Connexions

An online journal of cognitive science

ISSN 1368-3233

EDITORS: TOM DICKINS, KEITH FRANKISH, AND ADAM HUFFMAN

Issue 4 October 1998

Contents

Sponsored by the Hang Seng Centre for Cognitive Studies

The Enemy from Within : of Memes and Modules, Explanation and Confabulation

Lars Hall
Dept. of Cognitive Science
University of Lund

1. Introduction

After a self-proclaimed charitable reading of *Consciousness Explained*, Block (1994) concludes that Dennett has not produced a theory of *consciousness* at all, but rather a theory of the *whole mind* (and a false one, at that). Dennett of course takes issue with Block. If we want no tack with phenomenal properties, Cartesian theaters, *me*dia, qualia, raw feelings, real seemings, and their kin, then the question of consciousness naturally turns into a question of the whole mind. I agree with all this. Nevertheless I take myself to represent a hitherto silent class of students of cognitive science that tend to agree with Block's conclusion, but for a wholly different reason. That is, given that one already is a die-hard non-reductive materialist, and that one is aware of the largely uncommitted and incomplete nature of the multiple-drafts theory, the really *exciting* thing about Dennett is that in his writings a vision of the whole mind can be discerned[1].

Now, I do not think that Dennett would agree that he has ever tried to provide something like a rough, but complete picture of the mind. Even so, I gather his sizable writings over the years add up to such a grand project, and I think it would be interesting to momentarily set consciousness aside and evaluate the wider merit of some of the claims made by Dennett. As alluded to above I am in full agreement with Dennett on the issue of consciousness, and I also follow him in rejecting Intentional Realism and The Language of Thought, but my hunch is that *within* the bounds of his theorizing there is enough room for amendments that would make the overall picture look rather different in the end.

What I would like to argue here is that some of the very exciting and intriguing conclusions that Dennett has presented in his series of books on human intentionality and consciousness (1987, 1991a, 1995, 1996) are very much dependent on a detailed specification of the *interface* between our evolved cognitive architecture and cultural/memetic evolution, and that when we unravel the tangled causal web of human development and social interaction we might very well get quite different results out of Dennett's basic premises. To make a case for this I first identify what I take to be the most crucial components of Dennett's vision, and then go on to a discussion of the interaction between *memes and modules, explanation and confabulation*

2. Dennett's Mind

Obviously I can not do justice to thirty years of proliferate academic work in one short summary. What follows is nevertheless, in rough phylogenetic order, what I take to be the most important components of Dennett's theorizing outside the main area of consciousness:

- An appreciation of the body as a slow but reliable system of information exchange - a mind of its own - that continues to exert its influence over our cognitive processes, not primarily by being our bounded, stable vehicle of interaction in the world (with all its talents and limits), but by supplying much needed *valence* to our decision procedures (Damasio 1994 calls these prevalent emotional tones *somatic markers*). According to Dennett this (mainly hormonal) wisdom of the body, tried and tested through the eons, does not have to be (re)represented in some more "explicit" format in the brain to be usable: "By using the old bodily systems as a kind of sounding board, or reactive audience, or critic, the central nervous system can be guided - sometimes nudged, sometimes slammed - into wise policies" (1996, p 79).

- The evolutionary precedence of modular, cheap, environmentally coupled cognitive systems with a minimal amount of representation. Animal minds can often be described as "a collection of interlocking, special purpose minimalist subroutines" (1995, p 373, in this case a bird).[2] In accordance with the need-to-know-principle of espionage (give each agent as little information as

possible to accomplish its share of the mission), and for myopic reasons of economy and efficiency, evolution builds brains that are essentially "bags of tricks." Such brains might solve problems in ways totally different from what we would intuitively expect, as seen for example in the hunting behavior of snakes where three different sensory systems without any central organization are used: vision for striking, smell for locating the prey again (that normally runs away for some distance before dying), and finally tactile information for finding the head and to commence eating. Each behavior is exclusively wired to a particular modality with no apparent interaction between the different systems except via the behavioral loop out into the world (Gärdenfors 1996, see also discussion in Dennett 1994). In this vein Dennett (1996, 1997a) warns us of ascribing too much distinct, *propositional thoughts* to this type of cognitive architecture. Describing the well-known early experiments of Wolfgang Köhler in which chimpanzees had to arrange boxes and sticks to get to a banana that hung from the roof, he writes: "Now, were they *thinking* when they were fussing about in their cages? What were they manipulating? Boxes and sticks. It is all too tempting to suppose that their external, visible manipulations were accompanied by, and driven by internal, covert manipulations - of thoughts about or representations of these objects, but succumbing to this temptation is losing the main chance. What they were attending to, manipulating and turning over and rearranging were boxes and sticks, not thoughts" (1997a).

- A contention that we do not now any "plausible upper bound on the number of concurrent, semi-independent control structures that can coexist in a nervous system (1996 p 125)", and an admonition that we should be *very* careful in taking intentional stance characterizations of animal behavior at face value (see also above). So far it has not been proved that reputed examples of higher-order intentionality among animals (for example the distraction displays of the piping plover[3]) are produced by anything other than "dumb", attentional, modular mechanisms that are sensitive to contingencies of behavior, coupled with some traditional reinforcement learning (so called *killjoy* mechanisms, Dennett 1988). A representational gloss on these systems is therefore often misleading - information about the minds of other agents is embodied in their actions, but it is information *in* the system, not *to* the system.[4] This is of course still an open area of investigation, but recent data tend to favor Dennett's interpretation (Heyes 1998, Whiten 1997, Povinelli, Povinelli & Giambrone 1998), and I will not go into this debate in any more detail.

- What finally creates our kind of intentionality - where we have explicit, manipulable representations of the minds of ourselves and others - is the specific selection pressures set up by the advent of language and communication. Communicative acts *demand specificity* according to Dennett, and the agent confronted with shaping and executing such an act faces a problem of how to carve up its own "tangle of competing, enhancing, merging, intertwining, behavioral control circuits... into competing "alternatives" (1996, p 127). The only solution available given the impossibility of adding up the sum-total of one's behavioral tendencies over time and presenting them in a manageable format (which would be the same thing as having a full-blown design stance theory of ones own brain and behavior that could be divulged in a two minute lecture!), is to abstract away from the noise and declare a category. Dennett suggests that:

  > This task of carving where nature has not provided any salient joints is a problem the agent solves by what we might call *approximating confabulation*. The agent comes to label its tendencies *as if* they were governed by explicitly represented goals - blueprints for action - instead of trends of action that emerge from the interplay of the various candidates. Once such *representations of intentions* (in the everyday sense of intentions) have come into existence in this backhanded way, they may succeed in convincing the agent itself that it has these clear-cut prior intentions governing its actions. In order to solve its communication problem, the agent has created a special user-interface for itself, a menu of explicit options from which to choose, and then has been to some degree taken in by its own creation (1996, p 128, emphasis in original)

This takes us right back to Dennett 1978, and the distinction between *beliefs* and *opinions*. Beliefs are *patterns* in the behavior of people that one, so to speak, could get rich by betting on, and opinions are, in effect, those wagers. Restating the content of the quote above,

Dennett is saying that the ever ongoing business of predicting, explaining and manipulating the behavior of our fellow organisms is almost invariably conducted by means of linguistic tools - speech and writing, public and private - and that this very process creates a whole new class of *spuriously specific* beliefs and desires (linguistically infected opinions) that exist exclusively in the domain of human discourse (at least on this planet).

- This then provides all the conditions for (runaway) memetic, cultural evolution, which is a process more or less analogous to genetic evolution (even if all the tools of evolutionary biology do not carry over, see Dennett 1995). A meme, just like a gene, is the smallest unit of interest that replicates with reliability and feundicity.[5] The selective environment of memes consists of other memes competing for access to the limited supply of human brains/minds, and whatever "biases" are present in those brains before the first memes arrive(d) (this holds both on a phylogenetic and a developmental level). Dennett has said very little about the original adaptive landscape of memetic selection (then or now). In *Darwin's Dangerous Idea* he argues that sociobiologist who did not realize that memes could be universal not only due to patterns of descent, but also due to the manifest rationality and utility of the meme in question (like the idea of throwing spears with the pointy end first) tended to put too much innate content into our heads. On the other hand he is favorably inclined towards the investigations of Cosmides & Tooby, Keil, Jackendoff and Baron-Cohen (among others) that seems to show that there is an genetic component to (among other things) the concepts of social contracts, ownership, spatial representation, and intentionality (Dennett 1995, p 379). Instead he has put the meme-theory into use in three other (largely overlapping) ways:

1. To explain consciousness. To try to show that "our brains are *not* organized at birth, thanks to our animal heritage, in ways that automatically guarantee the sorts of mutual influence that is the hallmark of "our access" to conscious content" (1994, p 548); how memes might enable a virtual, functional machine in our brains, and how the "self" then may emerge as center of narrative gravity by our incessant private and public storytelling. This is all familiar, and I will not pursue this further here.[6]

2. To explain *how* language enables all this new functionality and "access to conscious content." This is a fairly recently blossomed project as far as I know (see Dennett 1996, 1997a, Clark 1998a), and the interesting new thing is an emphasis on the similarity between words and (other) worldly artifacts as stable *objects* to which we can target our native associative learning mechanisms and enable learning of higher-order relations; the famous software-links of the virtual machine (a prime example of this approach being the series of experiments by Thompson & Oden (1996) in which chimpanzees learn higher order matching task by associating them with stable tokens in the world). Here I think Dennett is in the business of restating, in less cryptic form, the idea of *autostimulation*, or talking to oneself, that formed one of the central pieces in his exposition of the evolution of cognition in *Consciousness Explained*. One important difference is salient though. He has lately begun to put much more emphasis on the fact that the skin is not an important boundary when it comes to locating cognition, and that there are immense cognitive gains to be had by offloading memory and decision-procedures into the world (a problem Clark (1996a), and Clark & Chalmers (1995) vividly discuss under the heading of cognition and mind "leaking" out into the world), as well as starting to appreciate all the fundamentally cognitive, social things in our environment. He can *almost* agree with the words of Dahlbom:

   > As a *brain* process, thinking is a natural process, and this makes it difficult for us to see that thinking today is about as artificial as anything else - communication, production, consumption - in our modern artificial world. Just as our society will grind to a halt when our artifacts break down, so thinking would be reduced to next to nothing were we to suffer a breakdown of our intellectual artifacts. What could we think, how could we reason, if we did not have words, books, diagrams, figures, concrete examples, algebra, logic, lisp, or legal systems? (1993 s.170)

3. To explain how language greatly enhances our powers of mental simulation, or *preselection* (which allows our hypotheses to die in our stead) in our inner environments. Here Dennett again leans hard on theoretical evolutionary considerations. Given the fact that the worldly knowledge used for preselection not only has to find its way into the head of the organism in question (by descent or learning), but also get itself into a position where it can influence global behavioral output, and that the primary mode of operation for animal cognitive architectures is highly decentralized (e.g. the snake example above), preselection is probably not going to come for free. He writes:

> This is where *earlier* decisions come back to haunt - to constrain - the designer. In particular, choices that evolution has already made between need-to-know and commando team [i.e. distributed or centrally organized systems] now put major constraints on the option for design improvement. If a particular species' brain design has already gone down the need-to-know path with regard to some control problem, only minor modifications... can be *readily* made to the existing structures, so the only hope of making a major revision of the internal environment to account for new problems, new features of the external environment that matters, is to *submerge* the old hard-wiring under a new layer of pre-emptive control... It is these higher levels of control that have the potential for vast increases in versatility. And it is at these levels in particular that we should look for the role of language... in turning *our* brains into virtuoso preselectors (1995, p 377, emphasis in original).

Although I do not have space to elaborate on it here it is interesting to note that Dennett's line of reasoning on mental simulation cannot lay claim to being the only respectable alternative that invokes evolutionary plausibility for support. For example Grush (1997, to appear) and Clark & Grush (1997) maintain that in solving the problem of skilled motor reaching, the mammalian brain has most likely availed itself of a solution that depends on *emulation* - "a mechanism (circuitry, software routine, whatever) that takes as input information about the starting (or current) state of a system... and about the control commands that are being issued. The emulator then gives as output a prediction of the next state of the system... The emulator thus models the target system and generates a kind of mock feedback that can be used instead of the laggardly feedback from the real system (Clark & Grush 1997, p 2)." In a countermove to Dennett's argument from plausible constraints above, they consider it to be a fairly *small* step (in evolutionary terms) to have such emulators running *off-line* for use in mental simulation, or preselection, and that a full scale migration of this type of architecture from motor skills to other domains of knowledge is not *that* hard to imagine (for more on this line of thinking, see Wolpert 1997, Wolpert, Miall & Kawato 1998, Jeannerod 1994, and Arbib & Rizzolati 1997).

2.1. Dennett's defense of intentional psychology

Adding all these pieces together, does it now make for a stable and unequivocal picture of the mind? For example, does the folk-psychological *agent* really survive unscathed in all this pandemonium, distribution, and cheap modulation? Has Dennett, so to speak, managed to balance the agent on the edge of chaos (using a well-known slogan of Kauffman 1993), or is it perhaps more plausible to think that sometime in the future when we paste the whole empirical picture together again, there will be enough cracks and gaps to warrant a much greater downplaying of our rationality, and the "unity" of the mind, compared to what Dennett so far has allowed (see Brown 1997 for a nod in this direction)? I think this situation can be profitably compared to the ruminations of Dawkins (1982) regarding the consequences of seeing the world of biological evolution through the theory of the extended phenotype. First he says he can dimly imagine a boundless crisscrossing network of replicator influences, inexplicably interlocked in, through and around the bodies of animals and plants, but then he quickly reassert the agent as a causal nexus in the image of bodies being transparent and the genes as neatly packaged glowing bundles inside. Even if its clear that both perspectives are correct in some sense if Dawkins theory is correct, the reader is left in the dark as to how to translate between them (Wilson & Sober 1994, with commentaries). In the case of Dennett this move is even more distressing because he has arguably done more then *dimly* imagined the mind as an evolving, distributed, fragmented, pandemonium system - before again drawing in on the whole agent/person, as a reassuring explanatory unit. We all know that Dennett has worked incredibly hard to dispel overgrown homunculi from subpersonal theories of cognition (even though he harbors no hope of

dispelling them all, given that ascription of intentional content to lesser homunculi is his own way of dissecting the brain, more on this below), but as I see it he has also worked pretty hard to salvage his own version of the rational, coherent, intentional agent (see for example 1991b,c). Since Agents with capital A:s do not exist, intentional agents have to be created somehow. Dennett's answer to this question is, as stated above, that the demands of communication and cultural evolution set up selection pressures for the creation of a special kind of linguistically infected beliefs; that these selection pressures also yield the use of intentional terms as an indispensable tool for prediction and explanation; and the outcome of all this incessant explanation and storytelling is a distinct self (or sometimes many selves, as in the case of multiple personality disorder) at the center of narrative gravity. At the same time, he stresses that an agent so created has no special privileges as to how to describe its own behavior. There is to a large extent *causal symmetry* between explanations by agents and observers:

> We postulate all these apparent activities and mental processes in order to make sense of the behavior we observe - in order, in fact, to make as much sense possible of the behavior, especially when the behavior we observe is our own /.../ each of us is in most regards a sort of inveterate auto-psychologist, effortlessly *inventing* intentional interpretations of our own actions in an inseparable mix of confabulation, retrospective self-justification, and (on occasion, no doubt) good theorizing (Dennett 1987 p.91, emphasis in original).

