal of Darwinian modularity too. Tooby and Cosmides make this clear:

> What is special about the human mind is… that it proliferated "instincts" – that is content-specific problem-solving specializations – which allowed an expanding role for psychological mechanisms that are (relatively) more function-general. These are presently lumped into categories with unilluminating labels such as "the capacity for culture," "intelligence," "learning," and "rationality." It is time for the social sciences to turn from

> a nearly exclusive focus on these embedded, more function-general mechanisms to a wider view that includes the crucial, and largely neglected, superstructure of evolved functional specializations. Equally, we need to explore how the two classes of mechanisms are interwoven so that their combined interactive product is the zoologically unique yet evolutionarily patterned breadth of functional behaviours. (pp. 113 – 114)

Here we see an explicit claim for evolved content-specific cognitive organisation being served by fairly domain-general computations. Given the presence of some domain-generality we can only see this position as weak massive modularity.

There is a potential confusion that could arise from reading the above quotation from Tooby and Cosmides. It has briefly been noted that Fodor's original work argued for computational modules that operated on specific inputs. Tooby and Cosmides have stated, in this quotation, that modules are to be seen as content-specific, not computationally specific, which is a different idea. Samuels, Stich and Tremoulet (1999) have also defined Fodorian modules as computational, and they contrast them with "Chomskyan" modules, which are a domain-specific set of *representations*. This conception arose from Chomsky's notion of a language faculty instantiating innately specified grammatical principles. Samuels et al. note that various developmentalists have also posited Chomskyan modules for systems of principles of physics, psychology and mathematics (see for example Carey and Spelke, 1994). The modules of Carey and Spelke are theory-like, but Samuels et al. point out that Chomskyan modules need not be theory-like (i.e. truth evaluable) but could for instance be "how to" systems of representations that demand no evaluation.

At this juncture it looks as if SBEP is arguing for a weak Massive Modularity hypothesis where modules are to be seen as content-specific (or Chomskyan), their content being established through natural selection and computed by some domain-general mechanisms. In the above quotation from Tooby and Cosmides it is noteworthy that they qualify the notion of function-general mechanisms with the term "relatively", thereby implying some restrictions on the operation of these processes. Elsewhere in the same paper they make more forceful claims about the nature of the computations that operate over the Chomskyan architecture.

> General mechanisms turn out to be very weak and cannot unassisted perform at least most and perhaps all of the tasks humans routinely perform and need to perform. Our ability to perform most of the environmentally

> engaged, richly contingent activities that we do depends on the guiding presence of a large number of highly specialized psychological mechanisms… Far from constraining, specialized mechanisms enable competences and actions that would not be possible were they absent from the architecture. This rich array of cognitive specializations can be likened to a computer program with millions of lines of code and hundreds or thousands of functionally specialized subroutines. (Tooby and Cosmides, 1992: 39)

Here Tooby and Cosmides have clearly made an argument for *computational* specialization that is not at odds with a Fodorian view of modularity, and one that does not rely upon content-specific bodies of knowledge, as in a Chomskyan architecture. None the less, this description is still compatible with a weak Massive Modularity hypothesis.

What are we to think about SBEP and the massive modularity hypothesis? A harsh critic would claim confusion for the position but a more charitable one would argue that SBEP proposes weak massive modularity, whereby modules are to be seen as content-specific and computationally-specific. In keeping with a weak version, there is also room for some domain-general computations that would allow for the global properties of cognition that concerned Fodor (1983). The rest of this paper will assume the more charitable reading.

Behavioural Data

The claims that SBEP makes for modularity are based on behavioural data. The work on the WST demonstrated a particular facilitation effect that was domain-specific. Humans appear to have an ability to perform conditional reasoning well for a social problem but not an abstract one. The principal assumption of the WST work is that the logic of the abstract and social version of the task is the same; however, Fodor (2000) has recently taken issue with this. Fodor uses two examples, one of a rule from an abstract version of the task, and one from a social version. The abstract rule example is "if a card has a vowel on one side then it has an even number on the other" and is of the form $P \rightarrow Q$. The social rule example is "it's required that if someone is under 18 (s)he drinks coke". According to Fodor this is a deontic form of the task. The social contract that is represented is one of legal age limits for alcohol consumption. However, if we want a coke rule that is of the form $P \rightarrow Q$, then we would be better writing it as "if someone is under 18 (s)he is drinking coke" (this is an indicative conditional version). The addition of phrases such as "it's required that" changes the logic from one of a conditional P→Q to one "where all that P does is determine on whom the prohibition falls" (Fodor, 2000: 30).

