Abstract

Cognitive Evolutionary Psychology Without Representational Nativism

Denise Cummins and Robert Cummins

University of California-Davis

Pierre Poirier

University of Quebec at Montreal

A viable evolutionary cognitive psychology requires that specific cognitive capacities be (i) heritable and (ii) "quasi-independent" from other heritable traits. They must be heritable, because there can be no selection for traits that are not. They must quasi-independent from other heritable traits, since adaptive variations in a specific cognitive capacity could have no distinctive consequences for fitness if effecting those variations required widespread changes in other unrelated traits and capacities as well.

These requirements would be satisfied by innate cognitive modules, as the dominant paradigm in evolutionary cognitive psychology assumes. But those requirements would also be satisfied by heritable learning biases, perhaps in the form of architectural or chronotopic constraints, that operated to increase the canalization of specific cognitive capacities in the ancestral environment (Cummins and Cummins 1999). As an organism develops, cognitive capacities that are highly canalized as the result of heritable learning biases might result in an organism that is behaviorally quite similar to an organism whose innate modules come on line as the result of various

1

environmental triggers. Taking this possibility seriously is increasingly important as the case

against innate cognitive modules becomes increasingly strong.

Cognitive Evolutionary Psychology Without Representational Nativism

Denise Cummins and Robert Cummins

University of California, Davis

Pierre Poirier
University of Quebec at Montreal

I. Introduction

Does a viable evolutionary cognitive psychology require the assumption of innate cognitive

modules?

A viable evolutionary cognitive psychology requires that specific cognitive capacities be (i)

heritable and (ii) 'quasi-independent'.[1] They must be heritable, because there can be no

evolutionary response to selection for traits that are not. They must also be quasi-

independent—i.e., there must be additive genetic variance–because adaptive variations in a

specific cognitive capacity could have no distinctive consequences for fitness if effecting those

variations required widespread changes in other traits and capacities as well. If, for example, the

emergence of a *theory of mind* in young children is simply the consequence of the operation of a

general purpose learning device, then the adaptive consequences of having a theory of mind

(assuming there are some) would simply be a factor in the selection of the general purpose

learning device in question. While this sort of relation between evolution and cognition would not be completely trivial, it would hardly justify a new sub-discipline of cognitive evolutionary psychology.

The dominant paradigm in evolutionary cognitive psychology attempts to satisfy these constraints by proposing that the mind can be characterized as a collection of *innate cognitive modules*. This position is succinctly captured by the following quotation:

> [O]ur cognitive architecture resembles a confederation of hundreds or thousands of functionally dedicated computers (often called modules) designed to solve adaptive problems endemic to our hunter-gatherer ancestors. Each of these devices has its own agenda and imposes its own exotic organization on different fragments of the world. There are specialized systems for grammar induction, for face recognition, for dead reckoning, for construing objects and for recognizing emotions from the face. There are mechanisms to detect animacy, eye direction, and cheating. There is a 'theory of mind' module .... a variety of social inference modules .... and a multitude of other elegant machines. (Tooby and Cosmides, 1995, pp. xiii-xiv)[2]

Since what is innate is heritable, the requirements of heritability and quasi-independence would be satisfied by innate cognitive modules on the plausible assumption that modules are relatively decoupled from each other and from other traits. By 'innate', proponents of the dominant paradigm do *not* mean 'present at birth', but rather, encoded in the genome. For example, secondary sex characteristics are innate in that their development is encoded in the human genome

but they are not present at birth. The relevant notion of a cognitive module derives from Fodor (1983). But, whereas Fodor held that modules were largely peripheral mechanisms, the modules at issue here know no such boundaries. Nor are all of Fodor's characteristics always, or even typically, assumed. Rather, the key features are (1) domain specificity, both informationally and computationally, (2) universality—i.e., present in every normal mind in the species (assuming it has gone to fixation), and (3) relative encapsulation—insensitivity to collateral information[3].

When we say that this paradigm is dominant, we do not mean to suggest that it is widely accepted but rather that advocates and critics alike take the view in question to be the main theoretical contender defining the field. This is unfortunate because reliance on many innate cognitive modules has been heavily criticized in psychology and philosophy for its logical and theoretical incoherence, as well as its incompatibility with what is known about developmental neurobiology (which we discuss in some detail below). These criticisms have been cited as evidence that an evolutionary cognitive psychology is not viable.

But a viable evolutionary psychology does not require the existence of such modules in order to satisfy the heritability and 'quasi-independence' constraints. These requirements could also be satisfied by heritable learning biases, perhaps in the form of architectural or chronotopic constraints, that operated to increase the *canalization* of specific cognitive capacities in the ancestral environment (Cummins and Cummins, 1999). Chronotropic constraints are constraints on the time course of development (Elman et al, 1996). A trait is said to be more or less canalized as its development is more or less robust across environmental variations (See Ariew, 1996; Waddington, 1957; 1975). As an organism develops, cognitive capacities that are highly canalized

as the result of heritable learning biases might produce an organism that is behaviorally quite similar to an organism whose innate modules come 'on line' as the result of various environmental triggers.

Taking this possibility seriously is increasingly important as the case against the view that domain specific knowledge/processors are genetically encoded (a view sometimes referred to as 'representational nativism') becomes increasingly strong (Elman *et al.*, 1996; see also Samuels, 1998, for a critique; Karmiloff-Smith *et al.* 1998 for an answer). The implications of the outcome of this debate are relevant not just to proponents and opponents of evolutionary cognitive psychology; they are directly relevant to the broader question of how we characterize knowledge and its acquisition. While behavior geneticists and social scientists continue to argue about the relative importance of genes and environment in the development of personality and behavior, the initial either/or character of this debate has given way to an interactionist consensus. Nevertheless, there is still considerable vagueness and confusion about the conceptual framework needed to articulate specific interactionist proposals. This is especially true with respect to higher cognition: it is not clear how knowledge can be 'jointly authored', and hence the debate tends to slip back into a debate about which knowledge, beliefs, or behavioral traits are innate, and which are learned. Conceptualizing matters in terms of learning-bias-induced canalization of domain/task specific cognitive capacities, however, offers the promise of extending the interactionist perspective to cognition. One might ask of a jointly authored paper who wrote which sections, or paragraphs, or even sentences. But it could happen that both authors are jointly responsible for every sentence, with the *degree* of responsibility varying from place to

place. This suggests thinking of all of knowledge as co-authored. It suggests that the question is not which concepts (or rules or whatever) are contributed by the genes and which by learning, but how canalized the development of a given concept or body of knowledge happens to be (Cummins and Cummins, 1999). Many important pieces of an interactionist story about cognition are already to be found in the literature (see below). In this paper, we take some steps toward synthesizing this material, and articulating the conceptual resources for making this proposal specific.