Now, to be able to call these interpretative practices the work of *rational*, *coherent* agents it also has to be the case that their heterophenomenological worlds - the things they say about what goes on in their minds - are not full of falsehoods. Maybe it turns out that the real patterns in the tales told by everyday people about what they believe, how they came to those beliefs, and so on; in short, what they *do* in their minds, correspond *very* poorly with what we find going on in their brains. What I would like to suggest is that Dennett himself in his enthusiasm for hormonal wisdom, subsumption architectures, cheap modulation, pandemonium, and cognitive artifacts has provided all the tools necessary for opening up this as a serious possibility. I suspect that for Dennett, the folk-psychological agent as such has never been in question, just what kind of explanation should be given of it (bureaucracy or pandemonium, distribution or centralization, etc.), and that he has not been alert to the possibility that his own successful endeavor to explain it in terms of pandemonium and distribution might also erode the very thing he is trying to explain.

It is important to realize that this is not an attempt at pitting different levels of description against each other. I am not out, like the Churchlands, to eliminate the use of intentional idioms because they have no real correspondence to anything in our connectionist brains. Given Dennett's instrumentalism (or patternalism) about beliefs we already know that there is nothing in the brain literally corresponding to, or *resembling* the terms of folk-psychology (so they are of course in principle eliminable). The theory of *content* that Dennett espouses for whole persons he espouses all the way in. That is, the principles of interpretation that are used to endow whole persons with content are used for subpersonal parts as well:

> The way in which personal-level attributions of belief gets confirmed (in the crunch) by subpersonal attributions of (nonordinary) intentional properties is roughly parallel to the way in which one might confirm one's attribution of culpable motives to say, the British Empire, or the CIA, or IBM, by discovering a pattern of beliefs, desires, intentions, among the agents whose joint activity compose the actions, beliefs, and intentions of the superpersonal agent (Dennett 1994, p. 528)

Using an example from Rudder-Baker (1994), and Dennett (1994), the procedure might look like this: Eve is on the phone talking to a researcher from the Intentional Truth Squad. Describing an incident in her home this morning she says: *I was suddenly conscious of the fact that I was not alone in the house.* In doing so she reports that she earlier had a certain first-order thought to the effect that she was not alone in the house. Now, did she really have such a thought? The ITS people staking out her house are skeptical. They have further behavioral evidence beyond her verbal report. The secret videotape of her behavior at the time reveals no visible signs of apprehension, no abrupt changes of pace or direction in her movements, and so on. Still, it is hard to counter her own sincere avowal to the contrary, but their suspicions are confirmed when the sensor readings from the neurocryptography lab finally arrives - according to their careful brain-mapping none of the homunculi in Eve's brain regularly engaged in detection of the presence of others (for example those hooked up to the vertical symmetry detector in the vision system) sounded any alarms during the period during which she claimed to have had the thought in question. Even though it seems counterintuitive Eve is clearly confabulating in this case.[7] The question now becomes: to what extent do real people engage in similar behavior, and how are we ever going to confirm if this is the case?

We know that intentional ascription is a *noisy* process that abstracts away from the messy details of the underlying cognitive operations, so that a more accurate explanation for a behavioral event always

could be manufactured if one is able and willing to incur the computational cost. We also know that in case of brain malfunction (when the brain as a syntactic engine no longer approximates a semantic, rational engine); the explanatory tools of folk-psychology fall short. What is at issue here though is whether we *regularly* misconstrue the happenings in our brains even in cases of normal, everyday operation. Before I go on to try to answer this question it should be noted that these hypothesized daily confabulations could have more or less serious implications for Dennett's theory, depending on exactly what is wrong with the intentional story:

- In the case of brain malfunctions, intentional explanations can interact with the deficit in question in several different ways. Take for example Capgras delusion - the belief that one or more of one's close relatives has been replaced by an exact replica or impostor. It is a completely baffling delusion from a folk-psychological perspective. How can one come to believe such a thing? It is not that it is impossible to imagine a story to account for the belief (it could for example be the work of an elaborate conspiracy), but rather that no such story would have any credibility whatsoever. The tentative neuropsychological explanation for Capgras offered by Stone & Young (1997) is that we have two visual recognition pathways in the brain, one explicit and one implicit, or one for recognition and one for valence, and that the right sided cerebral hemisphere injury (in the occipito-temporal and temporo-parietal regions, especially those involved in face perception, see Bruce & Young (1998)) suffered by almost all Capgras cases have damaged the second pathway. The outcome of this for the patient is that he/she still recognizes his family and peers, but that all the affection, or somatic markers are lacking (to use the terminology of Damasio). Now, in the light of the extraordinary weird perception brought about by the lesion, the intentional hypothesis that people in his/her close surroundings are impostors or aliens does not seem so incredible after all. It is simply an attempt to *make sense* of an overly strange world. Still, it might actually be the case that many more people than reported are suffering form the secondary affective loss that Stone & Young (1997) describes, but that their standards of attribution are more rigorous or conservative than the Capgras cases - i.e. faced with the strangely affectionless responses to their spouses, family and friends, they blame *themselves* rather than the world, or just settle for the fact that things feel weird, and thus they never show up in the medical literature (Stone & Young indeed hypothesizes that the Capgras cases involve additional damage to neurological systems involved in "reality testing"). Another favorite neurological deficit in the literature is prosopagnosia - the inability to *explicitly* recognize faces. Stone & Young (1997) consequently hold that the neurological damage involved in prosopagnosia is the inverse of that of Capgras, and many test have also shown that the ability to implicitly recognize faces is intact (as measured by forced-choice or inclusion, exclusion tests, etc.). Again this is a disorder that is completely unexplainable from a folk-psychological standpoint. There is nothing rational about not being able to recognize faces, and in this case it is *really* hard to come up with a background story to explain why it is rational after all (unless one thinks the prosopagnostic patient is feigning the whole thing).[8] For Dennett none of these cases are very damaging. Brain malfunctioning is after all exactly one of those situations where rationality is *expected* to break down (for more examples, see Stone & Young 1997, Ramachandran et al. 1996), and even I would be hard pressed to accept that a large fraction of the human population is walking around with neurological damage that has been covered under a thicket of ingenious intentional storytelling (it is possible though...). The reason I bring them up is because they highlight the possibility of malfunctioning that is not situated *in* the mechanism itself, but in the mismatch between it and the *environment* in which it is supposed to work. Mechanisms cannot be made to be rational in every context, and depending on its computational nature it might produce wholly different results when moved out of its intended domain of functioning. In section four below I explore the possibility of humans being not so adapted (or even positively maladapted) for certain environments, and how this might have a negative impact on the accuracy of the intentional ascriptions that are produced in those environments.

- An intentional story could also be wrong because the behavior it tries to rationalize is produced by a dumb killjoy mechanism. At the level of intentional explanation, on Dennett's own account, we really *ought* to revise our attributions of intentionality downward if we find too simple - killjoy - mechanisms at work. The primary criterion for something being a killjoy mechanism (there are no

> *defining* conditions of course) is that an intentional stance approach to the system gives us less predictive leverage than a design-stance approach, and that we do not get a much more computationally tractable instrument by treating it in intentional terms (translated to behavioral criteria it normally means that the mechanism is inflexible or limited in scope). Here I believe Dennett is in one sense working against himself when he is so skeptical of the prowess of animals, because every piece of evidence that he digs up to show the "dumbness" of homunculi involved in animal cognition might also be taken as an invitation to extend the findings to humans, and then use them in an engineering spirit to reveal the circumstances under which the rational agent falls apart. In section four I review some evidence to support this conclusion.

- A further way our intentional storytelling could be regularly off track is if the underlying adaptations are implemented in such a non-obvious way (especially in regard to their distribution between brains, bodies, artifacts and the wider social environment) that it becomes almost impossible to describe them accurately in intentional language - both in terms of it being difficult to actually divine the real cause, and in terms of it being difficult to describe it even if it was known. I pursue this line of argument in some detail in sections three and four.

3. Memes *and* Modules

How could it then possibly be the case that our heterophenomenological worlds are full of falsehoods? The answer lies in Dennett's reply to Millikan (1993) and Dahlbom (1993). They urge on him a dilemma between viewing thinking as a biological, modular, brain process (Millikan) and as an artificial, distributed, social phenomenon (Dahlbom, see quote above), and Dennett (1993) tries to use the theory of memes to reconcile the two perspectives - i.e. memes are both natural and artificial: natural by evolving on top of, and within the biological substrate, and artificial by comprising a world of wildly diverse artifacts. In Dennett's version we end up having a little bit of "nature" and a little bit of "nurture"; what he (pace Barkow et. al. 1992) calls his *Only Slightly Nonstandard Social Science Model* (1995 p 491). I agree with Dennett on this. Memes vs. Modules is not the most fruitful way of framing an attack on his theory (see Whitmeyer 1998 for such an attempt), but that does not automatically mean he has insulated himself against my type of approach. What should be obvious by now is that I instead think that *both* Millikan and Dahlbom are right, that they are even *more* right than they themselves think, and that the whole *person* might be the one "suffering" from this.[9]

3.1. What is in the head?

Starting from the biological side the first trend I want to review is that of putting ever more content into the head. Using the position of evolutionary psychologists like Cosmides & Tooby (1994a,b) rather than Millikan (her brand of functionalism being a bit too cryptic), we may asks ourselves: how are we to cash in all the functional "Darwinian Algorithms" that have been suggested to operate in human cognition in the currency of causal, developmental mechanisms? The manifesto of Barkow et al. (1992) was mainly a ferocious theoretical attack on the canons of Standard Social Science (Holcomb 1996). When it came to experimental data they provided a nice and suggestive lists of possible adaptations, but when presenting more detailed accounts they leaned hard on the subject areas of perception (where many years of successful neuropsychological research into modularity backed them up), language (where Chomsky and Pinker pulled the weight), and social *contract* reasoning (where they presented their revised Wason-experiments). With regards to these areas the picture of development that was hinted at was one of setting parameters, critical periods, poverty of the stimulus, inference engines (that had to be contentful to avoid combinatorial explosion), and of robustness in the face of environmental perturbations, but not much was said about actual mechanisms (Mundale & Bechtel 1996).

I think it is fair to say that since then the field of evolutionary informed experimentation and speculation has exploded.[10] To provide a flavor of the domains of behavior that are under investigation I quote at some length from Cosmides & Tooby (1994a):

> Textbooks in psychology are organized according to a folk-psychological categorization of mechanisms; "attention", "memory", "reasoning", "learning". In contrast, textbooks in evolutionary ecology are organized according to adaptive problems: foraging (hunting and gathering), kinship, predator defense, resource competition, cooperation, aggression,

> parental care, dominance and status, inbreeding avoidance, courtship, mateship maintenance, trade-offs between mating effort and parenting effort, mating systems, sexual conflict, paternity uncertainty and sexual jealousy, signaling and communication, navigation, habitat selection, and so on /.../ Twenty-first-century textbooks on human cognition will probably be organized similarly (Cosmides & Tooby 1994, p 53).

Within this field, developmental psychologists have perhaps been most successful in digging up kinds and varieties of "innate" content, and in his latest discussion of memetic evolution, Dennett (1995) allows that the (proto)concepts of being alive, ownership, spatial representation, and theory of mind have a "genetically imposed head start in the young child's kit of mind-tools" (p 379). To this list we might add numerals (Hauser 1997), self-propelled motion (Spelke, Philips, & Woodward 1995, Premack & Premack 1997), naive physics (Wellman, Hickling & Schult 1998), causal classifications of biological kinds (Wellman & Gelman 1997, Atran, to appear), mother-infant attachment (Kraemer 1995, Wilson & Daly 1997), and "fast mapping" in word learning (Bloom & Markson 1998).

Another large chunk of proposed innate content comes from the domain of sexual selection theory (what many regard as the current flagship of evolutionary psychology). Specific cues that influence choice of sexual partners universally across different individuals and cultures have been reliably identified, among them are; height, intelligence, walking speed, male jaw size, waist-to-hip-ratio, degree of genetic relatedness, kindness, scent, full lips, facial symmetry, and political status (from Miller 1997a,b,c, Miller & Todd 1998, see also Jones 1995, Cunnigham et. al. 1995). This might seem like a motley collection, but there is certainly no lack of theoretical resourcefulness around to explain these findings (see Miller & Todd 1998, Gangestad & Thornhill 1997, Möller 1997, Möller & Thornhill 1998 for detailed accounts).

Sex is not the only thing that drives us though, and Cosmides & Tooby (1997), and Gigerenzer (1997) argue persuasively that we have a whole adaptive faculty of *social cognition*, containing a slew of specialized inference procedures evolved to deal with dominance hierarchies, social contracts, permission schemas, precaution rules, and threat detection (the general who-did-what-to-whom-and-why structure of gossip that Dunbar (1997) identifies as one of the driving forces in the evolution of language).[11]

It is extremely hard to evaluate to what extent the results coming out of evolutionary psychology (widely conceived) are genuinely solid. A fair part of the results from the area of sexual selection theory suffer from the limitations stemming from heavy reliance on survey and questionnaire methodology, and many hypotheses are interesting just-so-stories but have not been put to any serious experimental tests (which of course is difficult given the fact that our society is rather different than that of hunter-gatherers). For the purposes of this paper it is enough that these results are very promising (and that they rest on *theoretically* solid ground), and that Dennett has not made any attempt to gauge in what manner they may affect his theory. The question for me is in what manner all this content can be said to be *innately represented*, and how this in turn might bear on Dennett's position. Lately there has been a large amount of discussion of innate content in various quarters of cognitive science, and the end result looks very interesting.[12] Take for example the classic, prototypical example of innateness - the imprinting behavior of chicks and ducklings. How does it work? It is common knowledge that under experimental conditions the newly hatched chicks can be fooled to imprint on just about any moving object. This should alert us to the fact that in order for the imprinting behavior to be an adaptation (which no one doubts) this kind of outcome must be regularly ruled out. What has been found is that two different neural systems are specifically involved in generating the behavior (see Clark 1998b, and Elman et al. 1996 for the details). The first (A) makes the chick prefer stimuli that display certain neck and head configurations common to roughly equally sized birds and mammals, and the second (B) develops a representation of the object attended to, so as to allow stable tracking under varying perceptual conditions (see Dennett 1996, Millikan 1998 for discussion of the significance of tracking as a prerequisite for fancier cognitive skills). No internal communication has been found to occur between these systems. B learns to identify the object of imprinting solely by being exposed to heavy doses of it. Here we thus have the kind of cheap modulation so persuasively put forward by Dennett (1996). The system makes a heavy bet on the environment to provide the right contingencies for it to work, and against a normal ecological backdrop this it of course what it does, partly because it has been *selected* to do so - for example the behavior of the mother hen and the location of the nesting site (Gray 1992, Griffith & Gray 1994). This fits well with evolutionary psychology, because it has always been out to explain primarily the *environmental* sources of variance in human behavior (Daly 1996, more on this below), to explain how our universally shared facultative adaptive algorithms "kick in" and interact under the input of complex environmental variables (compare the

discussion in Cosmides & Tooby 1992 of the concept of *evoked culture* and the example of food-sharing[13]). This original form of essentially *interactive* innateness was however downplayed by Cosmides & Tooby (1992, 1994) when it came to distinctly human skills, and instead they left the reader with the impression that human development was all about rigid chomskyean data structures, parameter setting, critical periods, etc.