Fodor's argument then is that the performance differences noted by Cosmides and Tooby on the WST are due to differences in the tests. This behavioural difference is not evidence of a special cheat-detection mechanism but instead it is, at best, evidence of subjects using different "inferential paths". Fodor does not think that the fact that we find deontic conditional reasoning easier than indicative conditional reasoning is in want of an evolutionary explanation. He thinks that these material differences are sufficient explanation.

Fodor’s (2000) point is a purely logical point about the validity of the experimental assumptions. It has some bite with regard to SBEP, as the WST work is the principal experiment supporting the whole modular view of an evolved cognitive architecture. The motivation for experiments within the cognitive sciences must be to explain, at the very best, or clarify the nature of, at the very least, a specific, regularly observed behaviour. In the case of the Wason work the initial phenomenon was the poor performance on the abstract indicative conditional version of the task. Cosmides’ and Tooby’s insight was to theorise about possible evolutionary explanations for the emergence of reasoning and to suggest that this might explain any biases observed. In short, it could be that we are only good at certain forms of reasoning because they were critical to life in the evolutionary past. The subsequent experiments do indeed seem to have uncovered a possible bias in favour of either social reasoning or, perhaps more precisely, deontic reasoning, and this has to be of interest in light of the very least aim of experiments within the cognitive sciences.

SBEP and Cognitive Science

If we leave Fodor (2000) to one side and assume that the WST findings are robust and that the materials well matched, we must now ask whether or not domain-specific behavioural differences are necessarily underlain by a similar cognitive architecture. In order to answer this we need to look at the view of cognition that is at work in SBEP.

Tooby and Cosmides (1992: 75; Cosmides and Tooby, 1992: 178) make explicit reference to Marr’s (1982) tripartite explanatory framework for the cognitive sciences. According to this view in order to have a full explanation of cognition we need to provide:

1. A functional decomposition that asks what the function of the particular cognition is
2. An algorithmic characterisation that hypothesises a formal procedure that can produce the function described in level 1
3. A physical implementation that realises the algorithm in the neural substrate

Marr's framework has been influential within the cognitive sciences and SBEP aims to provide a functional decomposition based on the adaptive function of a variety of cognitions.

The cognitive sciences have been dominated by the Computational Theory of Mind (CTM) and naturally a Marrian perspective has been brought to bear upon the CTM. CTM states that the mind is a computer, a physical symbol manipulator, consisting of a set of syntactically structured representations that are subject to formal procedures, or algorithms operating serially over them in a localised fashion (see Fodor 2000). Algorithms are intensional definitions of computable functions, recipes for producing a specified input-output relation. Turing introduced the notion of using a universal computer to run any number of different algorithms and, given the coupling between algorithm and computation, effectively implementing a number of different sub-machines within one piece of hardware. CTM sees the brain as an idealised Turing Machine. The hypothesis that our minds consist of functionally specialized subroutines, shaped by natural selection, that is referred to by Tooby and Cosmides in the quotation on p X is an example of this view.

Marr argued that a higher level of explanation was independent of those below it, such that information about function and algorithm could not be informed by neuroscience. Functional decomposition provides an extensional definition of ***what*** the function is; the algorithmic level has to specify ***how*** that function is produced. A universal Turing Machine can run any algorithm so individuated. Despite this higher-level independence claim, Marr (1982: 27) argued that, "an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied." But this is seen to be a question of determining the algorithm's goal, not as a belief that functional information can specify the exact algorithm. Given Turing computability, there are an infinite number of possible algorithms that could do the job.

Under this conception any behaviour that a cognitive system produces can be implemented in numerous ways, such that the behaviour itself cannot help to differentiate between hypotheses about algorithms.

Not all theorists agree with Marr's independence claim. Churchland and Sejnowski (1992) have argued that the neural hardware is not a uniform substrate but has many different levels of organisation from molecular structure through synaptic structure to neurones and beyond. All of these physical differences might act to constrain the kinds of computable functions possible within the brain. But, as Kennair (2002: 25) has recently pointed out:

> Cosmides and Tooby … argue against the idea that empirical findings of "low-level" neuroscience "will place strong constraints on theory formation at the cognitive level". They doubt that the properties of neurophysiology will lead to the discovery of cognitive programs, due to the fact that *"[t]he same basic neural tissue embodies all of these programs"*.