To understand the learning-bias-and-canalization (LBC hereafter) framework gestured at above, we begin (section II) by summarizing the reasons for scepticism about the dominant massive innate modularity paradigm. Evolutionary psychology has been criticized both for requiring massive innate modularity (see below), and for being conceptually vague, confused, lacking explanatory force, and ill-founded empirically. A careful and rigorous formulation and analysis of LBC meets objections to massive modularity as well as questions about conceptual clarity and explanatory force. Only the science proper can speak to the issue of empirical grounding. Next, in section III we survey some of the recent developments that we believe are promising pieces of an LBC interactionist synthesis. In section IV, be briefly sketch the LBC framework. Finally, as an aid to rigor and theoretical viability, we also propose a series of simulations designed to demonstrate the feasibility of the processes postulated by the framework, and to test the rigor of its formulations. The goal of this kind of simulation work is primarily to investigate how selection could lead to learning biases that increase the canalization of fitness enhancing cognitive capacities. To some extent, this has already been done (see the

literature review below). Much is required, however, and we describe what we take to be the next step in section V.

II. Arguments against the direct genetic specification of knowledge.

To understand the motivation for the framework we propose for cognitive evolutionary psychology, one needs to appreciate the challenges facing the innate modules framework that currently predominates in evolutionary cognitive psychology. We begin, therefore, with a discussion of two influential lines of argument against direct genetic encoding of knowledge: the poverty of genetic resources, and the plasticity of cortical tissue. We do not necessarily endorse these arguments, but they are sufficiently persuasive to motivate a careful and complete formulation of a framework for evolutionary cognitive psychology that can accommodate them.

The poverty of genetic resources. Innate cognitive modules will certainly require innate knowledge of some sort, either in the form of explicitly represented knowledge or in the form of knowledge implicit in the logic and structure of special purpose processors. Either way, direct genetic specification of an innate module will, as far as we know, require specifying synaptic connections in the cortex. (It might require more than this—e.g., neuron differentiation—but it will surely require at least a specification of synaptic connections.) Even if one supposes with Fodor and Pylyshyn (1988) that the level of synaptic connections is not the appropriate level for the specification of cognitive function, cognitive capacities still must be implemented as patterns of synaptic connectivity. If cognitive capacities are specified in the genome, the genome must encode for specific patterns of synaptic connections.

The problem is that the human genome does not appear to have the resources to directly

specify a significant amount of cortical connectivity. It is now known that human genotypes

contain many fewer genes than previously thought (around 30,000-40,000 instead of a 100,000;

International Human Genome Sequencing Consortium, 2001). Among these, it is estimated that

from 20-30% (Wills, 1991) to perhaps as many as half (Thompson, 1993) may be implicated in

brain development. However, our brains literally contain trillions of synaptic connections and

5,000 to 15,000 genes are clearly insufficient to directly encode all of these (Churchland, 1995;

Buller and Hardcastle, 2000 – see also McCullogh 1951). Moreover, it seems that very few of the

genes involved in brain development are concerned with cortical development. Most of the genes

involved in brain development are dedicated to making sure our sensory transducers are properly

hooked-up. Winberg and Porter (1998) report that fully 4% of them are concerned with the

sensory cells located inside our nose! If there are innate *cognitive* modules, these surely will be

found in the cortex and not, e.g., in the 'lowly' pons.

The plasticity of cortical tissue. The data just summarized meshes nicely with neuroscientific

data suggesting that the human cortex is deeply immature at birth, leading some to claim that we

are not really born with a cortex but with a proto-cortex (O'Leary 1989, 1997). If the pattern of

connectivity in some area of the cortex were genetically pre-specified, then it should be

impossible to transplant it to another area without bizarre consequences. If, for example, cells in

the visual cortex are pre-programmed to finely connect themselves so as to subserve specifically

visual processing, then surely they could not function properly when transplanted to the

auditory cortex. But it has been shown that such transplants are possible, and that visual cortex

cells transplanted to the auditory cortex do connect themselves in a manner that is relevant to auditory processing (O'Leary et al., 1992; see Samuels 1998 for a critique).

III. Alternatives to nativism.

It is one thing to cast doubt on the feasibility of nativism; it is quite another to provide a feasible alternative that accounts for the data. We therefore briefly review some theory and research that, perhaps, has the potential to account for highly canalized, early emerging cognitive modules.

Quartz and Sejnowsky (1997; see also Quartz, 1999) argue that 'cortical development involves the progressive elaboration of neural circuits in which experience-dependent neural growth mechanisms act alongside intrinsic developmental processes to construct the representations underlying mature skills' (Quartz, 1999: 48). The growth of dendritic trees is determined both by general endogenous mechanisms and by environment-driven activity. Moreover, it has been shown recently that dendritic trees have the computational properties of neural networks with hidden layers (Quartz and Sejnowsky, 1997: 549). Indeed, some researchers have suggested that patches of dendrite membrane might be the basic computing units of the brain (Jaslove 1992, Koch et al., 1982). It would follow that the representational capacities of dendritic trees grow as a result of general endogenous mechanisms interacting with environment-driven activity. These results enhance the plausibility of so-called constructive learning algorithms developed for neural networks—learning algorithms that allow a network to be dynamically reconfigured during training through the addition or deletion of nodes and weights

(Azimi-Sadjadi et al., 1993; Fahlman & Lebiere, 1990; Frean, 1990; Hanson, 1990; Hirose et al., 1991; Wynne-Jones, 1993).