3.1.1 Rethinking Innateness

With a little help from Elman et al. (1996) this impression can be countered. They presents a powerful conclusion to the effect that the literal content uncovered in all these experiments and studies in developmental and evolutionary psychology could not have been *pre-specified* in the cortex to begin with (what they term representational nativism). According to Elman et al. we have to separate the *mechanisms* of innateness from the *content* of innateness (as Karmiloff-Smith 1994 remarks, the domain-specific know-how uncovered in many of the studies mentioned above could equally well be the *outcome* of a learning process, as a description of the mechanism behind it). They present a tripartite argument for this position.

First they provide a range of existence proofs (well, analogies) that show how knowledge profiles with tightly structured domains, similar to those found in real life studies can be approximated by connectionist simulations in which the starting-state specifications of connectivity and the ensuing learning tries to mimic a developmental sequence devoid of genetically controlled hard-wired content. This of course only means that they did not spell out, or hand-code *fine-grained* content into their networks, not that the network had no biases at all (the old blank-slate of empiricism). Instead Elman et al. relied on global architectural biases (density of connectivity, structural topography, etc.), variations in timing of growth, and variations in the rate and presentation of the learning set (more on this shortly).

Second, they present detailed discussions of real world examples (like the chick imprinting above, or children learning to walk) where first impressions (and intuition) suggested large amounts of prespecification, but where investigations have revealed that the developmental story did not involve the content being there - at a representational level - to begin with.

Thirdly, and most important perhaps, they present evidence from comparative vertebrae brain-development that suggest that "innate specification of synaptic connectivity at the cortical level is highly unlikely" (1996, p 361). Further evidence for this conclusion comes from studies of human development (Quartz & Sejnovski 1997), large scale-neuronal computation (Phillips & Singer 1997), and from successful cross-species transplants (xenografting) of neural-tissue (Deacon 1997).

To be sure this is not the final word on "representational nativism." Elman et al. do not offer too much in the way of argument for why detailed knowledge - as a matter of empirical fact - is only realizable as fine-grained patterns of synaptic connectivity at the *cortical* level. There are many other brain areas that might serve this purpose (although the cortex is certainly the most obvious place to look). This caveat aside, what is more important is that the evolutionary, historical, *adaptive* aspect is almost totally absent from the discussion of Elman et al. The content identified in the experiments above does not cease to be domain specific adaptations (if they indeed *are* adaptations) just because it can be shown that the knowledge or skill was not initially specified in the head. Constraints flow from both the level of brain architecture *and* from the level of actual behavior (Estes & Bartsch 1997). Elman et al. should be seen as a "fascinating demonstration not of the inapplicability of the notion of innate knowledge but of the unexpected variety of ways in which nature might gift a being with innate knowledge of its world?" (Clark, to appear). Instead we are invited by Elman et al. to the exiting project of telling the story of how subcortical and environmental structures, together with constraints arising from architecture, timing, and the physical nature of peripheral organs "enslaves" the cortex (see Barton 1997 for a short discussion of such enslavement by the periphery) to yield the outcome produced in the experiments mentioned above. What is important here is that this kind of process will most likely not be easy to describe in ordinary "representational" terms. This is so partly because of the distributed character of the process and our lack of research tools to adequately describe such events (Hutchins 1995), and partly because the crucial interactions might not be the ones that are most intuitive or obvious (just like the mechanism of the hunting behavior of snakes described in paragraph 2 makes no sense from an "intuitive" point of view). A good example of this comes from Deacon (1997). In a discussion of language learning he makes the point that the whole traditional setup of the problem is fundamentally flawed just because it does not recognize that the "learning" might be unintuitively distributed between the children and the evolving environment:

Children's minds need not innately embody language structures, if languages embody the predispositions of children's minds /.../ Languages don't just change, they evolve [and here he refers to

memetic evolution as described by Dennett]. And children themselves are the rigged game. Languages are under powerful selection pressures to fit children's likely guesses, because children are the vehicles by which languages gets reproduced. Languages have had to adapt to children's spontaneous assumptions about communication, learning, social interaction, and even symbolic reference, because children are the only game in town. It turns out that in a curious sort of inversion of our intuitions about this problem, languages need children more than children need languages (1997, p 109).

An illustration of this kind of process is provided by Clark (1996b). He discusses an interesting series of simulations with genetic algorithms by S. Nolfi and D. Parisi where pairs of coupled neural nets passing "teaching signals" to each other in order to find a food source were allowed to evolve by natural selection in a population (which could be seen as a stand in for the coupling between for example infants and mothers, only with the difference that the fitness interest of each net was completely overlapping in the simulations, which is a highly unrealistic assumption for biological systems). Clark writes: "if we ask what exactly is the content of the innate knowledge embodied in the initial weights of the evolved nets, we find ourselves talking... of one net's knowledge of the others location in weight space, of initial positions which avoid local minima, etc, etc.: i.e. we are confronting a form of innate knowledge which is fully independent of the usual resources of daily language (p 126). Indeed I think this kind of innate knowledge is not only fully independent of the usual resource of daily language in cognitive science research, it is also most likely fully independent of the resources of intentional ascription, explanation, and "rational" decision-making. When the causal story is so cryptic that we have great trouble describing it even in low-level functional, computational terms, then how could a coarse folk-psychological abstraction capture the same process with any accuracy at all? I return to this question in paragraph 4 below, but before facing that discussion I introduce two more lines of argument congenial to my quest.

3.2. Rethinking Adaptationism

In their original formulation of evolutionary psychology Barkow et al. (1992) were very careful to stress the *non-genetic* nature of their research program. (see also Tooby & Cosmides 1990) We ought *not* to expect *adaptive* genetic variance among humans (except perhaps a thin filter of recent changes, like for example lactose tolerance among Europeans), because if the abilities in question were fitness-related then selection should tend to eliminate the variance (like it has done with regard to the possession of two eyes). Above that sexual recombination should tend to break up the large gene complexes that would be needed for genetic polymorphisms (and intermediate forms would be non-functional, thus strongly selecting against this trend). Under this conception the things studied in the neighboring field of behavioral genetics became largely uninteresting. A trait or behavior showing high heritability only indicated that it was due to the large amount of quantitative, non-adaptive variance that is known to exist (in for example eye-color), or to unknown environmental influences on the measurements. For someone in the field of behavioral genetics these arguments simply did not make sense though, given that just about every trait studied showed moderate, broad heritability (Gangestad 1997, Plomin et al. 1997), and the arguments of Cosmides & Tooby have now been shown to be far too strong in their conclusions. Adaptive genetic variations can be upheld both by sexual selection for reliable indicators of genetic fitness, and by assortative mating (Miller & Todd 1998, Miller 1997a), and by frequency dependent selection for distinct phenotypes (this is true even if intermediate types are selected out, Wilson 1994, 1997). Even if this sort of selection is weaker than traditional natural selection it does not follow that it must be unimportant: "Clearly, the importance of individual differences at the phenotypic level does not depend entirely on the complexity of the trait at the genetic level... local adaptations is a possibility for many human traits that are interesting from the intellectual and practical standpoints" (Wilson 1994, p 227). As an example take Mealy (1995) and her hypothesis that sociopathy is a permanent "cheater" strategy that has been maintained in the gene-pool by frequency-dependent selection, and which is regulated by a comparatively simple mechanism of discounting future expected utility (or rather inability to resist immediate rewards, see discussion in Deacon 1997 for experiments that demonstrates the immense difficulties apes and monkeys have in choosing future benefits over immediate rewards). She estimates that sociopaths, which comprise only 3-4% of the male population (and less than 1% of the female) are accountable for approximately 50% of all crimes in the united states (1995 p 523), which should be a large enough figure to make it interesting from a practical standpoint.

I do not intend to get into a detailed discussion of the results from behavioral genetics since I lack the knowledge to evaluate it fully (see Plomin et al. 1997 for an overview, and read Devlin, Daniels & Roeder 1997 for a suggestion that prenatal environmental influences have been greatly underestimated),

instead I want to suggest that here as well the results make much more sense when viewed from the explicitly interactionist perspective of development provided by Elman et al. (1996).

Two kinds of extended, interactive processes of reactive heritability can thus perhaps explain the puzzling fact that dizygotic twins raised in the same family often turn out very different, and monozygotic twins in different families often turn out very similar (Plomin et al. 1997). The first thing that we must realize is that features of the world *and* features of oneself (ones individual characteristics) may be assessed by an adaptively designed system. People who are big and strong might for example fare better in enacting an aggressive social strategy than relatively small individuals, and that is a stable feature of the environment to exploit for an adaptive system (this is of course also a stable feature of the world to exploit for a system that is not totally dumb, so I am not implying that it *must* be an adaptation). The second thing to notice is that similar characteristics tend to evoke similar environments, and that adaptive, genetic variance might exist that *provokes* a stable and reliable environmental backdrop that is subtly different from that of other individuals (so for example has it been suggested that most attempts at explicitly and knowingly teaching children to be alike, for example in kibbutz environments, have failed not primarily because the children were of different aptitude, but because the teachers treated them different without knowing it). If one does not have this kind of refined concept of shared environment (or extended phenotype) the findings from twin-studies would point towards a surprising amount of genetic prespecification - *even* though the monozygotic twins are placed in environments that on the face of it are very different (although they of course are similar in the important sense that the families involved have passed the filter of having been allowed to adopt at all) they turn out the same, and *despite* the fact that the dizygotic twins inhabit (at least) the same home environment they turn out very different. If we include reactive heritability in the picture we realize that at a less coarse scale the dizygotic twins do not experience the same environmental inputs at all - depending on their temperament and physical appearance (etc.) they evoke different responses from their parents (whether they know it or not, and even if they try to treat them similarly), and depending on how their Darwinian algorithms gauge their own abilities they (later) tend to invent or construe or place themselves in certain form of environmental situations - thus *exposing* themselves to different stimuli (starting a positive feedback loop that then serves as input to the species typical cognitive architecture). Conversely the monyzygotic twins separated at birth tend to evoke and seek out the same environmental inputs, and consequently share very similar environments.

Again, given that we have only recently begun to appreciate this kinds of convoluted gene/environment interaction, and still have very little to say about the specifics, it is hard to see how the knowledge thus represented could have made its way into folk-psychology and the reasons that go into our everyday intentional attribution and explanation - this seems to be information that is about as explicitly represented in our cognitive system as aerodynamics is in the wings of birds (to use an example from Dennett 1996).

3.3. What is outside the head?

The second trend I want to bring to the fore is the one so eloquently described by Dahlbom in the quote above - that of studying the *cultural* distribution of cognition among brains, artifacts and institutions. Dennett himself has been very keen on adopting this view since it effectively deflates some of the claims made on behalf of the traditional Fodorian symbol-crunching mind he has been combating over the years. If we look at the wider social system of artifacts we indeed find our serial, symbolic processes (Olsson 1994). On a polemic note Dahlbom writes:

Thus when Millikan argues that the language of thought, for reasons of organic alization, cannot be very language-like, she is not really saying anything about the medium in which we think. We think in all kinds of artificial media of course. We should not take her advice and "abandon the traditional image of the computing of mental sentences as like a formal system unfolding on paper", since it is on paper that much of our mental inference are unfolding" (1993, p 169).

Dennett has so far mainly used the notion of distributed cognition to argue for the utility of offloading memory and incorporating artifacts (like a pair of scissors) as invaluable *evolved* extensions of individual cognition, but "extension of individual cognition" is really not such a happy phrase, because it only connotes the massive gain in *functionality* made possible by the artifacts, and says nothing about the nexus of the causal process. These tools are not just *passive* instruments for cognition, and sometimes it is clear that most of the "cognition" is actually situated in the world rather than in the brain (see Ballard et al. 1995, 1997, Kirsh & Maglio 1994, Zhang 1997, Zhang & Norman 1994, Brooks et al 1998 for more on this). So far Dennett has not touched upon the ever so tempting notion of *group cognition* (except in his

discussion of science as the first really non-myopic evolutionary process). Perhaps we should be skeptical of such notions on grounds of the dubious history behind the concept, but recent and more sophisticated versions of group-selection in evolutionary biology and careful studies of culturally, evolved group-processes have convinced me that it deserves another airing.

My purpose in bringing up group cognition here has of course to do with a belief that it could play a role in diminishing the status of the rational, folk-psychological agent (as conceived by Dennett). Therefore I will not focus on those aspects of group cognition that might be most obvious, like the fact that more people can do more stuff, or that different people do different stuff in different ways (division of labor), and that these skills can be traded in ways that make it better for all involved - these are benefits we are all too aware of (see Clark 1997a, Kennedy 1998). Instead I want to bring attention to the potential interaction of memes and modules in producing genuinely distributed but non-obvious cognitive effects. In his discussion of the Hutterite sect (Dennett 1995) - a sect born in sixteenth century Europe explicitly "engineered" so as to curtail all expressions of selfishness and dissident thinking[14]- Dennett argues that this is a quite striking case of an evolved *cultural* phenotype:

These declarations make it clear that one way or another, Hutterite social organization is the effect of cultural practices quite vigorously arrayed *against* the very features of human nature Wilson & Sober wish to deny or downplay; selfishness and openness to reasoning. If group thinking were really as much a part of human nature as Wilson & Sober would like to believe, Hutterite parents and elders wouldn't have to say a thing. (Compare this to a case in which there truly is a genetic predisposition in our species; how often have you heard parents cajoling their children to eat more sweets? (1995 p, 474, emphasis in original)

I think this is a reasonable characterization of the Hutterite case, but I think it neither establishes that we have no "biological" group-phenotype adaptations, nor that more subtle "cultural" group-phenotypes are to be found among us common folks (it is really hard talking in ways that does not introduce the memes vs. module perspective). In the account of the Hutterites presented by Wilson & Sober (1994) it seems to me that this case-study was not supposed to carry much evidential load (given the extreme nature of the Hutterite social organization I think it was actually a very badly chosen example). The main inference was instead to proceed from shared fate among group-living individuals in ancestral conditions, to a modern human tendency for group adaptive behavior - i.e. behavior that has evolved for group-selective reasons, but that benefits all individuals in the group (here the individuals in the group stand to the group like genes stand to the organism, although the group clearly has very rudimentary and shallow adaptations compared to a single organism). In the words of Wilson & Sober:

Since humans have lived in small groups throughout their history, it is reasonable to expect the evolution of psychological mechanism that cause them easily to become "team players" in competition with other groups. We do not expect these to be the only motives that guide human behavior, but rather a module that is faculatively employed under appropriate conditions. In fact, there is abundant empirical evidence that humans coalesce into cooperative teams at the merest suggestion of a metapopulation structure in which groups can compete against other groups (1994, p 601).