Kennair goes on to outline SBEP's commitment to evolutionarily based functional decomposition as a route to delimiting the number of possible algorithms that might compute a given function. Indeed, Kennair makes it clear that this claim is of the strong form that suitable evolutionary functional decomposition can leave us with an empirically testable set of alternative algorithms. Kennair sees this as a result of limits on functionality set by natural selection – i.e. a result of a finite number of adaptive problems.

It would appear that SBEP is committed to CTM and, if Kennair is correct, the Marrian project in its pure form. None the less, looking at Cosmides' and Tooby's (1992) treatment of the classic Prisoner's Dilemma work on reciprocity (see below), what they produce is a set of design features that any putative algorithms would have to satisfy. This is quite different from producing an actual algorithm to perform reciprocal analyses. Indeed, it amounts to no more than a finessed functional decomposition, which is useful but again could be implemented by a vast number of different algorithms.

SBEP's Architectural Claims

Architectural claims about modularity are of a different order from claims about possible algorithms. Algorithms can be seen as functionally specific organisational elements of a computer but modularity refers to the organisation of representations, computations, or both, into "packages" that deal with a specific domain. So, the computational notion of modularity is about a set of algorithms, or sub-computers, devoted to a specific (adaptive) function. This overall function is likely to consist of a number of sub-functions, each of which could be serviced by a specific algorithm. It might seem that there is little space between the concept of modularity and the concept of algorithms – both being about functionally specialized computation. Why not look at behavioural data of the sort produced by the WST work and simply argue for specific social reasoning algorithms? What does modularity add to our understanding?

The notion of modularity has also arisen in the Artificial Intelligence literature in response to a specific problem that serial computational machines face. This is sometimes referred to as the frame problem, which is the problem of how a cognitive system can know which representation to access when confronted by a problem space that demands action.

Dennett (1992) neatly describes the number of assumptions that an Artificial Intelligence (AI) machine would require for solving a relatively simple problem such as making a midnight snack. These assumptions include knowledge about what is in the fridge, what ingredients can be successfully combined, what tools one needs to construct a meal, the fact that the physics of the universe allows food to be carried on a plate etc. The amount of assumptions that need to be ruled in and ruled out of consideration soon appear intractable after a few minutes' thought – although we know that it is tractable as we have all made midnight snacks successfully and, importantly, without much time spent considering what to do.

The frame problem is problematic for AI because any intelligent machine that is built to operate even in a very small toy domain has to be able to compute the relevant information successfully. Dennett offers a second example of a robot that has to save its own battery from a room with a time bomb in it. The first generation robot goes to the room, sees the battery on a wagon and pulls it out of the room. Unfortunately the bomb is on the wagon too and it explodes. It is at this point that the AI team realise they need a robot that does more than simply remove the battery from its location: it must weigh up the consequences of its actions. A second generation is developed, but this goes the way of its forebear because it spends too long weighing up all the potential consequences and non-consequences of its possible actions. For instance it might go through "thoughts" about the effect that removing the battery on the wagon will have on the football results next week. When it realises this will have no effect, the timer on the bomb reaches zero and the game is over.

The robots above fail because they are machines that search a vast space serially. The consequence of such local computation is that it is time-consuming when dealing with such spaces. One possible solution is to program the robot with domains, or specialised frames of reference, that reduce the amount of searching through representations by zeroing in on a narrower search space. This is the modularity solution. Each module is internally a serial machine and still performing local computation but this architecture makes for an overall system that is not generalist. In other words, the overall machine has a bespoke architecture that reflects the variety of problem spaces it will encounter, and at each such space the relevant modules switch on and perform the necessary serial computations for the appropriate input-output function. SBEP argues that, in nature, such bespoke tailoring can only be carried out by natural selection, which is a version of assumption 1. (For alternative views on ontogenetic tailoring, see Dickins and Levy, 2001; Elman et al., 1996; and Quartz and Sejnowski, 1997.)

It is worth noting that the frame-problem makes content a prime focus – modules are contentful, allowing the system to zero-in on the appropriate set of hypotheses to help solve a problem. As outlined above content is also a prime focus for a Chomskyan perspective within SBEP, but SBEP also appears to argue for computational specificity. None the less, one might expect a contentful module to have its own dedicated processing system too.