These ideas apply to the issue of nativism in a number of ways. For example, classic results from formal learning theory (e.g. Gold 1967) seem to show that language cannot be learned unless the child comes equipped with rich innate knowledge to restrict the set of grammars that have to be tested against evidence. These analyses assume that language learning is search through a fixed (time-invariant) hypothesis or representation space (a form of parametric regression). However, 'constructive learning', does not assume a fixed representation space. Indeed, its main feature is the *growth of the representation space*: 'The constructive learner builds its hypothesis space as it learns'. (Quartz and Sejnowsky, 1997: 553). Accordingly, the classic results from formal learning theory do not apply to constructive learning. This does not mean that constructive learners actually can solve learning problems that formal analysis appeared to rule out, but it does remove a formidable and influential barrier, and suggests a positive strategy for research. A number of formal and empirical results reviewed by Quartz and Sejnowsky provide reason for optimism.

Some of these results come from experiments with artificial neural networks. Since we cannot review all of the relevant literature here, we discuss what is perhaps the most influential example.

In their book, *Rethinking Innateness*, Elman and his colleagues (1996) distinguish three types of nativism (representational, architectural and chronotopic).[4] Representational nativism (RN) is the direct 'innate structuring of the mental/neural representations that underlie and constitute *knowledge*' (Bates *et al.* 1999: 591) Architectural nativism is the 'innate structuring of the information-processing system that must acquire and/or contain these representations' (Bates *et*

al. 1999: 592). These might include, for instance, a specification of the basic computing units (their types and internal properties), the local architecture (e.g., number of layers in a structure) and the global architecture (e.g. pathways between global regions). Finally, chronotopic nativism is the 'innate structuring of the timing of developmental events' (Bates *et al.* 1999: 592).

The role of chronotopic constraints is illustrated by a famous experiment of Elman's (1993). Elman showed that a simple recurrent network subject to no chronotopic constraints is unable to learn a relatively complex grammar (an artificial grammar with central embedding and long-distance dependencies). In this case, only the architecture of the network is innate (i.e., specified by the modeler). A representational nativist might point out that the negative result was predictable and suggest that innate knowledge, in the form of a pre-specified pattern of connectivity between the network's units, is necessary to reduce the search space that has to be explored by the learning algorithm. But Elman found out that he could get the network to learn the grammar without innate knowledge (pre-specified connection weights) by simply adding a *chronotopic constraint, viz., decreasing the noise faced by the network's context units as a function of time*. In a simple recurrent network, the context units function as a short-term memory (STM) so, in effect, high noise means a rapidly fading STM, and decreasing noise through time means augmenting the STM's fidelity. This suggests that a plausible developmental trajectory—more and more faithful STM—can enable otherwise impossible learning.[5] This developmental trajectory was studied independently, from a linguistics point of view, by Newport (1990, see also Goldowsky and Newport 1990).

IV. The LBC (learning-bias-and-canalization) framework

We believe the forgoing discussion opens up possibilities for evolutionary explanations and analyses of cognitive capacities that are considerably more liberal and complex than the reigning innate modules approach exemplified by the work of Cosmides and Tooby (1987, 1994a, b, 1995, 1997). While the elements of LBC are not novel, we believe the particular systhesis we propose, and the correlative analysis of fundamental concepts, is.

LBC does not require a genetically problematic representational nativism, yet it does appear to satisfy the two fundamental criteria set out above for the viability of evolutionary cognitive psychology: heritability and quasi-independence. We discuss these in turn.

Heritability. The LBC framework assumes that there was variability in learning biases in ancestral populations. A learning bias can not only make acquisition of a cognitive trait possible, it can make it more likely. That is, it can increase the canalization of a cognitive trait. When such a trait proves adaptive, there will be a tendency for selection to spread the bias(es) responsible for increased canalization through the population. The end result is a population with a highly canalized cognitive trait.

Learning biases might themselves be the result of a kind of limited RN, since encoding these would not be as costly as encoding specific cognitive traits. More attractive, however, is the idea that variation in such traits as cortical architecture and its chronotopic properties amount to learning biases. This makes it more difficult to conceive of the resulting biases as specific to particular learning tasks, e.g., language or cheater detection, but it makes their genetic specification and heritabilty much more plausible, given that evolving a set of representations

requires genetically coding each connection. While this solution might work well for small networks, its applicability decreases drastically as the number of connections increases.

While the heritability of such traits as architectural and chronotopic constraints in ancestral populations is eminently plausible, it is, in the present state of knowledge and technology, difficult—perhaps impossible—to demonstrate the historical heritability of particular architectural and chronotopic constraints bearing on the acquisition of specific cognitive capacities. This is primarily because heritability is zero where there is no variance, and we can expect such things as the developmental course of STM involved in language learning to have gone to fixation in the current population. In this instance, then, we must be satisfied by two lines of argument. The first involves appeal to the fact that a strong genetic component in the architecture of the cortex and its chronotopic properties seems undeniable. Second, we can construct simulations that illustrate how heritable constraints of this kind can, in fact, give rise to learning biases that canalize observed effects. Language, for example, does get learned, and, given the growing body of evidence against innate modules, there is a strong presumptive argument in favor of constraints and biases that can, in principle, account for the learning observed, including not just the end result, but its typical developmental profile and observed departures from it.

Quasi-Independence. If we assume that learning biases are not themselves representationally implemented, then, as remarked above, they will not be intrinsically specific to particular learning tasks, and this will impact the degree to which the cognitive traits they enable can evolve independently. This is both good news and bad news. The good news is that a demonstrable coupling between cognitive traits would provide strong evidence for a bias that grounded both,

since the existence of such a bias would immediately explain the coupling. The hypothesis thus has the capacity to generate some interesting and falsifiable predictions. The bad news is that, as we noted in the introduction, some degree of independence among cognitive traits seems required to make evolutionary analysis a useful and interesting tool in cognitive psychology. It is worth noting that, even if learning biases proved to be quite general, it would not follow that they could not give rise to a modular functional architecture. A single bias could canalize development of several domain specific capacities. For instance, Jacobs *et al.* (1991) have shown 'how mixtures of expert networks exposed to a what/where problem […] will always assign the 'where' task to the expert network which possesses a linear activation function. The implication is that networks do not necessarily need to be designed to carry out particular tasks. Rather, the task will select the network which has the appropriate (i.e. innate) computational properties'. (McCleod *et al.* 1998)

Two considerations tend to modify the point just made. First, architectural modularity might allow for learning biases that are not representationally based to have consequences for only one or a few cognitive traits. If, for example, there is a STM specific to language learning, or some aspect of visual processing, then the chronotopic constraints on it might evolve primarily or exclusively under the control of the adaptive consequences they have for language learning or that aspect of visual processing. A second and closely related consideration is that specific types of input are known to be capable of causing the specialization of the local neural circuitry that receives it early in development (Elman, et. al., 1996). Hence, an innate specification of input connections can serve to specialize circuitry while effectively decoupling it from circuitry grounding other types of processing.