With a slight flavor of overstatement they are saying that we have a tendency to become *groupish* (as compared to selfish) under certain circumstances (Baumeister & Leary 1995, Wilson, Near & Miller 1996, Richerson & Boyd 1997). The first thing to scrutinize in the chain of inference that leads to this conclusion is whether the description of ancestral conditions holds up. I am not the one to fully evaluate this claim either, but at least it fits nicely with the views of the leading evolutionary psychologists Cosmides & Tooby:

Ancestral hominids were ground-living primates; omnivores, exposed to a wide variety of plant toxins and having a sexual division of labor between hunting and gathering; mammals with altricial young, long periods of biparental investment in offspring, enduring male-female mateships, and an extended period of physiologically obligatory female investment in pregnancy and lactation. They were a long-lived, low-fecundity species in which variance in male reproductive success was higher than in female reproductive success. They lived in small, nomadic kin-based bands of perhaps 20-100; they would rarely (if ever) have seen more than 1000 people at one time; they had little opportunity to store provisions for the future; they engaged in cooperative hunting, defense and aggressive coalitions; they made tools and engaged in extensive amounts of cooperative reciprocation; they were vulnerable to a large amount of parasites and pathogens (1994b, p 53).

The second thing to critically appraise is whether conditions for shared fate were strong enough to overpower individual (or genic) level selection. Boehm (1997a,b, see also Knauft 1991) argues that in the community living in ancestral times shared fate was explicitly *induced* and *enforced* by cultural

mechanisms. He calls this the *egalitarian behavioral syndrome*, and considers it to be an cultural adaptation (a meme-complex) that has been prevalent long enough for genetic changes to have taken place (i.e. in the vicinity of at least 100,000 years). The egalitarian behavioral syndrome has in Boehm's account three main (one might say) *moral* components. The first one is suppression of competition and leveling of hierarchies within the band by the sharing of food and resources. The second is consensus seeking at the group level on important decisions of foraging, dispersion, aggression, and environmental threats (drought, etc.), as well as internal social disruptions (for example willful deviance or insanity), and explicit punishment of free-riders.[15] This would then constitute evidence that we could have evolved to exploit group dynamics when the situation called for it. Like Wilson (1994) does when it comes to adaptive, genetic polymorphism Boehm stresses that such adaptations need to be very complex to have substantial effects. An interesting comparison can be made to the more substantial group level adaptations of honeybees and other eusocial insects. Seeley (1997) has found that the behavior of individual bees in service of group-level effects tend to be very simple, while the outcome at the group level (foraging decisions, etc.) is very complex (incidentally this is also one of the main points of Elman et al., they stress that insignificant changes at the level of mechanisms can produce dramatic effects at the level of behavior). Here the comparison to the Hutterites becomes positively misleading, for as Wilson (1997) remarks, given the simplicity of such behavior we might not even notice it as a fundamentally group-oriented event. Indeed, if we take Hutchins' (1995) analysis of team ship navigation seriously we should instead *expect* such group-level processes to be non-transparent to the individuals participating. This is so for the very reasons Dennett himself has provided in his discussion of the architectural principles involved in biological evolution (1995, p 373). Since ship-navigation in the real world is not an explicitly designed activity, but instead a practice that has coevolved with the advent of new technologies, we should expect that a version of the need-to-know principle of espionage (see paragraph 2 above) works similarly, one level up, at the group level.

What I am getting at here is that a lot *more* distribution of cognition seems warranted than Dennett so far has allowed. If we indeed have evolved to "coalesce into cooperative teams at the merest suggestion of a metapopulation structure in which groups can compete against other groups," if memetic practices like moral rules have evolved in conjunction with this tendency, and if group-behavior in complex tasks can be as non-transparent as Hutchins suggests, then I again wonder if it is fair to assign such a central role to the intentional agent? As Humphrey remarks in a review of Dennett (1996): if Dennett already has allowed such an important role for the body as an arbiter among otherwise unvalenced alternatives, why does he not extend this to incorporate the evolved, cultural, linguistic body of experience and knowledge? Humphrey clearly sees a deep commonality between the role of the body and culturally transmitted know-how and common sense: "Hence the body of language, just like the body of flesh and blood, can act as a sounding board against which to test the mind's evolving answers to its problems: expressing preferences, arbitrating between possible alternatives, and generally bringing matters to a quick and sensible conclusion" (1997, p 101). Taking the meme-eye perspective he continues: "Language is not just a medium *in which* we think, it actually *does* some of the thinking with us and for us" (op cit.). And it seems as if it does most of this thinking without us knowing it!

In a sense this is not fair to Dennett of course. I am trying to push him further down the cheap modulation, pandemonium, distribution path in hope that this will reveal the rational folk-psychological agent as something of a fiction, but perhaps in all my speculative suggestions I have reached the "plausible upper bound on the number of concurrent, semi-independent control structures that can coexist in a nervous system," and that such a maximally modular and distributed mind as I am envisaging is not a real possibility. Unfortunately I really can not answer this question. All I know is that Dennett has never seriously considered any of the results coming out of evolutionary psychology (including the possibility of group-cognition) and behavioral genetics, or the studies done by Hutchins (and others) on culturally evolved group processes, and that the scenarios I have presented have none of the dreaded connotations of rigidity and determinism that evolutionary explanations have carried in the past. According to this story, in our heads we (mostly) find different learning mechanisms that prey upon subtle initial biases, architectural constraints, timing sequences, somatic markers, and specific environmental cues. In the (rest of the) world we have a causal nightmare of a process involving these lightly biased brains and the coupled, memetic environment; and like Dennett (1995) says, memes and cultural evolution can indeed overcome (exploit, attenuate, etc.) these weak constraints. I am not saying that intentional language always distorts the causal process, or that there can be no rationality at a personal level. What I am saying is that even if language-learning and cultural evolution to a large extent are processes of progressive "explicitation" of knowledge (Dennett 1997) and we engage in some form of endogenous driven representational redescription to the

same end (Karmiloff-Smith 1994, Clark & Thornton 1997, Greco & Cangelosi 1996), many of the explanations we use in our daily lives - explanations that influences our preferences and decisions greatly - are not carving nature anywhere near its joints.

4. Memes and Modules,Explanation and Confabulation

My argument so far has tried to establish that this process of everyday attribution is at serious risk of swinging loose from the actual causal happenings of life once we allow the existence of cognitive content with such diverse, complex and unintuitive developmental trajectories as the ones described above. In section two I outlined some different ways in which intentional ascriptions may be false or unwarranted, and now it is time to scrutinize the dynamics of intentional explanation more closely.

Reflecting on the nature of the wisdom dispersed throughout the body and in our encapsulated brain modules Dennett (1996, p 80) notices that it sometimes can feel as if the "body" has a mind of its own that tries to enact policies in competition with *our* projects (see also Ekman 1992). We are all familiar with these (often embarrassing) situations, and hence we can *plan ahead* to prevent or avoid them (or embrace them on specially sanctioned occasions). Consequently we have also armed ourselves with an arsenal of cultural artifacts invented to tame and augment the wills of the body (makeup to hide blushing, Viagra to ward of impotence, etc.), but these are exactly the cases I am *not* thinking about. The extended influences of the modular/memetic mind I have tried to review above seem to be much more insidious - just because we *do not* know they exist, we can not actively work against them. *We* might think we believe and desire X and Y, and do it for reasons Z and Q, but this might in fact be just intentional confetti on top of the modular, computational machinery. As an illustration of this, take the old (but still notorious) social psychology experiments of Nisbett & Wilson (1977). In one of these experiments that took place in a clothing department-store the subjects were presented with an array of five identical pairs of socks, and were given instructions to choose the one they liked best and then to state their reason for choosing that particular pair. By using a between-group design and randomly moving around the socks Nisbett & Wilson found that a strong bias towards choosing the rightmost sock existed, and that this cause or reason figured nowhere in the verbal reports of the subjects involved (instead people said things like "it had a good texture," "the quality seemed superior"). They did a whole range of similar experiments in which they manipulated environmental variables that they thought would influence the subjects behavior, and then simply asked the persons involved to state the reasons *why* they did what they did (see also Nisbett & Ross 1980, Wilson & Stone 1985, and Ross & Nisbett 1991), and as striking as a series of experiments have ever been, these reasons regularly turned out to have nothing whatsoever to do with the causes Nisbett & Wilson had independently identified. Clearly the subjects were *approximating confabulation* (in the terminology of Dennett), making up a reasonable answer without knowing it themselves. At that time the experiments were almost buried by the ensuing critique (Fiske & Taylor 1991); because they had used a between-group design, it could still be the case that the subjects actually were correct about what went on in their minds, and that the cause Nisbett & Wilson had identified were an artifact of the averaging between groups. Now that we are all eliminative materialists (well...) regarding consciousness, qualia and privileged access, this critique can not bear much weight any longer. Just because people say they believe something for a particular reason does not mean we have to take their word for it. Interestingly enough this is exactly what we do in everyday life though. When Nisbett & Wilson confronted some of the subjects after the sock experiment and asked them if it could not have been the case that they actually chose the sock they did because it was the rightmost one (after all, all the socks were the same) they got only incredulous stares as an answer - who were Nisbett & Wilson to say what *personal* reason they had for choosing the particular way they did. What is even more interesting is that this kind of case fits perfectly with the process Dennett (1996) describes as language users being *taken in* by their own linguistic creations - i.e. *coming to believe the things they say* (for why would they have said them if they were not true?), so from that point on, after Nisbett & Wilson asked their question about the socks (and particularly after they confronted the subjects with their renegade cause), these people had a new, and specific *opinion* lingering in their minds ready to influence behavior on a later occasion (what Hacking (1995) calls the *looping-effect* of human kinds[16]).

Now, this kind of sense-making process would tend to produce "rational" results after all if the conditions identified by Nisbett & Wilson were only laboratory anomalies with no everyday significance (even if we have no privileged access to the causes of our behavior we are bound to come up with a good causal explanation if someone punches us in the face and we start crying). Here I think one of the other main lines of criticism of Nisbett & Wilson is extremely helpful in deciding whether this is the case or not. This critique centered on the fact that Nisbett & Wilson had made many variations of their experiments in

which they had gotten no results. That is, these were variations of the environmental manipulation in which no uniform effect could be seen across the two groups. Thanks to this the actual results could be interpreted as due to freak circumstances which Nisbett & Wilson laboriously had managed to set up (this is the indeed the opinion of Dennett when discussing these experiments in The Intentional Stance). Given my arguments above another more plausible interpretation is available. The great difficulty Nisbett & Wilson had in manipulating potential causes had nothing to do with "personal" reasons nullifying their setup, but could instead simply be seen as a reflection of the true difficulty in divining the obscure causal workings of the modular/memetic mind (as researchers they had almost as little insight into these causal processes as the naive subjects). If human cognition, as I have suggested, owes a lot more to cryptic "innate content," "distributed cognition" and "group phenotypes" (biological and cultural), than Dennett earlier envisaged, then it might actually be the case that Nisbett & Wilson (unknowingly!) made one of the very few valid observations as to how beliefs and desires are formed and maintained in the dynamics of everyday life, and that a modest downplaying of the unity and rationality of the intentional agent is in order.

4.1.What is in the process of intentional ascription?

To really bring the conclusion from the paragraph above to the fore I think it is necessary to once more draw together the components of Dennett's view of intentionality and folk-psychology. Starting with the advanced "behavior reading" capabilities of animals (particularly primates) we need language to get to our kind of (infinite order) intentionality: but it is also clear that this could never be accomplished without the underlying simple mechanisms of behavior reading (Heyes 1998, Whiten 1997). First we seem to have a very basic tendency - visible within the first month or so - of treating self-propelled, animate objects, or biological kinds, as if they had a causal *essence* (Gelman, Coley & Gottfried 1994, Wellman, Hickling & Schult 1998, Hirschfeldt 1996, Atran, to appear). Building on this many researchers now think that the ability of infants to track and follow gaze direction (Baron-Cohen 1994, 1997), especially to look in the direction of gaze of the parents (or other adults) when they utter novel words (Baron-Cohen et al. 1997), to understand and use deictic pointing (Gomez. 1998), to take-turns in conversational situations (Johnson 1997) are also highly modular abilities that do not involve much in the way of "learning" (remember the simple case of the imprinting chicks however). Thus there are elements of killjoy in any account of human intentionality.

This much is needed to get language learning off the ground. In Dennett's version (as far as I understand it) theory of mind is then an effect of the demands of the communicative context, which in essence forces the uses of intentional abstractions (as a kind of approximation of behavioral tendencies). If as I suspect this is supposed to be primarily a *phylogenetic* account, and that we all can agree on the adaptive value of having intentional concepts, then we can suppose that the actual developmental pathway children take when learning intentional concepts involve similar kinds of non-obviously distributed processes as the ones described above. With the help from the scaffolding that the ability to track gaze and pointing and take turns provides it seems like the use of mental state terms actually comes along with language. This is evidenced by the strong correlation between the degree of use of mental state terms by parents and the corresponding use by their children (Montgomery 1997), by the radically different conception of "mind" that preschool children have (Flavell et al. 1997, Flavell, Green & Flavell 1995), and by the fact that explicit training in the use of mental state terms can produce much improved results on classic theory of mind tasks like false-belief (Slaughter & Gopnik 1996, Gopnik & Meltzoff 1997, see also Diaz & Berk 1992, Berk 1994). But there are most probably also less obvious processes that conspire to create a world of intentionality. Hendricks-Jansen (1996) provides an example of this in the mother-infant interaction. Human infants are the only mammalian infants that make pauses in their breast sucking-behavior, and human mothers seems to be biologically primed to respond to these pauses by starting to jiggle the child. They strongly believe this jiggling to encourage the infant to resume sucking, but in independent testing it has been shown that infants resume sucking after a few seconds anyway (in fact, it is only after *cessation* of jiggling that they *can* resume sucking). Hendricks-Jansen suggests that this behavior is adaptive just *because* it encourages the mother to ascribe *more* intentionality to the child than is actually warranted (and here it is easy to imagine that the sucking behavior is governed by an all too simple killjoy mechanism).