Fodor (2000) notes that a massively modular mind could only successfully avoid the frame problem if the appropriate modules were activated when a given problem space was encountered. Fodor looks at a number of ways of activating a massively modular mind that runs on CTM principles. For instance, one could have a set of modules that translated input into the appropriate form of representations to then be fed into the relevant higher-order module. So, a cheat detection module would be fed with representations of social contracts, which had been extracted from the input received, and this module would discern whether or not the contracts were being violated. But what process governs the arrangement of input into representations that can then be fed onto relevant higher order modules? As a strong massively modular theorist one would have to propose another set of modules and this would begin the decline into an infinite regress. In short there seems no way around falling into this kind of regress if one rigorously adheres to the defining principle that all is modular. If one does not, and instead takes a *weaker* form, then one can build in some order of domain-general thinking system that can augment some modularity. *But* a domain-general system that can take input, and infer which module to activate in order to produce a relevant response, is one that is subject to the frame problem. Fodor concludes that we are forced to take some version of domain-general CTM and accept that it is explanatorily inadequate in the face of the frame problem. So, we could envisage a mind with a good deal of modularity that is supported by a domain-general computational system. However, invoking domain generality at any stage in this argument for modular architecture considerably weakens this architectural hypothesis, as it re-introduces the frame problem.

As we have seen, SBEP argues for weak massive modularity and this introduces some domain-generality into the proceedings. It would further appear from the initial quotation from Tooby and Cosmides (p. X) that this generality is a computational trait, not a content-trait. Given the argument from the AI field we can see that this domain-general computational capacity has to account for global cognition and get around the frame problem.

Tooby and Cosmides (1992: 105–106) do use the above arguments from AI to support their claims:

> It is the perennial hope of SSSM advocates within the psychological community that some new technology or architecture (wax impressions, telephone switching, digital computers, symbol-processing, recursive programming languages, holograms, non-von Neumann architectures, parallel distributed processing – a new candidate every decade or so), will free them to return to empiricism, associationism, domain-generality and content-independence (where SSSM tells them they should go). Never the less, the functional necessity of content-specificity emerges in every technology because it is a logical inevitability. Most recently, researchers are establishing this all over again with connectionism…. the frame problem (*is an obstacle*) that can only be overcome by endowing computational architectures with contentful structure. This is because the world itself provides no framework that can decide among the infinite number of potential category dimensions, the infinite number of relations, and the infinite number of potential hypotheses that could be useful to analyse it. (Italics added.)

This is clearly not an argument from evolutionary theory. It would appear that SBEP holds the view that modularity is a solution to a problem suffered by serial-computational systems and that we can look for the domains of our modules through an evolutionarily inspired functional decomposition. This assumes that functional decomposition at the behavioural level matches functional organisation at the cognitive level. In other words, there is an assumption that an extensional description of behaviour will inform us about the organisation of our mental contents and associated computations. This amounts to a stronger version of my first assumption, which could now be read as:

1.1 Cognition has to be organised in a domain-specific fashion and the only way that organisation could have been instantiated is through natural selection

The evolutionary analysis provides a description of the domains. However, there is still support for assumption 3, which is not incompatible with assumption 1.1.

Adaptive Behaviour and Direct Cognitive Control

> Natural selection permits the evolution of only certain strategies for engaging in social exchange. To be selected for, a design governing reasoning must embody one of these strategies – in other words, it must meet an "evolvability criterion"…. (Tooby and Cosmides, 1992:170)

SBEP makes explicit reference to game-theoretic analyses of the emergence of certain strategies. In particular, Cosmides and Tooby (1992) discuss the classic work on one-shot and iterated Prisoner's Dilemma games that were used to model the emergence of various evolutionarily stable strategies. Evolutionarily stable strategies are defined as those strategies that, once in place, cannot be supplanted by another. This claim is not the claim that other strategies will not exist alongside the evolutionarily stable strategy, but simply that any other strategy cannot beat it, given the defined constraints of the evolutionary game being played. Thus an evolutionarily stable strategy is an optimal strategy, given these constraints. Within behavioural ecology, sociobiology and evolutionary psychology it is important to note that the strategy part of the term refers to behaviour, not to a derived plan. Such strategies do not necessarily come under the control of conscious or deliberative systems.