LBC has a number of special features that require emphasis.

1. Learning has its costs. Learning takes time, energy and cognitive resources. During the learning period, an organism is at a disadvantage vis-à-vis a counterpart that does not have to learn the target capacity. Learning typically entails making mistakes, and these can be costly. If learning is done in the infant/juvenile period, care must be provided by parents, etc.

2. Learning has its advantages. Learning allows for the possibility that an organism can adapt to new environments. Hence, learning can sometimes give rise to a capacity that is more highly canalized than one that is completely innate in conditions in which the novel environment fails to 'trigger' development of the innate capacity, or leads to the development of a 'capacity' that isn't capable of fulfilling its function in the novel environment.

3. Learning allows for the possibility of overcoming neural defects due to trauma or developmental problems. Direct specification of synaptic connections and structure in the genome would seem to leave the organism vulnerable in this respect.

4. LBC would appear to allow more easily for gradual evolution of cognition, since the relevant architectural and chronotopic constraints can be introduced and gradually modified antecedent to the appearance of the target cognitive capacity.

5. LBC is compatible with innate modules (though the arguments that motivate us to take it seriously are not), as well as with the innate developmental recipe approach associated with Pinker (Pinker, 1994, 1997) Direct comparison of LBC with other frameworks at both the conceptual level and the simulation level is a major goal of the project.

V. The Role of Simulation in this research

The simulation of adaptive behavior (SAB) is now a recognized sub-field of psychology, with its own periodicals (e.g., *Adaptive Behavior*, *From animals to animats*, both from MIT Press) and series of conferences (*International Conference on Adaptive Behavior*). And as Miller and Todd (1994) rightly point out, there is a clear link between SAB and work in evolutionary psychology, a link that goes both ways: human evolutionary psychology can inform SAB by providing target behaviors to simulate and SAB can inform evolutionary psychology by testing the computational realizability of putative evolutionary stories (i.e., whether a proposed story could have taken place by the application of known evolutionary mechanisms). This aspect of SAB is particularly important in the present context since, as all researchers who have attempted the simulation of complex nonlinear processes know only too well, the dynamics of nonlinear processes is often counter-intuitive and difficult to rationally anticipate from the armchair.

Simulation plays several different related roles in our work.

- *Rigor and precision*: It has repeatedly been our experience that simulation exposes lacunae, vagueness and ambiguity in the formulation of theories and hypotheses at every level of detail.

- *Hypothesis generation*: simulation not only helps to refine and elaborate hypotheses, it often suggests novel hypotheses. The example of chronotopic constraints from Elman rehearsed above is a striking case in point.

- *Prediction generation*: the interaction of architectural and chronotopic constraints on network design and development with learning and selection is extremely complex. We

suspect the only sure way to reliably generate predictions from hypotheses about such interactions is to run simulations.

- The simulation of adaptive behavior provides us with the means to test the sufficiency, of the LBC framework, and to compare it in detail with competitors.

In the literature on simulations, two distinct algorithms are alternatively at work. (1) A neural network 'learns' various environmental contingencies during the life of the organism whose behavior it directs, and develops a set of representations through its interaction with that environment. (2) A genetic algorithm 'evolves' a population of networks through differential reproduction and mutation[6] of some of the network's parameters. In what follows, we briefly review some the relevant literature in this field and explain how our work proposes to adapt this research to our study of cognitive modularity.

Three degrees of RN have emerged in this literature: *full* representational nativism, *minimal* representational nativism (after Clark 1993[7]), and no representational nativism (or *representational* constructivism).

Full RN is the view that systems are born with a full set of representations or develop such a set without the help of learning mechanisms. Work in the simulation of adaptive behavior has shown that it is possible to evolve representations in connectionist nets as 'easily' as it is to teach them: any set of representations (underlying such behaviors) that can be learned by a neural network can also be evolved by a genetic algorithm (see Yao 1993 for a review). In an evolved population of neural networks, all or most individuals in the population are born with the representations, which are therefore literally innate. For instance, Nolfi, Elman and Parisi (1994)

evolved a population of animats (simulated organisms), controlled by a neural network and

selected for their ability to find food items on the basis of information about distance and

direction to food. Neural networks can easily be trained by backpropagation to achieve such

behaviors and, as Nolfi, Elman and Parisi have shown, they can also be evolved to do so. An

experiment by Floreano and Mondada (1996) shows how sophisticated these evolved

representations can be. They used a simple recurrent network (Elman 1990) to let some of the

robot's representational capacities be carried by the network's internal dynamics. Although the

network only contained four hidden units, they found that two of them sustained an intricate

topological map of the robot's environment, a map that varied systematically as a function of the

robot's current orientation and the state of its rechargeable battery. These hidden units, in fact,

seemed to function as 'place cells', which are known to exist in the rat's hippocampus (O'Keefe

and Nadel, 1978), indicating, perhaps, that this type of internal organization is an especially

efficient way to implement navigation behavior. Full representational nativism is thus a viable

option from a purely computational point of view. However, to get a genetic algorithm to evolve

a set of representations, each connection weight must be coded on an (artificial) chromosome.

While this solution might work well for small networks, its applicability decreases as the number

of connections increases. Moreover, maintaining a large genotype is not without costs of its own

(both in computers and in nature), and it is reasonable to think that, if the same end state can be

reached with a smaller genotype, the very process that evolves the set of representations will

also select those networks that can achieve the result with as small a genotype as possible.