Now, on Dennett's account intentional ascription, whether it is targeted towards other people or oneself, has manifest predictive utility. An agent believes and/or desires X if one gets predictive leverage by ascribing that certain belief/desire profile to that particular agent, because then one has captured a *real pattern* in the behavior of that agent. Beliefs and desires are not all there is to folk-psychology though,

there is also all "occurrent" mental content, like the sudden thought about the presence of someone else that Eve claimed to have when talking to ITS (in the example in section two), and all the causal why:s of the Nisbett & Wilson experiment. Of these I have mainly talked about the risk of making false attributions of cause due to the highly distributed and modular nature of the underlying cognitive architecture, and how this in turn might (unconsciously) lead to the creation of new beliefs and desires. So far in my exposition, despite all the talk about distribution and group cognition, it has seemed as if this process is something that takes place solely at an individual level, but that is clearly not the case. In fact many researchers have argued that there is something distributed about the process of intentional explanation itself (Goody 1996, Strum, Forster & Hutchins 1997, see also Carporeal 1996, Wilson & Keil 1998, Gopnik 1993), that the predictive utility comes as it were from it being a self-sustaining norm:

The value of thinking of people in terms of beliefs, desires, and personality, or in terms of social position, need, and the dictates of the gods, is conditional on other people thinking of them in the same terms. The descriptions and the explanations have to mesh. And they have to mesh with the rest of what people believe" (Morton 1996, p 134)

Dennett (1991) uses this view as a bogey in discussing Richard Rorty's alleged irrealism about the predictive patterns seen from the Intentional Stance (in Dennett 1993 he admits that this is not a fair description of Rorty, but I think the quote is telling in and of itself):

Rorty might hold that the predictive "success" we folk-psychology players relish is itself an artifact, a mutual agreement engendered by the egging-on or consensual support we who play this game provide each other. He would grant that the game has no rivals in popularity, due - in the opinion of the players - to the powers it gives them to understand and anticipate the animate world. But he would refuse to endorse this opinion (1991b p 50)

Dennett then compares the Intentional Stance to the Astrological Stance (that deals with patently unreal patterns) and claims that Rorty would have no way of explaining why it is that from one of these stances one can get rich by betting on the patterns, and from the other one could only get rich by selling them. Regardless of what Rorty believes this strikes me as not being true. Professional astrologists do not primarily churn out predictions based on nonsense data, instead they tailor their predictions carefully to their customers preferences and situation, in essence making a informed guess as to what they *could* and *want* to become, before putting forward their prediction. In this they often *create* the very patterns they divine (given that the subjects are true believers and not skeptics that actively try to nullify the predictions). I do not think a comparison to what goes on in the discourse of everyday life is too strained. Even though folk-psychology is a genuine predictive device we do not usually go around actively predicting each other, most of the time we just retrodict and never really get around to testing our beliefs, not the least because doing so would normally intimately *involve* the person under prediction, which of course would ruin the whole procedure. Dennett's (1991b) emphasis on idealized prediction contests (to tell if a pattern is real) obscures the fact that in everyday life prediction, retrodiction and manipulation is one and the same process - a process that we are very seldom even aware of participating in (Shotter 1997)

Further complications arise from the fact the "method" of folk-psychological attribution - the *who* did *what* to *whom* and *why* questions that are so characteristic of gossip (Dunbar 1997) - that might have worked reasonably well to calibrate life under ancestral conditions, do not function nearly as well in our modern society. This is so for two main reasons. Firstly, the structure of everyday interactions has changed considerably. Now we are not only expected to have a range of beliefs and desires about the maximally 50-100 or so people in our hunter-gatherer group (our family, friends, neighbors and close enemies, people that we normally meet face to face in interaction). We are also expected to have beliefs and desires concerning just about any subject imaginable. We constantly have to form opinions about people who we have never met, situations we have never experienced, and things we never even knew existed. Under these conditions it is impossible to even begin to sort out which patterns are real, and which ones that are merely apparent (like the ones of astrology). The standards of evidence that we have inherent from our evolutionary past are seriously maladapted to a world without personal *history* to most of the interactions talking place. The second reason has to do with the effects of cultural evolution upon the vast memosphere of competing beliefs we face everyday through the use of fast and more wide-spread media (books, newspapers, TV, CD-ROMs, internet, etc, including secondary influences from other people exposed to these sources). In the discussion of meme-theory in Dennett (1995) he stresses that most memes after all replicate in concert with us (share fate with us) - i.e. because we deem them to be somehow good for us. This includes all the extended tools of rationality, the "words, books, diagrams, figures, concrete examples, algebra, logic, lisp, and legal systems." Unfortunately this also highlights the great and pressing *need* for these tools. As largely

institutional, distributed tools they enjoy only a marginal existence in the standard kit of folk-psychology, and not very many people can afford to spend the time, energy and money to get a sufficient amount of these tools *into their heads*, or even the pointers, remainders and procedures to access them in the world when needed (Kuhn et al. 1995, Kuhn 1996, but see also Gopnik 1998). On the other hand, as individual folk-psychologists we are awash in a sea of ferociously competing "psychologies" (Holland & Quinn 1987, Hirschfeldt 1995, Morris, Nisbett & Peng 1995); purported truths about our mental lives outstandingly adapted to the task of making more copies of themselves (astrology, psychoanalysis, the practice of "revealing" repressed memories of child abuse, religions, soap operas, etc, see Dawkins 1993, and Hacking 1995 for more examples[17]), but with no particular friendly relationship to the issue of truth (more than in the looping-effect sense of *making* the carrier of the meme into an existence proof of the content of the meme). This, together with the material reviewed above, leads me to conclude that even though folk-psychology might very well be an unparalleled predictive device, it is more than that. In a way that I would be hard pressed to call *rational* it is above all a process nontransparently distributed between many unknowing users, and intimately bound up in the *creation* of our selves.

5. Concluding remarks

If I am on the right track in my discussion there is reason to believe that the coherent, rational, intentional agent that lies at the core of our folk-psychological conception of ourselves is something of a fiction. In my view Dennett has in his extensive writings provided all the tools necessary for this conclusion. What I have done is to try to push him further down the path of seeing cognition as being made up of cheap modulation, pandemonium, and social distribution, and then try to use this position to expose a crack or two in his conception of the unity, coherence and rationality of the folk-psychological agent. It is not wholly clear what moral to draw from this conclusion though. Dennett has always been a conservative philosopher when it comes to results that could be interpreted as having consequences for our personal lives. As cognitive scientists we must give up on phenomenal properties, Cartesian theaters, *me*dia, qualia, raw feelings, and real seemings, otherwise the progress of science will be hampered, but he believes that as individuals we can still can have our free will (1984), our intentional folk-psychology (1987), our selves (1991a) our heterophenomenological worlds (1991), and our moral rights and obligations (1995). I on the other hand tend to think that the really important conclusion to draw from the work of Dennett (and my extrapolations) lies squarely in the moral, personal and political sphere. Like Rorty (1991) I think it is about time that we changed the revered image of ourselves as private, privileged, ineffable subjects, and instead found some other more profitable ways of speaking, some ways of allowing us to relinquish some of the authority we claim over our daily lives. Apart from reviving the research program of Nisbett & Wilson, and their special brand of rather invasive and manipulative experiments, I have very few specific suggestions as a cognitive scientist about how we are to proceed and really put the speculations and conjectures of this article to the test, because they already fall within what might be both the most hotly debated and the most challenging topic in contemporary cognitive science - that of scrupulously dividing up the causal influences of behavior between brains, bodies, artifacts and the wider social environment.

References

Akins, A. K. (1996). Lost the Plot? Reconstructing Dennett's Multiple Drafts Theory of Consciousness. *Mind & Language*. 11. 1-43.

Akins, A., K., & Winger, S. (1996). Ships in the Night: Churchland and Ramachandran on Dennett's Theory of Consciousness. In K. Akins (Ed.) *Perception*. New York: Oxford Univ. Press. p 173-198.

Arbib, M. A., & Rizzolati, G. (1997). Neural Expectations: A Possible Evolutionary Path From Manual Skills to Language. *Communication & Cognition*. 29:3/4. 393-424.

Atran, S. (to appear). Folk Biology and the Anthropology of Science: Cognitive Universals and Cultural Particulars. *Behavioral and Brain Sciences*

Ballard, D. H., et al. (1997). Deictic Codes for the Embodiment of Cognition. *Behavioral and Brain Sciences.*20. 723-767.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience*. 7:1. 66-80.

Barkow, J., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary Psychology and the generation of Culture.* New York: Oxford University Press.

Baron-Cohen, S. (1994). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *Current Psychology of Cognition.* 13:5. 513-552.

Baron-Cohen, S. (1995). *Mindblindness: an Essay on Autism and Theory of Mind*. MIT-Press.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do Children with Autism Use the Speaker's Direction of Gaze Strategy to Crack the Code of Language? *Child Development*. 68:1. 48-57.

Barton, R. A. (1997). Neural constructionism: How mammals make modules. *Behavioral and Brain Sciences*. 20:4. 556-557.

Baumeister, R., F., & Leary, M., R. (1995). The Need to Belong. *Psychological Bulletin*, 117, 497-529.

Berk, L. (1994). Why children talk to themselves. *Scientific American*. 271-5. 78-83.

Block, N. (1994). What is Dennett's Theory a Theory of? *Philosophical Topics*. 22, 23-41.

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences*. 2:2. 67-73.

Boehm, C. (1997a). Egalitarian behavior and the evolution of political intelligence. In A. Whiten & R Byrne (Eds). *Machiavellian Intelligence II*. Cambridge University Press.

Boehm, C. (1997b). Impact of the Human Egalitarian Syndrome on Darwinian Selection Mechanics. *American Naturalist*. Vol 150. July 1997.

Brooks, R., et. al. (1998). Alternative Essences of Intelligence. *MIT Artificial Intelligence Lab Technical Report*

Browne, D. (1997). Cognitive Versatility. *Minds and Machines*. 6.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. Oxford University Press.

Caporeal, L. R. (1996). Coordinating Bodies, Minds and Groups: Evolution and Human Social Cognition. *Journal of Social & Evolutionary Systems*. Vol. 19. Iss. 3. pp 261-276.

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

Clark, A. (1996b). Dealing in Futures: Folk Psychology and the Role of Representations in Cognitive Science. In R. N. Cauley. (ed). *The Churchlands and their critics*. Oxford: Basil Blackwell.

Clark, A. (1998). Magic Words: How language augments human computation. In. P. Carruthers & J. Boucher. (Eds.). *Language and Thought: Interdisciplinary Themes*. Cambridge University Press.

Clark, A. (1998b). Twisted Tales: Causal Complexity and Cognitive Scientific Explanation. *Minds & Machines*. 8. 79-99.

Clark, A. (to appear). Review of Elman et. al. Rethinking Innateness. Mind and Language.

Clark, A., & Chalmers, D. (1995). The Extended Mind. Philosophy-Neuroscience-Psychology Research Report. Washington University, St. Louis.

Clark, A., & Grush, R. (1997). Toward a Cognitive Robotics. Philosophy-Neuroscience-Psychology Research Report. Washington University, St. Louis.

Clark, A., & Karmiloff-Smith, A. (1993). The Cognizer's Innards. *Mind & Language.* 8, 487-515.

Clark, A., & Thornton, C. (1997). Trading Spaces: Computation, Representation and the Limits of Uninformed Learning. *Behavioral and Brain Sciences*. 20. 57-90.

Cosmides, L., & Tooby, J. (1994a) Beyond Intuition and Instinct Blindness: Toward an evolutionary rigorous cognitive Science. *Cognition*, 50, 41-77.

Cosmides, L., & Tooby, J. (1994b). Origins of domain specificity: The evolution of functional organization. In. L. A. Hirschfield, & S. A. Gelman (Eds.) *Mapping the mind: Domain specificity in cognition and culture.* Cambridge, NY: Cambridge University Press.

Cosmides, L., & Tooby, J. (1997). Dissecting the computational architecture of social inference mechanisms. In. M. Daly. (Ed.). *Characterizing human psychological adaptations.* Ciba Foundation Symposium 208. Chichester: Whiley

Crawford, C., & Krebs, D. (Eds). (1997). *Handbook of Evolutionary Psychology: Ideas, Issues and Applications*. New Jersey: Erlbaum Associates.

Cronk, L. (1994a). Group selection's new clothes. *Behavioral and Brain Sciences*, 17, 615.

Cunningham, M, R., et. al. (1995). "Their Ideas of Beauty Are on the Whole, the Same as Ours": Consistency and Variability in the Cross-Cultural Perception of Female Physical Attractiveness. *Journal of Personality and Social Psychology*. 68. 2. 261-279.

Dahlbom, B. (1993) Mind is Artificial. In B. Dahlbom (Ed.), *Dennett and his Critics: Demystifying Mind.* Cambridge, Mass: Basil Blackwell.

Daly, M. (1996). Evolutionary adaptationism: another biological approach to criminal and antisocial behavior. *The Ciba Foundation Symposium.*

Daly, M. (1997). (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley.

Damasio, A. (1994). *Descartes' Error.* Grosset Putnam.

Dawkins, R. (1982). *The extended phenotype: The gene as the unit of selection.* Oxford: Oxford University Press.

Dawkins, R. (1993). Viruses of The Mind. In B. Dahlbom (Ed.), *Dennett and his Critics: Demystifying Mind*. Cambridge, Mass: Basil Blackwell.

Deacon, T. (1997). *The Symbolic Species: The coevolution of language and brain.* Cambridge, Mass: MIT-Press.

Dennett, D. C. (1978). *Brainstorms*. Bradford Books.

Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Oxford: Oxford University Press.

Dennett, D. C. (1987). *The Intentional Stance*.. Cambridge, Mass: MIT Press.

Dennett, D. C. (1988). Précis of The Intentional Stance. *Behavioral and Brain Sciences, 11*, 495-546.

Dennett, D. C. (1991a). *Consciousness Explained*. Boston: Little, Brown & Company.

Dennett, D. C. (1991b). Real Patterns. *Journal of Philosophy, 89*, 27-51.

Dennett, D. C. (1991c). Two Contrasts: Folk Craft versus Folk Science and Belief versus Opinion. In. J. Greenwood (ed.). *The future of Folk Psychology: Intentionality and Cognitive Science*. Cambridge University Press.

Dennett, D. C. (1993). Back from the Drawing Board. In B. Dahlbom (Ed.). *Dennett and his Critics: Demystifying nd*. Oxford: Blackwell.

Dennett, D. C. (1994). Get Real. *Philosophical Topics*. 22, 505-568.

Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* New York, NY: Simon & Schuster.

Dennett, D. C. (1996). *Kinds of Minds: Toward an Understanding of Consciousness.* Basic Books.

Dennett, D. C. (1997). How to do other things with words. *Philosophy: the journal of the Royal institute of philosophy*. Iss. 42. p219.

Devlin, Daniels, & Roeder. (1997). The heritability of IQ. *Nature*. 338. 31 July.

Diaz, R. & Berk, L. (Eds.). (1992). *Private Speech: From Social Action to Self-Regulation*. Erlbaum.

Dunbar, R. I. M. (1997). Groups, Gossip, and The Evolution of Language. In Schmitt et al. (Eds.) *New Aspects of man Ethology*. Plenum Press.

Ekman, P. (1992). *Telling Lies. Clues to deceit in the marketplace, politics, and marriage*, *2. ed*. New York. Norton.

Elman, J., et al. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, Mass: MIT Press.

Estes, D., & Bartsch, K. (1997). Constraining the brain: The role of developmental psychology in developmental cognitive neuroscience. *Behavioral and Brain Sciences.* 20:4. 562-563.

Fiske, S. T., & Taylor S. E. (1991). *Social Cognition*. New York, NY: McGraw-Hill.

Flavell, J. H., et. al. (1997). The Development of Children's Knowledge about Inner Speech. *Child Development*. 68:1. 39-47.

Flavell, J. H., Green. F. L., & Flavell, E, R. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development*. 60: 1.

Gallistel, R. (1990). *The Organization of Learning*. MIT-Press.

Gallistel, R. (1994). Interview. *Journal of cognitive neuroscience*. 6:2. 174-179.

Gangestad, S. W. (1997). Evolutionary psychology and genetic variation: non-adaptive, fitness-related, and adaptive. In. M. Daly. (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley.

Gangestad, S. W. (1997). Evolutionary psychology and genetic variation: non-adaptive, fitness-related and adaptive. In. M. Daly (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley.

Gangestad, S. W., & Thornhill, R. (1997). Human sexual selection and developmental stability. In. J. A. Simpson & D. T. Kendrick. (Eds.). *Evolutionary Social Psychology*. Erlbaum.

Gelman, S., Coley, J., & Gottfried, G. (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In. L. A. Hirschfeldt, & S. A. Gelman (Eds.) *Mapping the mind: Domain specificity in cognition and culture.* Cambridge, NY: Cambridge University Press.

Gergely, G., et. al. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165-193.

Gigerenzer, G. (1997). The modularity of social intelligence. In. A. Whiten & R Byrne (Eds). *Machiavellian telligence II*. Cambridge University Press.

Gómez, J-C. (1998). Some thoughts about the evolution of LADS, with special reference to TOM and SAM. In. P. Carruthers & J. Boucher. (Eds.). *Language and Thought: Interdisciplinary Themes.* Cambridge University Press.

Goody, E. N. (1997). Social intelligence and language: Another Rubicon? In A. Whiten & R Byrne (Eds). *Machiavellian Intelligence II*. Cambridge University Press.

Gopnik, A. (1993). How we know our Minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1-14.

Gopnik, A. (1998). Explanation as Orgasm. *Minds & Machines*. 8. 101-118.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, Thoughts and Theories*. Cambridge, Mass: MIT-Press.

Gray, R., D (1992). Death of the gene: Developmental systems strikes back. In. P, Griffiths (Ed.) *Trees of life: Essays in the philosophy of biology*. Kluwer.

Greco, A., Cangelosi, A. (1996). *A representational redescription method using competitive learning.* Technical report. University of Genoa.

Griffiths, P., E., & Gray, R., D. (1994). Developmental systems and evolutionary explanation. *Journal of Philosophy*. XCI (6).

Grush, R. (1997). The Architecture of Representation. *Philosophical Psychology*. 10:1. 5-23.

Grush, R. (to appear). Perception, Imagery, and the Sensimotor Loop. In Esken & Heckman (Eds). *A Consciousness Reader*. Schoeningh Verlag.

Gärdenfors, P. (1996). Cued and detached representations in animal cognition. *Behavioral Processes*, 35, 263-273.

Hacking, I. (1995). The Looping Effect of Human Kinds. In. D. Sperber, D Premack & A. J. Premack. (Eds.) *Causal Cognition: An Interdisciplinary Debate.* Claredon Press. Oxford.

Hauser, M. (1996). *The Evolution of Communication.* Cambridge, Mass: MIT-Press.

Hendriks-Jansen., H. (1996). *Catching ourselves in the act*. MIT-Press.

Heyes, C. M. (1998). Theory of Mind in Nonhuman Primates. *Behavioral and Brain Sciences.* 21. 101-148.

Hirschfeldt, L., A. (1995). Anthropology, psychology and the meanings of social causality. In. D. Sperber, D Premack mp; A. J. Premack. (Eds.) *Causal Cognition: An Interdisciplinary Debate.* Claredon Press. Oxford.

Holcomb III, H, R. (1996). Just So Stories and Inference to the Best Explanation in Evolutionary Psychology. *Minds d Machines*. 6. 525-540.

Holland, D., & Quinn, N. (1987). *Cultural models in language and thought.* Cambridge, NY: Cambridge University Press.

Humphrey, N. (1997). Review of Dennett, *Kinds of Minds*. Journal of Philosophy.

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

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. Behavioral and Brain Sciences. 17. 187-245.

Johnson, M. K., & Raye, C, L. (1998). False memories and confabulation. *Trends in Cognitive Science*. 2:4. 137- 145.

Johnson, M., H. (1997). *Developmental Cognitive Neuroscience.* Cambridge, Mass: Blackwell.

Jones, D. (1995). Sexual Selection, physical Attractiveness, and Facial Neotony. *Current Anthropology*. 36. 5. 723-745.

Karmiloff-Smith, A. (1994). Précis of Beyond modularity: A developmental perspective on cognitive science. *Behavioral and Brain Sciences*. 17. 693-745.

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

Kennedy, J. (1998). Thinking is Social: Experiments with the Adaptive Culture Model. *Journal of Conflict Resolution*. 42:1.

Kirsch, D., & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science.* 18. 513-549.

Knauft, B., M. (1991) Violence and Sociality in Human Evolution. *Current Anthropology*, 32, 391-419.

Kuhn, D. (1996). Is good thinking scientific thinking? In. D. R., Olson & N. Torrance. (Eds.) *Modes of Thought: explorations in culture and cognition*. New York: Cambridge University Press.

Kuhn, D., et. al. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*. 60:4.

McCrone, J. (1994). Don't Forget Your Memory Aide. *New Scientist*. feb, 32-36.

Mealey, L. (1995). The Sociobiology of Sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, 18, 523-599.

Miller, G. F. (1997a). A Review of Sexual Selection and Human Evolution: How mate Choice shaped Human Nature. In C. Crawford & D. Krebs. (Eds.). *Handbook of Evolutionary Psychology: Ideas, Issues and Applications*. New Jersey: Erlbaum Associates.

Miller, G. F. (1997b). Protean Primates: The Evolution of Adaptive Unpredictability in Competition and Courtship. In A. Whiten & R Byrne (Eds). *Machiavellian Intelligence II*. Cambridge University Press.

Miller, G. F. (1997c). Mate choice: from sexual cues to cognitive adaptations. In. M. Daly (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley.

Miller, G. F., & Todd, P. M. (1998). Mate choice turns cognitive. *Trends in Cognitive Science*. 2:5. 190-199.

Millikan, R. G (1998). A Common Structure for Concepts of Individuals, Stuffs, and Basic Kin: More Mama, More Milk, and More Mouse. *Behavioral and Brain Sciences.* 21. 55-100.

Montgomery, D. E. (1997). Wittgenstein's Private Language Argument and Children's Understanding of the Mind. *Developmental Review*. 17. 291-230.

Morris, M., W., Nisbett, R., & Peng, K. (1995) Causal attribution across domains and cultures. In. D. Sperber, D Premack & A. J. Premack. (Eds.) *Causal Cognition: An Interdisciplinary Debate*. Claredon Press. Oxford.

Morton, A. (1996). Folk Psychology is not a Predictive Device. *Mind*. 105:417. 119-135.

Mundale, J., & Bechtel, W. (1996). Integrating Neuroscience, Psychology, and Evolutionary Biology through a Teleological Conception of Function. *Minds & Machines*. 6. 481-505.

Möller, A, P. (1997). Evolutionary conflicts and adapted psychologies. In. M. Daly (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley.

Möller, A. P. (1997). Evolutionary conflicts and adapted psychologies. In. M. Daly. (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley

Möller, A., & Thornhill, R. (1998). Bilateral symmetry and sexual selection: a meta-analysis. *American Naturalist*. 151. 174-192.

Nisbett, R., & Ross, L. (1980). *Human Inference: Strategies and shortcomings of social judgement*. Engelwood Cliffs, NJ: Prentice-Hall.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Olson, D. R. (1994). *The world on paper: the conceptual and cognitive implications of writing and reading.* Cambridge, New York: Cambridge University Press.

Philips, W. A., & Singer, W. (1997). In search of common foundations for cortical computations. *Behavioral and Brain Sciences*. 20. 657-722.

Plomin, R, et. al. (1997). *Behavioral Genetics*. 3rd. ed. Freeman Press.

Povinelli, D. J., Povinelli T. J., & Giambrone, S. (1998). Review of Reaching into Thought: The Minds of the Great Apes. *Trends in Cognitive Science*. 2:4. 158-159.

Premack, D., & Premack, A. J. (1997). Infants attribute value to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience.* Vol. 9. Iss. 6. pp 848-859.

Quartz, S. R., & Sejnowski, T. J. (1997). The Neural Basis of Cognitive Development: A Constructionist Manifesto. *Behavioral and Brain Sciences.* 20. 537-596.

Ramachandran, V. S., et. al. (1996). Illusions about Body Image: What They Reveal About Human Nature. In R, Llinas & P. S. Churchland (Eds.). *The Mind-Brain Continuum*. MIT Press: Cambridge, Mass

Richerson, P. J., & Boyd, R. (1997). The Evolution of Human Ultra-Sociability. In I. Eibl-Eibisfeldt & F. Salter (Eds.). *Ideology, Warfare, and Indoctrinability.*

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Rorty, R. (1991). *Objectivity, Relativism and Truth.* Cambridge: Cambridge University Press.

Ross, L., & Nisbett, R. (1991) *The person and the situation - Perspectives of social psychology*. Temple University Press.

Rudder-Baker, L. (1994). Content Meets Consciousness. *Philosophical Topics*. 22:1-2. 1-23.

Schacter, D. L. (Ed.). (1995). *Memory distortion: how minds, brains, and societies reconstruct the past*. Cambridge, Mass: Harvard University Press.

Seeley, T. D. (1997). Honey Bee Colonies are Group-Level Adaptive Units. *American Naturalist*. Vol 150. July 1997.

Shotter, J. (1997). Dialogical Realities: The Ordinary, the Everyday, and Other Strange New Worlds. *Journal for the Theory of Social Behavior*. 27:2/3. 345-357.

Slaughter, V., & Gopnik, A. (1996). Conceptual Coherence in the Child's Theory of Mind: Training Children to Understand Belief. *Child Development*. 67. 2967-2988.

Spelke, E. S., Philips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In. D. Sperber, D Premack & A. J. Premack. (Eds.) *Causal Cognition: An Interdisciplinary Debate*. Claredon Press. Oxford.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In. L. A. Hirschfeldt, & S. A. Gelman (Eds.) *Mapping the mind: Domain specificity in cognition and culture.* Cambridge, NY: Cambridge University Press.

Sperber, D. (1997). *Explaining Culture: A Naturalistic Approach*. Cambridge Mass: Blackwell.

Stone, T., & Young, A, W. (1997). Delusions and Brain Injury: The Philosophy and Psychology of Belief. *Mind & Language*. 12:3/4. 327-364.

Strum, S., Forster, D., & Hutchins, E. (1997). Why Machiavellian intelligence may not be Machiavellian. In. A. Whiten & R Byrne (Eds). *Machiavellian Intelligence II*. Cambridge University Press.

Thompson, R., & Oden, D. (1996). A profound disparity re-visited: Perception and judgment of abstract identity relations by chimpanzees, human infants and monkeys. *Behavioural Processes*. 35. 149-61.

Thornhill, R. (1997). The concept of an evolved adaptation. In. M. Daly (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley

Tooby, J., & Cosmides, L. (1990) On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*. 58, 17-67.

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

Tooby, J., & Cosmides, L. (1992). The psychological Foundations of Culture. In J. Barkow., L. Cosmides., & J. Tooby (Eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Oxford: Oxford University Press.

Wellman, H. M., & Gelman, S. A. (1998) Knowledge Acquisition in Foundational Domains. In. D. Kuhn & R. Siegler (Eds.). *Cognition, Perception and Language.* New York: Wiley.

Wellman, H., Hickling, A., & Schult, C. (1998). Young Children's Explanations: Psychological, Physical, and Biological Reasoning. In H. Wellman & K. Inagaki (Eds.). *Children's Theories.* San-Francisco: Jossey-Bates.

Whiten, A. (1997). The Machiavellian mindreader. In A. Whiten & R Byrne (Eds). *Machiavellian Intelligence II*. mbridge University Press.

Whitmeyer, J. M. (1998). On the relationship Between Memes And Genes: a Critique of Dennett. *Biology and Philosophy*. 13. 187-204.

Wilson, D. S. (1994). Adaptive genetic variation and human evolutionary psychology. *Ethology and Sociobiology*, 15, 219-236.

Wilson, D. S. (1997). Altruism and Organism: Disentangling the Themes of Multilevel Selection Theory. *American Naturalist*. Vol 150. July 1997.

Wilson, D. S., Near, D., & Miller, R. R. (1996). Machiavellianism: A synthesis of the Evolutionary and Psychological Literatures. *Psychological Bulletin*, 119, 285-299.

Wilson, D., S., & Sober, E. (1994). Reintroducing Group Selection to the Human Behavioral Sciences. *Behavioral and Brain Sciences*. 17, 585-654.

Wilson, M., Daly, M. (1997). Relationship-specific social psychological adaptations. In. M. Daly (Ed.). *Characterizing human psychological adaptations*. Ciba Foundation Symposium 208. Chichester: Whiley.

Wilson, R, A., & Keil, F. (1998). The Shadows and shallows of Explanation. *Minds & Machines*. 8. 137-159.
Wilson, T., & Stone, J. (1985). Limitations of self-knowledge: More on telling more than we can know. *Journal of personality and social psychology*,
Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Science*. 1:6.
Wolpert, D. M., Miall, C., & Kawato, M. (1998). Internal models in the cerebellum. Trends in Cognitive Sciences. 9. 338-347.
Zhang, J. (1997). The Nature of External Representations in Problem Solving. *Cognitive Science.* 21:2.
Zhang, J., & Norman, D. (1994). Representations in distributed cognitive tasks. *Cognitive Science*: 18 -122.