Cosmides and Tooby (1992) look at the canonical game-theoretic work of Axelrod, Hamilton and Trivers in the domain of reciprocal altruism. They argue that reciprocity must be underpinned by "cognitive programs that actually evolved in the human lineage" (1992: 176-177) and this is tantamount to an endorsement of assumption 3. As discussed above, the uncovering of such programs is seen as an empirical question and this question has its focus narrowed by paying attention to the nature of the evolutionarily stable strategy in question. So, the claim is that only certain evolutionarily stable strategies can emerge in a given niche. Those cognitive mechanisms that allow the production of these strategies will be selected for. The level of cognitive internalism inherent within this hypothesis is held to be tenable for reasons relating to a commitment to CTM and the frame problem argument. It is regarded as computationally intractable for a domain-general architecture to arrive at an evolutionarily stable strategy, given the vast number of potential behavioural hypotheses it could generate for any niche. Accordingly, adaptive behaviour has to come under direct cognitive control. However, as Grantham and Nichols (1999: 60) point out:

> (S)ome adaptive problems don't even require a *cognitive* solution. For some animals, the adaptive problems of predator-avoidance and aggression-deterrence are solved by physiological features. And in humans, it's possible that the adaptive problem of attracting potential (mates) was solved by physiological features.

It is perfectly possible that some of our adaptive problems were cognitively solved, and others were not, but which were and which were not is in itself an empirical question. It might be the case, for example, that the evolutionarily stable strategies outlined in the work on reciprocal altruism are implemented not by specific algorithms but by simpler associative learning mechanisms. Furthermore, in line with the assumptions of Turing computability, Grantham and Nichols (1999) point out that "each of these … behavioural solutions can be cognitively realised in a multitude of ways", which is the problem of multiple-realisation that allows for a given function to be carried out by multiple lower-level mechanisms. Evolutionary theory will not help us to decide between which of the possible realisations has been instantiated[2].

There is a related problem. As implied, the emergence of the tit-for-tat strategy in the iterated Prisoner's Dilemma game could arguably depend only upon an organism with the ability to register reward and punishment, and the ability to associate those things with salient contingencies. The innate architecture required would be sensitivity to the tokens used in the game as rewards or punishments; beyond this the recurrent structure of the world, i.e. the constraints and basic learning skills (which are importantly domain-general) would be sufficient to produce tit-for-tat. Indeed, we might be happier explaining the reciprocal altruism we see in vampire bats in these sorts of terms rather than in complex Marrian ones. Obviously those organisms that learn better (quicker, more accurately) will be favoured and this will have an effect upon relative gene frequencies too – but not upon specific genes for specific behaviours, because there are no such genes in this scenario and no such devoted behaviour production mechanisms either. How could an evolutionary argument of the order proposed by SBEP help us to decide this issue?

It must be remembered that the above arguments are all about the general lessons that SBEP seeks to draw about doing evolutionary cognitive science. It may well be the case that the constraints on human reciprocal altruism have been unchanging for sufficient time to allow natural selection to winnow away at the variance in behaviours such that there are very few strategies open to us. This is an open question. The answer to this question of the limits on strategic availability does not entail a specific cognitive argument, for that too is an open question. All told, for every behaviour we have to carefully look to the intricacies of the case and answer these questions – there is no overall epistemological blueprint to follow save always asking them and not being surprised if each case varies significantly. It is also critically important to realise that the SBEP position does not boil down to a finessed evolutionary argument at all, but to a specific cognitive position about the nature of cognition, the frame problem and the notion of direct cognitive control of adaptations. All of these assumptions are questionable and quite possibly wrong. None of these questions can be answered by evolutionary theory, because all this can tell us about is the optimality or not of a given behaviour, and none can be directly tackled by data about behaviour, for these

[2] As Grantham and Nichols (1999: 60) point out Cosmides and Tooby do concede this point, which makes their commitment to this program unusual.

very assumptions are about the production of that self-same behaviour. The extensional description will not tell us about the intensional mechanism.

If this bleak conclusion is true, then what of the role of evolutionary theory within the cognitive sciences?

Some Concluding Comments on Parsimony

Lurking in the above discussion is the idea that the recurrent structure of the world might actually play more of a guiding role in the production of behaviour than previously thought by SBEP theorists. This argument has been clearly made elsewhere (Ariew, 1999; Cummins and Cummins, 2000) in terms of the role of canalisation. Canalisation models assume less innate structure as a result of evolution than SBEP and argue that this innate structure forces certain behavioural outcomes by delimiting possibilities given recurrent structure. Such models can rely upon domain-general learning processes as well as innate structure.