Minimal RN is the view that systems are born with (or develop without the help of learning

mechanisms) the minimal set of representations necessary to learn the full set underlying a given

behavior. Nolfi and Parisi (1997) constructed a network comprised of two modules, one (the

standard module) whose function was to drive the wheels of a small robot and one (the teaching

module) whose function was to send a teaching signal to the other module.[8] The connection

weights of both modules were evolved through a genetic algorithm, but the weights of the

standard module were also adjusted throughout the life of the robot by backpropagation (on the

basis of a teaching signal it received from the teaching module). The weight changes caused by

backpropagation were discarded at reproduction time, in order to insure Darwinian (not

Lamarckian) evolution. Finally, successive generations of robots were alternatively 'raised' in

bright and dark environments (a difference that has a major impact on their infrared sensors).

Robots were selected as a function of their ability to navigate the environment quickly while

avoiding obstacles. As a control, Nolfi and Parisi also evolved a population of simpler robots

only comprised of a standard module. The robots in the control population are born with the

innate representations needed to behave successfully in both environments. However the fitness

of the control population was quickly surpassed by that of the two-module network. The control

networks evolve a strategy that works in both environments but that is optimal in neither. The

two-module robots adopt different and superior strategies depending on whether they are in a

bright or dark environment. Evolution has given the teaching module the representations

necessary to differentially teach the standard module according to the environment in which the

robot is learning. It is interesting to note that neither module evolved a set of representations

efficient for navigating while avoiding obstacles; their success remained dependent on learning.

Unlike the control population, the two-module robots are not born knowing how to behave. They are born with a set of representations that predispose them to learn the set of representations underlying the best strategy for the environment they happen to occupy.

Representational constructivism is the view that all representations are learned from the interaction between the network and its environment. Too often, however, representational constructivism is confused with the view that everything cognitive is learned. Elman *et al.*'s (1996) distinction between representational, architectural and chronotopic constraints on development allows us to rethink the nature/nurture debate within cognitive science. Perhaps evolution only fixes the architectural and chronotopic constraints necessary to canalize learning towards the set of representations underlying the modules' cognitive abilities.[9] Indeed, Elman's famous (1993) experiment (described above) is one such instance, where only architectural and chronotopic constraints were innate. However, these were not 'evolved' but rather fixed by trial and error. It would be interesting to see which constraints a process of simulated evolution like the one described above would fix in the network's genotype. In any case, some researchers have evolved only chronotopic or architectural characteristics of learning networks. For instance, Floreano and Mondada (1996) devised an experiment in which a robot that isn't born with a navigation module learns it very quickly (in less than 10 sensory-motor loops – i.e., 10 sweeps of the back propagation algorithm) *because of what it is born with*. Instead of evolving the connection weights to and from the hidden unit, Floreano and Mondada evolve a set of 'meta-properties' for each connection: whether the connection is driving or modulatory, whether it is excitatory or inhibitory, what learning rule it uses (one of four variants of the Hebb rule) and

what its bias is. Because the meta-properties are evolved in the environment in which the robot's

network controller will eventually have to learn the task, and because the connection weights are

not coded on the chromosome, the genetic algorithm doesn't evolve a capacity to navigate but an

ability to quickly learn navigation. The capacity is not innate but highly canalized through a

motor learning bias:

> Learning of the evolved controller relies on simple genetically inherited abilities. For example,
>
> the controller analyzed above always starts by moving backward until it finds some object;
>
> then it rotates to the right until its rightmost sensors become active and synapses begin to
>
> change. These two simple motor programs result from weak sensory signals (mostly noise)
>
> filtered and amplified through the synaptic matrix of excitatory and inhibitory weights. They
>
> represent the basis from which learning can start to operate and are similar to instincts in that
>
> they are both genetically inherited and represent primitives upon which the organism can start
>
> to explore its environment. In other words evolution not only shapes the learning modality,
>
> but also bootstraps learning by selecting basic behaviors useful for learning (Nolfi and
>
> Floreano 2000:172).

To test the in-principle sufficiency of the LBC framework, and to investigate its implications,

simulation experiments need to be designed to explore the way in which the development of a

given adult state (in particular, the representations that are though to comprise a cognitive

module) can be canalized by the presence of innate learning biases. In such simulations, learning

and evolution interact as previously, but evolution (as simulated by the genetic algorithm) is

constrained not to evolve a set of representations (or a *full* set of representations) but a set of

architectural and/or chronotopic (or some *minimal* representational) learning biases which, in any standard environment, will lead to the acquisition (via backpropagation or some other connectionist learning algorithm) of the necessary representations. These simulations should be designed to allow comparisons between innate-module scenarios (Full-RN) with LBC scenarios (Minimal or no-RN). Note that an LBC scenario involves architectural and chronotopic nativism, but no representational nativism. We describe two such simulations here to give the flavor of the work that needs to be done.

1. Simple Food Finding Task

In evolutionary biology, fitness is defined as reproductive success -- that is, the number of living offspring an individual produces that go on to reproduce themselves. Inclusive fitness means an individual's own fitness plus his or her effect on the fitness of any relative. Maximizing fitness means maximizing reproductive success. In engineering applications, fitness is typically defined *prior to reproduction* in terms of a 'fitness function', which evaluates the performance of each individual phenotype with respect to some predetermined criterion. Those individuals that achieve the highest fitness values are allowed to reproduce. In the simulations described below, we choose food consumption as the criterion of interest, and define our fitness function in terms of the number of food items eaten during a specified period of time. A natural strategy is to use a sigmoid function (which is common in neural network models) wherein fitness values change very little until a certain threshold of intake is achieved, then fitness increases linearly with food intake until an upper threshold is achieved at which point it tapers off. Those achieving the highest fitness values will be allowed to reproduce.