Notes

[1] Dennett was my primary gateway into the subject of cognitive science. As an undergraduate I read and appreciated *Consciousness Explained* (1991) but I already belonged to the minority of people that were already convinced that privileged first-person access and qualia was a hopeless position (not so much due to careful, technical argument perhaps, having instead been swept off my feet by the rhetorical glamour and eloquence of Rorty's *Philosophy and The Mirror of Nature* (1979)). Since then I have been somewhat annoyed that the debate after Consciousness Explained never came round to addressing the issues that fascinated me the most - the specific story of the evolution of our cognitive capacities, the pandemonium model of word-production, and the self as a center of narrative gravity.

[2] This is also the approach that Dennett (1994, 1997b) and Brooks et. al. (1998) has adopted in their attempt to build a humanoid, conscious robot at MIT (Cog).

[3] For more on this see Hauser (1996).

[4] In the terminology of Clark & Karmiloff-Smith (1993) and Karmiloff-Smith (1994) one could say that the amount of *representational redescription* taking place in these animal minds is minimal.

[5] See Sperber (1995, 1997) for an alternative account of cultural evolution along the same lines.

[6] For a detailed and revealing discussion of these aspects, see Akins (1996), Akins & Winger (1996).

[7] The only other possible (and dreadful) interpretation is that the thought rather occurred in her immaterial mind (thus evading any imaginable probe or scan by the neurocryptography unit).

[8] Here is also one of those rare cases where the emptiness of the competing psychoanalytic explanation is clear to see for all involved. The psychoanalytic account of prosopagnosia is of course that the person is unconsciously *denying* the faces he/she cannot recognize (and that this is so for some grueling reason from the past). Since no normal intentional story can be told for the behavior of the prosopagnosic it is easy to see that the only thing the psychoanalytic account does is to invoke a subagent with competing motives and considerable power (a *huge* homunculi), and then move the whole apparatus of normal interpersonal, intentional conflicts into the head. I think this is actually the main reason for the popularity of psychoanalysis: everyone is already so deeply familiar with the methods of explanation involved that it does not take much to accept it - it is in fact only traditional folk-psychology that has been pushed inside the mind (spiced with some dark, sexual motives among the agents involved).

[9] Following Dennett (1994) in his useful characterization of "knobs" that can be turned in the hypothetical intuition pumps of philosophers (such as Searle's Chinese Room), the following can be read as an experiment that, hopefully without doing violence to the basic assumptions of Dennett, cranks up all the knobs to maximum. Like a little kid I want to see what happens if we allow more innate content, more distribution of cognition, more confabulation and retrospective reconstruction in intentional explanations, and more uncontrolled cultural evolution.

[10] It is hard to say how much of this that is due to the Barkow et al. (1992) volume though.

[11] More hypothesized types of evolved, adaptive content are to be found in the recent volume edited by Crawford and Krebs (1997), but see also Cosmides & Tooby (1994), Gallistel (1990, 1997), and Daly (1997). A quick look at the abstracts from the three latest annual meetings of the *Human Behavior and Evolution Society* that are posted online reveals an even more breathtaking range of speculation as to what kind of "Darwinian algorithms" there might be.

[12] For useful overviews of this debate, see Clark (1996a, 1998), and Elman et al. (1996)

[13] Responsivity to local environmental contingencies is a source of cultural *variability* that has nothing to do with transmission. Cosmides & Tooby (1992) illustrate the concept of *evoked culture* by way of a hypothetical example of compact disc jukeboxes dispersed around the globe. These juke boxes have thousands of songs in their repertoire and are all identically equipped with a clock, a device that measures latitude and longitude, and a mechanism that selects what song to play on the basis of location, time and

date. What emerges is the same pattern of in-group similarities and between-group differences as is observed in humans. Jukeboxes in Rio play the same song, which is different from the songs played in Stockholm, etc. Each jukebox's behavior would change over time, and if moved to another location would seem to adopt local "fashions" etc., but the whole process is conspicuously devoid of pure transmission.

[14] "The Hutterites are a fundamentalist religious sect that originated in Europe in the sixteenth century and migrated to North America in the nineteenth century to avoid conscription. The Hutterites regard themselves as a human equivalent of a bee colony. They practice community of goods (no private ownership) and also cultivate a psychological attitude of extreme selflessness /.../ Nepotism and reciprocity, the two principles that most evolutionists use to explain prosocial behavior in humans, are scorned by the Hutterites as immoral. Giving must be without regard to relatedness and without any expectation of return" (Wilson & Sober 1994, p 602).

[15] And here is a dissenting voice: "although it might be true that selfless behavior may be encouraged by a social organization that plays the same role as the genetic rules of meiosis there is precious little evidence that such social organizations typified the human environment of evolutionary adaptness, and certainly they do not typify recently documented band and tribal societies" (Cronk 1994, p 616)

[16] Similar conclusions about the "creative" role of attribution processes (but with a slightly different emphasis) can be found in the literature about memory distortion over time (see Schacter 1995, McCrone 1994, Johnson & Raye 1998).

[17] Soap operas are an interesting case in that they derive their popularity partly from introducing exactly those features of folk-psychology that have been lost in modern society, namely a limited number of interacting people with an interconnected history.

Behaviourism as a Standpoint in Linguistics

Ullin T. Place
Thirsk, North Yorkshire

Introduction

The thesis of this paper is that behaviourism is the only adequate scientific foundation for the disciplines of psychology, linguistics and linguistic philosophy. Behaviourism in psychology is presented as a convergence of *six* principles: (1) behaviour as the subject matter of psychology, (2) the objectivity principle, (3) the rejection of mentalistic explanation, (4) the three-term contingency, (5) the distinction between discriminative stimuli and establishing conditions, and (6) learning theory.

Behaviourism in linguistics and linguistic philosophy is seen as resting on *ten* principles: (1) language as communication in the service of technology, (2) language and thought, (3) the sentence as the functional unit of linguistic communication, (4) novel sentence-construction, (5) novel sentences and the representation of unfamiliar contingencies, (6) sentence-construction and the win-shift/fail-stay contingency, (7) the picture theory of the meaning of sentences, (8) the associative learning of word and phrase meaning, (9) lexical words, syntactic words and Bickerton's "proto-language", (10) mutations and the facilitation of language learning.

Behaviourism

It is my view that behaviourism is the only scientifically acceptable foundation, not only for the science of psychology, but also for the science of language or linguistics, as it is usually called. Moreover, since I hold that philosophy, in so far as it is capable of contributing to the body of human knowledge, is a branch of the science of language, it follows that I also think that that discipline too requires a behaviourist foundation.

What Behaviourism is not

That said, the behaviourism which is needed as the foundation for psychology, linguistics and philosophy is something very different from the popular stereotype. It is no part of the behaviourism to which I subscribe that it denies either the existence of consciousness in the sense of private experience or the possibility of studying such experiences scientifically. It is no part of my behaviourism to reject explanations of the behaviour of organisms in terms of what is going on in the brain and nervous system. Nor is it committed to explaining that behaviour in terms of mechanical reflexes or stimulus-response connections.

Behaviourism in Psychology

Behaviourism as an approach to psychology is committed to the following principles:

B1. *Psychology as the Science of Behaviour*.
It was accepted by every school of psychology until the advent of the "cognitive revolution" that an empirical science must be defined by reference to the *phenomena* it observes and studies, rather than by reference to the *theoretical constructs* it currently uses to explain those phenomena. It is by applying this principle that the behaviourist concludes that the proper definition of psychology is as the science of the behaviour of living organisms, rather than as the science of the mind.

B2. *The Objectivity Principle*.
The *Objectivity Principle* holds that the observation statements which provide the foundation of empirical knowledge are and can only be observations of permanent states of affairs in the public domain whose correct description will be agreed by any competent observer who is also a competent speaker and interpreter of the language and technical code in current use. It is a consequence of this principle that conclusions about the nature of events, whether in the public or in the private domain, must be based on objective records whose relation to the events they record must be understood before they can be relied on. This means that before we can make serious evidential use of an objective record of a subject's "introspective report", we need to understand both the nature of what is being reported and the semantics of the process whereby the report is generated.

B3. *The Rejection of Mentalistic Explanation*.
When explaining the behaviour of living organisms for scientific purposes behaviourists have traditionally objected to the use of what have been called "*mentalistic explanations*". Exactly what kinds of explanation qualify as "mentalistic" and why they are objectionable has never been made entirely clear. I take it that a mentalistic explanation is one which invokes the kinds of process, instantaneous event and ongoing dispositional state which are referred to in the "common sense" or "folk psychological" explanations of human and animal behaviour which we encounter in ordinary non-technical discourse. If so, my view is that such explanations are scientifically objectionable *only* in so far as they presuppose that the behaving organism is linguistically competent and rely for an explanation of why the agent acted as she did on a quotation of what she might be supposed to have said to herself when deciding what to do. It should be obvious that such explanations are scientifically unacceptable when the behaviour to be explained is that of a pre-linguistic organism (animal) or in the case of linguistic behaviour itself. But the principle of the unity of science requires a theory that will apply to all forms of human and animal behaviour. Consequently, a form of explanation which is restricted in its scope to those aspects of human behaviour which are indeed controlled by linguistic formulations of the contingencies involved is unacceptable for the purposes of scientific theory; though equally unacceptable is a theory which cannot accommodate the phenomenon whereby a great deal of human behaviour is controlled in this way.

B4. *The Three-Term Contingency*.
At the molar level of analysis all instrumental/operant behaviour including verbal behaviour, both that of the speaker and that of listener, is acquired, maintained and abandoned in accordance with the principle of the *three term contingency* (Skinner 1969), consisting of

1. a set of Antecedent conditions
2. the Behaviour emitted or omitted under those conditions, and
3. the Consequences of so behaving

B5. *Discriminative Stimuli and Establishing Conditions.*
A distinction needs to be drawn within the *antecedents* of behaviour between *discriminative stimuli* whose effect is to alert the organism to the impending presence or availability of a particular behaviour-consequence relation and the *establishing conditions* (Michael 1982), such as food-deprivation or the lack of the appropriate utensils required in order to eat the food set before one in a socially acceptable manner, which determine the valence of both the anticipated and the actual consequences of emitting or omitting the behaviour, i.e. whether the effect of those consequences is to strengthen (*reinforce*) or weaken (*disinforce* - Harzem & Miles 1978) the organism's disposition to emit similar behaviour on similar occasions in the future. It should be apparent that this distinction corresponds to that traditionally drawn between "cognition" and "motivation".

B6. *Learning Theory*.
While no one could seriously deny that there are aspects of both animal and human behaviour that are innate, and other aspects that are learned, behaviourists have traditionally attached greater importance to the latter than to the former, and would insist with many contemporary neuroscientists that there are no "hardwired" connections in the brain, i.e., nothing that is not susceptible to modification by subsequent learning. In particular they have always insisted that, although its acquisition has undoubtedly been facilitated by genetic mutations, linguistic competence, both that of the listener and that of the speaker, is acquired in accordance with the same principles as are observed in experimental studies of animal learning. Contemporary learning theory, however, has moved on in a number of respects from the position adopted by most behaviourists in the 1950s and early 1960s before behaviourism was swamped by the cognitive revolution. One principle which at that time had only begun to impress itself on the scientific community was that expounded in Ferster & Skinner's (1957) book *Schedules of Reinforcement*, namely that instrumental/operant reinforcement is as much a matter of maintaining ongoing behaviour as of acquiring new patterns. Another principle which had been around for a long time, but which has only recently begun to make its mark is that demonstrated by Miller & Konorski (1928; Konorski 1948) when they showed that underlying every case of instrumental/operant learning there is a classical conditioning to the kinaesthetic feedback from the instrumental/operant response as it develops of the autonomic response elicited by the

instrumental/operant reinforcer in the case of a positively reinforced response and by the aversive stimulus in the case of a negatively reinforced response. More recently, this observation has been given added significance by the work of the associative learning theorists (Rescorla & Wagner 1972) which suggests, if it does not actually demonstrate, that what the organism learns in a classical conditioning situation is to "expect"(1) the US, given the CS, that such expectations develop in accordance with the Principle of Association by Contiguity (Hume 1739), whenever a stimulus of one type is regularly followed by a stimulus of another type, and that the function of the autonomic UR in the classical conditioning situation is simply to make this expectation visible in the form of an autonomic CR. Combining this view of classical conditioning with the Miller-Konorski evidence on the one hand and the evidence (Adams & Dickinson 1981) of the effect of reinforcer-devaluation on an instrumental/operant response on the other shows that in instrumental/operant learning the organism learns to expect certain consequences, given a particular combination of discriminative stimulus and the feedback of an instrumental/operant response. It also suggests that the way the organism behaves in response to that expectation will depend on the value it currently attaches to those consequences (Rescorla 1991). From this it would seem that the learning principle that applies depends on the level of analysis under discussion. Thus at the neuro-synaptic level some version of the Hebb (1949) Principle would seem to apply. I find the version proposed by Montague, Gally & Edelman (1991) the most convincing. At the mental process or, to use Edelman's (1987) term, the "neuronal group" level the Association by Contiguity or Stimulus-Stimulus Expectancy Principle applies; while at the molar behavioural level it is the Law of Effect (Thorndike 1911).

Behaviourism in Linguistics

The behaviourist approach to the science and philosophy of language, the position Place (1996) calls "*linguistic behaviourism*", is committed to the following principles:

L1. *Language as Communication in the Service of Technology.*
Language is a form of behaviour which has evolved in the first instance by virtue of allowing the development of the kind of *social control* and *information transmission* within the social group which are needed for the implementation of the *technology* required to adapt to the environment in which the group finds itself.

L2. *Language and Thought.*
Language has an equally important but ontologically derivative function of allowing its possessors to *solve problems*, both technological and personal, through *symbolic representations* of the consequences of doing one thing rather than another. It seems tolerably certain that the ability to solve problems by means of *iconic representations* (mental images) of the past consequences of emitting or omitting different kinds of behaviour in circumstances resembling those currently prevailing is part of our common mammalian heritage. But, not only are such iconic representations incommunicable from one organism to another, in the absence of language they are necessarily restricted to behaviour-consequence relations which the organism has experienced in its own case.

L3. *The Sentence as the Functional Unit in Linguistic Communication.*
The *sentence* is the functional unit of language, the unit which must be completed if an utterance is to effectively control the listener's behaviour and secure the reinforcement from the listener on which the continuation of the speaker's turn and hence the conversation as a whole depends.