Recent cognitivists (e.g. Karmiloff-Smith, 1992) have taken similar canalisation positions and discussed ontogeny as an interaction between minimal innate constraints and recurrent external structure leading to a modularised mind. Such modularity is very different from Darwinian modularity and is tantamount to the claim that humans are capable of developing specific behavioural skills and the underlying cognitive control of them. What is more, this literature is steeped in a connectionist view of cognitive processing rather than a CTM one (cf. Elman, 1993; Elman et al., 1996).

It would be very easy to lapse into assumptions about just how much recurrent external structure is required to produce a given trait, and just how much innate constraining will force a connectionist architecture to pump out appropriate behaviour, in much the same way as SBEP has simply assumed no role for recurrent external structure other than as the adaptive niche in which the organism in question finds itself. For SBEP all that counts is the degree of truth contingency between the innate internal cognitive economy of the organism and the recurrent external structure. Just so long as the latter is represented in its entirety by the former, then all is well. But it is perhaps this very issue that evolutionary theory can help us decide upon – not upon what the exact nature of the internal cognitive architecture is, but upon what degree of innate structure is minimally required.

To put this another way, we can use evolutionary theory to impose theoretical parsimony upon our cognitive speculations. This is, of course, already being done in the form of computational modelling work, perhaps most notably in the field of language origins (for examples see Knight, Studdert-Kennedy and Hurford, 2000) and also evolutionary robotics (see Hendriks-Jansen, 1996, and Nolfi and Floreano, 2000). The crucial difference between this approach and that of SBEP-style theorists is that what are being modelled is putative phylogenies, that do not have to be "true", in order to produce minimal cognitive architectures to perform a given function. In the order of approach discussed so far in this paper, evolutionary theory has taken on a top-down role in terms of supporting core assumptions about how minds must look. As we have seen these assumptions are questionable. Evolutionary modelling, from the bottom up allows a more empirical test of assumptions in terms of producing the most parsimonious model possible. By reducing the number of starting constraints in a modelled agent we can see just how well this agent fares in an adaptive problem space and see just how much structure we need to add to improve performance. The nature of the structure – connectionist or serial computational – is down to the pre-theoretical biases of the researchers to some large extent, although it might be possible to begin thinking about the emergence of such architectures over evolutionary time too. For now, such architectures can be pitted against one another to see which can produce the more parsimonious account.

The use of parsimony is perhaps best seen as a post hoc extra-theoretical commitment within science, a commitment that is not always applied (Boyd, 1991). It is used to help decide between theories that explain the same data equally well and that fit the general metaphysical assumptions abroad at the time. Within the biological and cognitive sciences it is not the case that organism in question will be the optimum design, and hence most parsimonious, as it will be the result of an evolutionary history of optimal design "decisions". The application of parsimony to decide between two possible theories cannot be performed, then, without a view on the possible phylogenetic histories of each model. It is at this point that more standard applications of parsimony might be useful, at the point of deciding between assumptions about the evolutionary economics that governed both the establishment of an adaptive problem and the abilities of the organism in question before the selection event. This use of parsimony and the establishment of an appropriate metric to so do presents a hard problem and one that I cannot pretend to have an answer for.

Summary

This paper has argued that the architectural commitments of SBEP are in large part the product of a specific commitment to the computational theory of mind. Furthermore, the paper has shown how this and related positions are questionable. More crucially than this, the paper has argued that there is nothing in evolutionary theory that will help us to decide between views on computation and cognition. More positively, I have suggested that perhaps evolutionary considerations can be used as a parsimonious guide to theory building and perhaps can aid us in decisions about the appropriate strength of our commitment to nativism, but this in itself is a difficult problem that requires work.

Acknowledgements

I should like to thank Elizabeth Valentine and two anonymous reviewers for their useful comments on an earlier draft. I should also like to thank Leda Cosmides for an enlightening conversation about the matters discussed in this paper, which took place at the annual conference of the Human Behaviour and Evolution Society in 2002. Despite these inputs all errors within this paper are entirely my own.