A typical simulation would consist of a main program and four subroutines embedded in each other: a neural net embedded in a critter embedded in an environment embedded in a genetic algorithm. The main program attaches a net to a critter, puts the critter in an environment where it gets to move around and eat, then applies the genetic algorithm. Computations proceed as follows: There is an input (a direction to move), which causes activation to spread through a simple recurrent net (4X4X4 in the current pilot). The output is a move. The critters (n=20) make 27 moves and then fitness (amount of food consumed) is measured. (For details about learning, see below.) The one top individual is allowed to reproduce. For the others, a roulette scheme (commonly used with genetic algorithms) is used: the roulette represents the total sum of fitness in the population, and each critter gets a portion of the roulette proportional to its own fitness. The roulette is spun 20 times and whichever critter 'wins' gets to reproduce. Mutation occurs randomly, with a few individuals being mutated a great deal, while others are not mutated at all. While biologically unrealistic, this is more realistic than very slight mutations to every synapse of every individual (common in the connectionist literature) and has a substantial engineering track record. Other mutation plans need to be tested as well. Unfortunately, crossover is not currently feasible with neural nets

The environment consists of a surface overlaid by a grid. The critters can move forward, backward, left, and right on the grid. The environment is contoured in various ways so that moves can be classified as up or down various gradients. This might be, e.g., smell (stronger, weaker, same), altitude (up, down, same), light (brighter, darker, same), etc. Food can be found

(with greater or less statistical likelihood) by following certain gradients up or down, and ignoring others.

Learning. All critters start out with a set of random weights. Full RN critters don't learn, only LBC and mixed condition critters learn LBC and mixed critters are trained using back-propagation to predict the results of possible moves—up, down, same—across the gradients in the environment. The prediction task is useful because the correct response is simply (some part of) the next input, making back propagation learning more realistic. They are not trained to make moves that take them towards food, since, a priori, there is no reason to prefer one gradient over another, or going 'up' to going 'down'. Rather, they are reinforced for moves that generate/increase some 'pleasant' stimulus. Evolution determines which stimuli can be discriminated and which, among those, are 'pleasant'—i.e., which generate reinforcement signals. This allows the critters to learn to follow a gradient even though the payoff is remote.[10]

Selection: What is reproduced. In the pure LBC condition, connection weights are not reproduced. Instead, architectural constraints (e.g., number of nodes, number of layers, connection patterns, connection meta-properties, and activation functions), chronotopic constraints (e.g., time course of realized changes in the network over the lifetime of the individual), and learning rule parameters (e.g., learning rates and intervals) are reproduced. In the mixed condition, some connection weights are also reproduced, viz., those in some subnetwork. In the full RN condition, the connection weights are reproduced *in addition to* what is reproduced in the pure LBC and mixed conditions.

24

2. Social Encounters.

Critters travel randomly through an environment, collecting food. When one critter encounters another, it can share food, withdraw, or attempt to take the other critter's food by force. Both critters have the same options. Outcomes are determined by relative dominance rankings. Dominance ranking is encoded as a property list, which includes size, health, facial expression, posture, and observed encounters with other critters. When one critter encounters another, their respective property lists are encoded in the input activation vectors to each other. A property list is used rather than a simple relative ranking because it allows relative dominance to be learned from the list. Advanced variations include the possibility of being aided by 'friends' recruited by sharing resources.

Selection works analogously to the first task.

Learning. As in the first task, networks are trained to predict the outcomes of possible moves, given dominance information, or information bearing on dominance. In this instance, we can assume immediate feedback concerning the move actually made, since the outcome of an encounter is determined immediately.

Simulations like these are designed to test whether the LBC condition can be at least as effective as the RN condition in generating a population with highly canalized cognitive capacities. To do this, we compare critters whose target capacities evolve in the RN condition, with critters that evolve learning biases that enable effective acquisition of the target capacities,

either by evolving representations (connections) in a learning subnetwork (the mixed condition), or by evolving non-representational constraints that fast-track learning of the target capacity.

3. Evaluation

Populations are compared in terms of average fitness, where the fitness of an individual is, as before, assumed to be proportional to task success during a lifetime. Individuals and populations can also be tested for plasticity, since one of the manifest advantages of learning is that it has the potential to allow the organism to acquire the target capacity in altered environments, thus opening up the possibility that the LBC scenario might lead to organisms with more highly canalized capacities than RN. This is an important consideration, since it tends to undermine universality as an argument for representational innateness. A universally acquired capacity is likely to be one acquired in a variety of environments. Too much environmental variability has the potential to lead to the extinction of capacities dependent on representations specified in the genome.

Simulations like those just described would hopefully advance our understanding a step beyond the results reviewed above by making direct fitness and canalization comparisons between RN, mixed and non-RN scenarios, and by introducing social encounters (generally studied only in RN conditions) in a way directly relevant to evolutionary hypotheses concerning innate social reasoning modules.

VI. Conclusion

We believe that the emergence of evolutionary cognitive psychology is an important development in cognitive science, one that may leave a lasting effect on the whole field. This

promise can only be realized, however, if the research is conceptualized in a framework that is consistent with what is known about the complex interactions between evolution, neural architecture and its development, and learning. The LBC framework described here shows considerable promise of effecting the needed 'interactionist' synthesis in connection with the evolution of cognition. It avoids massive innate cognitive modules and the problematic representational nativism it engenders. In its place, it motivates us to look for the evolution non-representationally based learning biases that interact with development and the enviornment to produce hignly canalized and early emerging cognitive capacities.

Evolutionary cognitive psychology is intrinsically problematic because brains are soft, leaving no historical record, and because the traits of interest have (presumbaly) gone to fixation and cannot be readily studied in fruit flies or bacteria. It can, however, be studied in artificial creatures. Ethology, especially primatolory, can provide invaluable clues. By telling us which capacities develop when and under what environmental conditions, developmental psychology continues to provide explananda that can, in principle and in fact, be used to distinguish among evolutionary hypotheses when these are framed with sufficient detail and rigor. (see D. Cummins, xxxx for some telling examples.) This will not be enough for those who want the scientific equivalent of a smoking gun. To those, we offer the following reminder: no one has ever predicted a tide from Newton's theory of gravity (or any other theory of gravity) and the hypothesis that lunar gravitation causes the tides. Yet this is a text-book example of good science because it allows us to understand the tides in terms of simulations ranging in complexity form passing a magnet over a newspaper with another magnet under it, creating a wave, to computer

simulations e.g., http://webphysics.ph.msstate.edu/jc/library/9-PP. (We are grateful to Franz-Peter Griesmier for this example.) No one dismisses this explanation as a 'just so' story because it cannot be directly tested. Interactions between evolution, development and learning of the sort contemplated within the LBC framework suggested here have the potential to allow us to understand cognition in a ways that standard cognitive and developmental psychology and neuroscience cannot hope to achieve alone.

Footnotes

[1] The term is Lewontin's (1978). 'Quasi-independence means that there is a great variety of alternative paths by which a given characteristic may change, so that some of them will allow selection to act on the characteristic without altering other characteristics of the organism in a countervailing fashion; pleiotropic and allometric relations must be changeable.' The point is that, from a genetic point of view, an intermediate grade of trait interaction/independence is required: Too much interaction (universal pleiotropy) means that there is basically one evolutionary unit that must evolve as a unit; too little interaction (complete additivity) blocks the developmentally concerted integration of complex phenotypes because the component genes/effects are always being divided up in reproduction. In between is the realm of quasi-independence.

[2] Cosmides and Tooby are not the only proponents of this view. See also, e.g., Carey & Spelke (1994) and Leslie, 1992.

[3] This characterization differs somewhat from the 'Darwinian module', ascribed to evolutionary psychology by Samuels, Stich and Tremoulet (1999). They define a Darwinian module as 'an innate, naturally selected, functionally specific and universal computational mechanism which may have access (perhaps even unique access) to a domain specific system of knowledge of the sort we've been calling a Chomskian module.' Encapsulation is not mentioned in this quote, but we retain this characteristic from Fodor's original formulation because, without it, it is difficult to distinguish a module from a mere 'subroutine'.

[4] The original distinction between representational and architectural nativism is from Narayanan (1992).

[5] An added bonus is a possible explanation of the so-called critical learning period for language, since it is the plausibly irreversible change in STM that matters, not its ultimate fidelity.

[6] Mutation is only one of the genetic parameters that can be used to implement such a search procedure. Another well known parameter is crossover, which is known to work well computationally (Holland 1975) but is difficult to apply to neural networks (Yao 1993).

[7] Although we adopt Clark's expression, what we describe as minimal representational nativism is not exactly what he describes with the term since, contrary to us, he refuses to distinguish architectural and representational constraints on development.

[8] For a similar system, see Ackley and Littman (1991).

[9] Or perhaps evolution fixes those first, and as much as possible, resorting to fixing representations (connection weights) only when the space of possible architectural and chronotopic biases has been fully explored and no set of biases has been found that could canalize learning towards the proper set of representations; a solution that would correspond to minimal representational nativism (see above).

[10] Learning has costs, e.g., delayed reproduction, and the costs of prenatal and juvenile investment by parents.

References Cited

Ackley D.H. and Littman M.L. (1991). Interaction between learning and evolution. In C.G. Langton et. al (Eds.), *Proceedings of the Second Conference on Artificial Life.* Reading, MA: Addison-Wesley.

Ariew, A. (1996). Innateness and Canalization. *Philosophy of Science*, 63 (Proceedings), S19-S27.

Ariew, A. (1999). Innateness is canalisation: In defense of a developmental account of innateness. In Valerie Hardcastle (ed.) *Biology Meets Psychology: Constraints, Connections, Conjectures*. Cambridge, MA: MIT Press, p. 117-138.

Azimi-Sadjadi, M.R., S. Sheedvash and F.O. Trujillo (1993). Recursive Dynamic Node Creation in Multilayer Neural Networks. *IEEE Transactions on Neural Networks*, 4, 242-256.

Bates, E, J.L. Elman, M.H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett (1999). Innateness and Emergentism. In *A Companion to Cognitive Science*. W. Bechtel and G. Graham (eds). Oxford: Blackwell.

Buller, D.J. and V.G. Hardcastle (2000). Evolutionary Psychology, Meet Developmental Neurobiology: Against Promiscuous Modularity. *Brain and Mind*, 1, 307-325.

Carey, S. and E. Spelke (1994). Domain-specific Knowledge and Conceptual Change. In *Mapping the Mind*. S. Gelman and L. Hirshfield (ed.). Cambridge: Cambridge University Press, pp. 169-200.

Churchland, P.M. (1995). *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge (MA): MIT Press.

Clark, A. (1993). Minimal Rationalism. *Mind*, 102, 587-610.

Clark, A. (1998). What's Knowledge Anyway? *Mind and Language*, 13, 571-575.

Cosmides, L. and J. Tooby. (1987). "From Evolution to Behavior: Evolutionary Psychology as the Missing Link" In *The Latest on the Best*. J. Dupré (ed.) Cambridge: MIT Press, pp. 278-306.

Cosmides, L. and J. Tooby. (1994a). "Beyond Intuition and Instinct Blindness: Toward an Evolutionary Rigourous Cognitive Science". *Cognition*, 50, 41-77.

Cosmides, L. and J. Tooby. (1994b). "Origins of Domain Specificity: The Evolution of Functional Organization". In *Mapping the Mind*. S. Gelman and L. Hirshfield (ed.). Cambridge: Cambridge University Press, pp. 85-116.

Cosmides, L. and J. Tooby. (1995). "From Function to Structure: The Role of Evolutionary Biology and Computational Theories in Cognitive Neuroscience." In *The Cognitive Neurosciences*. M. Gazzaniga (ed.). Cambridge (MA): MIT Press.

Cosmides, L. and J. Tooby. (1997). "The Modular Nature of Human Intelligence". In *The Origin and Evolution of Intelligence*. A. B. Scheibel and J. W. Schoff (eds.). Boston: Jones and Bartlett Publishers, pp. 71-101.

Cummins, D.D. and R. Cummins (1999). Biological preparedness and evolutionary explanation. *Cognition*, 73, B37-53.

Elman, J.L. (1990). Finding Structure in Time. *Cognitive Science*, 14, 179-211.

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

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

Fahlman, S.E. and C. Lebiere (1990). The Cascade-Correlation Architecture. In Advances in Neural Information Processing Systems, D. Touretsky (ed.). Morgan Kaufmann.

Floreano D. and Mondada F. (1996). Evolution of plastic neurocontrollers for situated agents. In P. Maes, M. Mataric, J-A. Meyer, J. Pollack and S. Wilson. (Eds.), *From Animals to Animats 4, Proceedings of the International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Fodor, J.A. (1983). *The Modularity of Mind: an essay in faculty psychology*. Cambridge (MA): MIT Press.