References

Andrews, P W, Gangestad, S W & Matthews, D (Forthcoming): Adaptationism – How to carry out an exaptationist program. Behavioral and Brain Sciences

Ariew, A 1999: Innateness is canalisation: In defence of a developmental account of innateness. In: Hardcastle, V G (ed) Where Biology Meets Psychology: Philosophical Essays. Cambridge MA: MIT Press

Boyd, R 1991: Observations, explanatory power and simplicity: Toward a non-Humean account. In: Boyd, R, Gasper P & Trout, J D (eds) The Philosophy of Science. Cambridge, MA: MIT Press

Carey, S & Spelke, E 1994: Domain-specific knowledge and conceptual change. In: L A Hirschfeld & S A Gelman (Eds) Mapping the Mind: Domain Specificity in Cognition and Culture Cambridge: Cambridge University Press

Churchland, P M & Sejnowski, T J 1992: The Computational Brain. Cambridge, MA: MIT Press

Cosmides, L 1989: The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition, 31, 187-276

Cosmides, L & Tooby, J 1992: Cognitive adaptations for social exchange. In: Barkow, J H, Cosmides L & Tooby J (eds) The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford: Oxford University Press

Cosmides, L & Tooby, J 1994: Beyond intuition & instinct blindness: toward an evolutionarily rigorous cognitive science. Cognition, 50, 41-77

Cosmides, L, Tooby, J & Barkow, J H 1992: Evolutionary psychology and conceptual integration. In: Barkow, J H, Cosmides L & Tooby J (eds) The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford: Oxford University Press

Cummins, D D & Cummins, R 1999: Biological preparedness and evolutionary explanation. Cognition, 73, 37-53

Dawkins, R 1976: The Selfish Gene. Oxford: Oxford University Press

Dennett, D C 1992: Cognitive wheels: The frame problem of artificial intelligence. In Boden, M (ed) The Philosophy of Artificial Intelligence. Oxford: Oxford University Press

Dickins, T E & Levy J P 2001: Evolution, development and learning - a nested hierarchy? In: French, R M & Sougné, J P (eds) Connectionist Models of Learning, Development and Evolution: Proceedings of the Sixth Neural Computation and Psychology Workshop. London: Springer-Verlag.

Elman, J L 1993: Learning and development in neural networks: The importance of starting small. Cognition, 48, 71-99

Elman, J L, Bates, E A, Johnson, M H, Karmiloff-Smith, A, Parisi, D & Plunkett, K 1996: Rethinking Innateness: A Connectionist Perspective on Development. Cambridge MA: MIT Press

Fodor, J 1983: The Modularity of Mind. Cambridge MA: MIT Press

Fodor, J 2000: The Mind Doesn't Work That Way. Cambridge MA: MIT Press

Grantham, T & Nichols, S 1999: Evolutionary psychology: Ultimate explanations and Panglossian predictions. In: Hardcastle, V G (ed) Where Biology Meets Psychology: Philosophical Essays. Cambridge, MA: MIT Press

Hendriks-Jansen, H 1996: Catching Ourselves in the Act: Situated Activity, Interactive Emergence, Evolution, and Human Thought. Cambridge MA: MIT Press

Karmiloff-Smith, A 1992: Beyond Modularity: A Developmental Perspective on Cognitive Science. Cambridge MA: MIT Press

Kennair, L E O 2002: Evolutionary psychology: An emerging integrative perspective within the science and practice of psychology. Human Nature Review, 2, 17-61

Knight, C, Studdert-Kennedy, M & Hurford, J R (eds) 2000: The Evolutionary Emergence of Language: Social Function and Linguistic Form. Cambridge: Cambridge University Press

Marr, D 1982: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: Freeman

Nolfi, S & Floreano, D 2000: Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines. Cambridge MA: MIT Press

Quartz, S R & Sejnowski, T J (1997) The neural basis of cognitive development: A constructivist manifesto. Behavioral and Brain Sciences, 20, 537-596

Samuels, R (1998) Evolutionary psychology and the massive modularity hypothesis British Journal of the Philosophy of Science, 49, 575- 602

Samuels, R (2000) Massively modular minds: the evolutionary psychology account of cognitive architecture In: Carruthers, P & Chamberlain, A (eds) Evolution and the Human Mind: Modularity, Language and Meta-Cognition Cambridge University Press

Samuels, R, Stich, S & Tremoulet, P D 1999: Rethinking Rationality: From Bleak Implications to Darwinian Modules. In: LePore, E & Pylyshyn, Z (eds) What is Cognitive Science? London: Blackwell

Tooby, J & Cosmides, L 1992 The psychological foundations of culture. In: Barkow, J H, Cosmides L & Tooby J (eds) The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford: Oxford University Press