Fodor, J.A. and Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture. *Cognition*, 28, 3-71.

Frean, M. (1990). The Upstart Algorithm: A Method for Constructing and Training Feedforward Neural Networks. *Neural Computation*, 2, 198-209.

Gold, E.M. (1967). Language Identification in the Limit. *Information and Control*, 10, 447-474.

Goldowsky, B. and E. L. Newport (1990). The Less Is More Hypothesis: Modeling the effect of processing constraints on language learnability. Unpublished Manuscript. University of Rochester, Rochester, N.Y.

The Journal of Philosophy, vol XCI, no. 6, p. 277-304.

Hanson, S. J. (1990). Meiosis networks. In D. S. Touretsky (Ed.), Advances in Neural Information Processing Systems II (p. 533), San Mateo: Morgan Kaufman.

Hirose, Y., K. Yamashita and S. Hijiya (1991). Backpropagation algorithms which varies the number of hidden units. *Neural Networks*, 4, 61-66.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor (MI): University of Michigan Press.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.

Jacobs, R.A. M.I. Jordan, A.G. Barto (1991) *Task Decomposition Through Competition in a modular Connectionist Architecture: The what and where vision tasks*. Cognitive Science, 15, 219-250.

Jaslove, S.W. (1992). The Integrative Properties of Spiny Distal Dendrites. *Neuroscience*, 47, 495-519.

Johnson, M.H. (1997). Developmental Cognitive Neuroscience: an introduction. Oxford: Blackwell.

Karmiloff-Smith, A.; K. Plunkett; M.H. Johnson, J.L. Elman; and E. Bates (1998). What Does It Mean to Claim that Something Is 'Innate'? Response to Clark, Harris, Lightfoot and Samuels. *Mind and Language*, 13, 588-597.

Koch, C. T. Poggio and V. Torre (1982). Retinal Ganglion Cells: A Functional Reinterpretation of Dendritic Morphology. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 298, 227-263.

Leslie, A. (1992). Pretense, Autism, and the 'Theory of Mind' Module. *Current Directions in Psychological Science*, 1, 18-21.

Lewontin, R.C. (1978) Adaptation. *Scientific American.* Sept. (239), 157-169.

McCleod, P., K. Plunkett and E.T. Rolls (1998). *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford: Oxford University Press.

McCullough, W. S. (1951). "Why the mind is in the head' In Lloyd A. Jeffres, *Cerebral Mechanism in Behavior. The Hixon Symposium.* New York: Hafner.

Miller, G.F. and P.M. Todd (1994). A bottom-up approach with a clear view of the top: how human evolutionary psychology can inform adaptive behavior research. *Adaptive Behavior*, 3, 83-95.

Narayanan, A. (1992). Is Connectionism Compatible with Rationalism? *Connection Science*, 4, 271-292.

Newport, E.L. (1990). Maturational Constraints on language learning. *Cognitive Science*, 14, 11-28.

Nolfi S., Elman J.L. and Parisi D. (1994). Learning and evolution in neural networks. *Adaptive Behavior*, 3, 5-28.

Nolfi, S. and D. Floreano (2000). *Evolutionary Robotics*. Cambridge (MA): MIT Press.

Nolfi S. and Parisi D. (1997). Learning to adapt to changing environments in evolving neural networks. *Adaptive Behavior,* 5: 99-105.

O'Keefe, J., and L. Nadel, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.

O'Leary, D.D.M. (1989). Do Cortical Areas Emerge from a Protocortex? *Trends in Neuroscience*, 12, 400-406.

O'Leary, D.D.M. (1997). Areal Specializationof the Developping Cortex: Differentiation, Developmental Plasticity and Genetic Specification. In *The Lifespan Development of Individuals: Behavioral, Neurobiological and Psychosocial Perspectives: A Synthesis*. Magnusson et al. (eds). Cambridge : Cambridge University Press, pp. 23-37.

O'Leary, D.D.M., B.L. Schlaggar, and B.B. Stanfield (1992). The Specification of Sensory Cortex: The Lessons from Cortical Transplantation. *Experimental Neurology*, 115, 121-126.

Pinker, S. (1994). *The Language Instinct*. New York: Harper-Collins.

Pinker, S. (1997). *How the Mind Works?* New York: Norton.

Quartz, S.R. (1999). The Constructivist Brain. *Trends in Cognitive Sciences*, 3, 48-57.

Quartz, S.R. and T.J. Sejnowsky (1997). The Neural Basis of Cognitive Development: A Constructivist Manifesto. *Behavioral and Brain Science*, 20, 537-596.

Samuels, R. (1998). What Brains Won't Tell Us About the Mind : A Critique of the Neurobiological Argument Against Representational Nativism. *Mind and Language*, 13, 548-570.

Samuels, R., S. Stich and P. Tremoulet. (1999). Cognitive Science and Human Rationality: From Bleak Implications to Darwinian Algorithms. In Lepore, E., & Pylyshyn, Z. (eds.), *What is Cognitive Science?* Malden, MA: Blackwell, 74-120.

Thompson, R. F. (1993) *The Brain: A Neuroscience Primer*. New York: Freeman.

Tooby, J. and L. Cosmides (1995). Forward to Simon Baren-Cohen, *Mind Blindness: an essay on autism and theory of mind.* Cambridge, MA: MIT Press (pp. Xi-xviii.)

Waddington, C.H. (1957). *The Strategy of the Genes*. London: Allen and Unwin.

Waddington, C.H. (1975). *The Evolution of an Evolutionist*. Ithaca: Cornell University Press.

Wills, C. (1991). *Exons, Introns, and Talking Genes: The Science Behind The Human Genome Project*. New York: Basic Books

Winberg, J. and R.H. Porter (1998). Olfaction and Human Neonatal Behavior: Clinical implications. *Acta Paediatria*, 87, 6-10.

Wynne-Jones, M. (1993). Node Splitting : A Constructive Algorithm for Feed-forward Neural Networks. *Neural Computing and Applications*, 1, 17-22.

Yao, X. (1993). A Review of Evolutionary Artificial Neural Networks. *International Journal of Intelligent Systems*, 8, 539-567.