L4. *Novel Sentence Construction.*
As Chomsky (1957 etc.) has always insisted, though they are made up of units, words, phrases and sentence frames, which *are* repeated, sentences, in so far as they are used to convey information, are seldom repeated word for word, and are typically *constructed anew* on each occasion of utterance. But although any reasonably intelligent speaker can break down her sentences into their constituent words when called on to do so, in practice, sentences are assembled, not from their individual constituent words, but from what Miller (1998) calls "prefabricated chunks", phrases, embedded sentences and sentence frames which, in the overwhelming majority of cases, are borrowed by imitation from other speakers.

L5. *Novel Sentences and the Representation of Unfamiliar Contingencies.*
The speaker's ability to construct and the listener's ability to construe sentences the like of which neither party need have encountered before, allows the speaker both to induce the listener to emit behaviour which he or she may never have emitted before and communicate to listener information about contingencies (antecedent-behaviour-consequence relations) the like of which he or she need have had no previous experience.

L6. *Sentence-Construction and the Win-Shift/Fail-Stay Contingency*.
Unlike the situation presented to the organism by the Skinner box which is a *win-stay, fail-shift* contingency, at the level of sentence construction verbal behaviour is on *win-shift, fail-stay* contingency in which what are variously referred to as "back-channels". "response tokens" or "verbal reinforcers" play an essential role in indicating to the speaker that his or her sentence has been successful, thus allowing the speaker to proceed to the next sentence. Only when the listener's response indicates failure to communicate or persuade is a sentence repeated. Even when it is repeated, it is usually with a different intonation, if not with different words.

L7. *The Picture Theory of the Meaning of Sentences.*
According to the *picture theory of meaning* (Wittgenstein 1921/1961) in the form in which I subscribe to it, a sentence is complete when it depicts, in a manner capable of being deciphered by any competent listener, a complete *situation* (Barwise and Perry 1983). A situation is either an *event* whereby a change occurs in either the properties of an individual entity or the relations between two or more individuals, or a *state of affairs* in which a property of an individual or a relation between two or more individuals remains unchanged for a period of time. The nature of what persists or changes is indicated by the *verb phrase* or *predicate*, the individual or individuals involved by *noun phrases* or *arguments*.

L8. *The Associative Learning of Word and Phrase Meaning.*
Unlike the sentences of which they form part and to whose meaning they contribute, the constituents of sentences, *words*, *phrases* and *sentence frames*, are repeated. Unlike the sentences of which they form part and to whose situation-depicting function they contribute, words, phrases and sentence frames acquire their semantic function by virtue of a *repeated past association* between the linguistic unit (word or phrase) and the aspect of the various situations into whose depiction it enters and which it depicts.

L9. *Lexical Words, Syntactic Words and Bickerton's Proto-Language.*
Words as the smallest functionally discrete constituents of phrases and sentences are of two kinds; *lexical words* (nouns, verbs, adjectives and adverbs) and *syntactic words* or, as Skinner (1957) has it, "autoclitics". The developmental evidence shows that the earliest sentences that a child produces are ones consisting only of lexical words and in the first instance only of verbs and nouns (Bickerton's 1990 "*proto-language*"). There is also evidence (Morford et al. 1993: Morford 1996) from studies of "*homesigning*" (individual sign-languages developed by the deaf who have no access to other sign-language users) that *iconic signs*, those where the sign resembles the natural signs which regularly accompany and thus signal the presence of a particular object or kind of object or event, precede *symbolic signs*, those where the connection between the sign and what it stands for is arbitrary. Needless to say, all syntactic words, prefixes and suffixes are symbolic in this sense.

L10. *Mutations and the Facilitation of Language-Learning*.
There is evidence both from developmental events, such as the "*naming explosion*" and a "*critical period*" for the acquisition of syntax, and from the existence of corresponding structures in the human brain, *Wernicke's area* in the case of naming and *Broca's area* in the case of syntax, of mutations that have been selected in the course of human evolution which, together with the changes to mouth and larynx which have made speech possible, have given our species an ability to acquire language that no other species possesses. Nevertheless, the lack of continuity between these structures and the innate communication system of pre-linguistic organisms, a system which survives in our species in the form of the so-called "*language of emotion*", emphasises the fact that language is something that has to be *learned* by both speaker and listener. If it were not, it would not be able to adapt and contribute to the development of a new technology as the group moves into a new environment.

References

Adams, C. D. & Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology*, 33 B, 109-112.

Barwise, J. & Perry, J. (1983) *Situations and Attitudes*. Cambridge, MA: M.I.T. Press.

Bickerton, D. (1990) *Language and Species*. Chicago: University of Chicago Press.

Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton.

Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Ferster, C. B. & Skinner, B. F (1957) *Schedules of Reinforcement*. New York: Appleton-Century-Crofts.

Harzem, P. & Miles, T. R. (1978) *Conceptual issues in operant psychology*. New York: Wiley.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Hume, D. (1739/1978) *A Treatise on Human Nature*, L.A. Selby-Bigge (Ed.), 2nd Edition, P.H. Nidditch (Ed.). Oxford: Clarendon Press.

Konorski, J. (1948) *Conditioned Reflexes and Neuron Organization*. English translation by S. Garry. Cambridge: C.U.P.

Michael, J. (1982) Distinguishing between discriminative and motivational functions of stimuli. *Journal of the Experimental Analysis of Behavior*, 37, 149-155.

Miller, J. (1998) Acquiring spontaneous spoken language: the role of simple syntax and ready-made phrases. Paper presented to the 6th International Pragmatics Conference, Reims, 20th July 1998.

Miller, S. & Konorski, J. (1928) Sur une forme particulière des réflexes conditionelles. *C. R. Soc. Biol. Paris*, 99, 1155-1157. For a more detailed description of these experiments in English see Konorski (1948) pp. 211-235.

Montague, P. R., Gally, J. A., & Edelman, G. M. (1991) Spatial signalling in the development and function of neural connections. *Cerebral Cortex*, 1, 199-220.

Morford, J. P. (1996) Insights into language from the study of gesture: a review of research on the gestural communication of non-signing deaf people. *Language & Communication*, 16, 165-178.

Morford, J. P., Singleton, J. L. & Goldin-Meadow, S. (1993) The role of iconicity in manual communication. In K. Beals, G. Cooke, D. Kathman, S. Kita, K.-E. McCullough & D. Testen (Eds.) *Papers from the Chicago Linguistic Society*, 29, Volume 2: The Parasession, pp. 243-253.

Place, U. T. (1996) Linguistic behaviorism as a philosophy of empirical science. In W. O'Donohue & R. Kitchener (Eds.) *The Philosophy of Psychology*. London: Sage, Chapter 9, pp. 126-140.

Rescorla, R. A. (1991) Associative relations in instrumental learning: The eighteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 43B, 1-23.

Rescorla, R.A. & Wagner, A.R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.) *Classical Conditioning*, Vol. 2: *Current Research and Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B. F. (1957) *Verbal Behavior*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1969) *Contingencies of Reinforcement*. New York: Appleton-Century-Crofts.

Thorndike, E. L. (1911) *Animal Intelligence*. New York: Macmillan.

Wittgenstein, L. (1921/1961) Tractatus logico-philosophicus. *Annalen der Naturphilosophie*. *Tractatus Logico-philosophicus* With second English translation by D. F. Pears & B. F. McGuiness. London: Routledge & Kegan Paul.

Notes

1. Provided its grammatical object is not an embedded sentence in *oratio obliqua* or indirect reported speech, and unlike the verb `predict' which is often used in this context by associative learning theorists, the verb `expect' is one of the mentalistic concepts that does *not* presuppose linguistic competence on the part of the organism concerned and is therefore immune to the objection raised to such expressions in Section B3 above.

Evidence for the Role of Operant Reinforcement in the Acquisition and Maintenance of Linguistic Competence

Ullin T. Place
Thirsk, North Yorkshire

Abstract: *Linguistic competence is defined, following Chomsky, as the ability both to construct and construe novel combinations of familiar words and phrases which serve to depict situations the like of which neither speaker nor listener need have encountered in their own case. The view that linguistic competence in this sense is acquired and maintained according to the principle of selective operant reinforcement is defended, partly on grounds of evolutionary probability and the special nature of human environmental adaptation, and partly on the basis of two strands of empirical evidence: experimental evidence from studies of "verbal conditioning" and observational evidence of naturally-occurring verbal interactions in the work of discourse and conversation analysts. Although this form of selective operant reinforcement differs in a number of respects from that typically observed in the Skinner box, taken individually, there are precedents in animal learning for all the distinctive features of this form of learning. But if linguistic competence is acquired according to the same principles of selective operant reinforcement as apply in the case of animal learning, that principle by itself cannot explain why only humans have developed language and why apes can, at best, attain to the linguistic competence of a human two-year-old.*

Language is learned

Contrary to the opinion of Chomsky (1965), language, both the speaker's utterance and the listener's response, is learned behaviour. Linguistic competence, defined following Chomsky (1957 etc.) as the ability to construe and construct indefinitely many novel well-formed sentences in a particular natural language, is acquired and maintained by the same principle as that which applies in the case of animal learning and much of the learning in connectionist networks, the principle formulated by Thorndike (1911) as the "Law of Effect", by Skinner (1981/1984) as the principle of "Selection by Consequences" and by the connectionists (Rosenblatt 1959; Widrow & Hoff 1960; McClelland & Rumelhart 1988) as the "error-correcting or `delta' learning rule."

Why language must be learned

That language *must* be learned is evident from the following considerations:

(a) The pre-linguistic forms of communication analyzed by the ethologists (Tinbergen 1951) vary little within species, are demonstrably innate and permit learning only within narrow constraints, as illustrated by the phenomenon of `imprinting' (Lorenz 1935/1957).

(b) Whereas our nearest primate relatives, the anthropoid apes, are confined to a narrow ecological niche within the tropical rain forest, *homo sapiens*, and to a lesser but nevertheless significant extent, earlier hominid species have succeeded in colonising a variety of different habitats, not by developing physical characteristics adapted to those environments, but by devising and learning a new technology appropriate to that currently occupied. Moreover, if the Tower of Babel legend is to be believed, it is by learning to modify their language so as to talk about the new environment and the technology used to adapt to it that different natural languages have evolved from a single parent stock.

Why language must obey the law of effect

The following considerations show that if language is learned, it must be learned in accordance with the principles described by Thorndike, Skinner and the connectionists:

(i) The Law of Effect is an application of Darwin's principle of variation and natural selection to the process of learning. The only differences are that in this case the development is ontogenetic rather than phylogenetic, and that the units that survive or are eliminated are patterns of behaviour rather than individuals with certain inherited characteristics.

(ii) The Law of Effect is the only learning principle at the molar level of analysis that can effectively promote the survival and reproduction of the social group.

(iii) The introduction of a new set of learning principles to cope with the learning of language would be a biological extravagance, if there already exists, as there clearly does, a set of such principles tried and tested over millions of years of evolutionary history

Experimental evidence for the role of reinforcement in verbal behaviour

The proposition that linguistic competence is acquired and maintained by the Law of Effect is supported by two kinds of evidence: experimental evidence and evidence from the observation of naturally-occurring conversations and business transactions. The *experimental evidence* consists of a series of studies of what was then called "verbal conditioning" (Krasner 1958), published between 1954 and 1962 of which three stand out as both representative and evidentially persuasive.

The first and best known study is that of Greenspoon (1954; 1955). Greenspoon asked his subjects to say all the words that came into his or her head without repeating and without using sentences or phrases. He then selected a particular category of word, either plural nouns or non-plural nouns, for either reinforcement or disinforcement.[(1)] In two groups of subjects every time the subject produced a plural noun in one group or a non-plural noun in the other group the response was reinforced by the sound *Mmhmm*. In another two groups the same responses were disinforced by means of the sound *Uhuh*. In both cases there was a significant increase in the number of plural nouns or non-plural nouns in the groups in which these responses had been reinforced and a corresponding decrease in those categories of response in the groups in which the responses had been disinforced. Although it appeared from questioning after the completion of the experiment that 10 of the 75 subjects were able to verbalize the relation between the category of response and its reinforcing and disinforcing consequences as the case the case may be, the remaining 65 subjects were apparently unaware of this relationship and its effect on their behaviour.

Greenspoon's experiment can be criticised (Cf. Spielberger & Levin 1962) on the grounds that its artificiality is liable to prevent the consequences applied from having the effect that they do in the course of normal conversation, in particular, making the consequences much more salient than they are in normal conversation. This objection cannot be raised against Verplanck's (1955) experiment in which he persuaded a group of his students to engage an unsuspecting fellow student in one-to-one conversation for half an hour. After an initial "baseline" period of 10 minutes during which the number of opinions given by the subject was recorded, supposedly without any reinforcement being given, the experimenter was instructed during a further 10 minute period to make a point of agreeing with every opinion expressed by the subject. In a final 10 minute period, the experimenter was instructed to withdraw this reinforcement either by failing to respond to opinions expressed or, in some cases, by actively disagreeing with the opinion. In an alternative version of the same experiment some experimenters were asked to agree with the opinions expressed during the first ten minute period, withdraw reinforcement during the second ten minute phase and restore it during the final phase. In both cases there was a significant increase in the number of opinions expressed by the subject while the experimenter was agreeing with him; whereas during the non-reinforced phases, particularly after a preceding period of reinforcement, not only did the number of opinions expressed decline, but

"some *S*s got angry at *E* and commented on his disagreeableness, or noted his `lack of interest.'"

Apart from this,

"No *S* ever gave evidence that he was `aware' that his behavior was being deliberately manipulated and recorded, or that there was anything peculiar about the conversation." (Verplanck, 1955, p. 671).

The third piece of experimental evidence comes from a paper by Azrin, Holz, Ulrich & Goldiamond (1961) which records a series of attempts to repeat the Verplanck experiment. Of these I shall mention only the last in which the experimenters were trained animal behaviour researchers rather than undergraduate students, as in Verplanck's original study. With these experimenters

"Out of 12 attempts, not one of the four *E*'s could complete his experiment. It may be recalled that the procedure requires that the *E* restrict himself to agreement (or disagreement) of opinions, and stipulates no questions, statements, nods, smiles, or other types of interaction. The reason for forbidding such behavior proved to be obvious: *E*'s reaction, however subtle, could often be seen to exert profound but

uncontrolled effects upon the conversation of the subject. In the absence of any reaction by the four *E*s, however, all of the twelve *S*s terminated the conversation within 10 minutes by leaving the room where the conversation was taking place." (Azrin *et al.* 1961, p. 29)

This result is taken by the authors as casting doubt on the reliability and validity of the original Verplanck experiment. In fact what it shows is that you cannot maintain a conversation unless the speaker's verbal behaviour is reinforced in some way by the listener. It is clear from this that the student experimenters in the original Verplanck experiment and in the two earlier repetitions of it reported by Azrin *et al.* "instinctively knew this", i.e., they followed their deeply ingrained linguistic habit of invariably supplying such reinforcement in the form of head nods or what they took to be non-committal utterances, such as "I see", for the most part without being aware of so doing.

Observational evidence of reinforcement in naturally -occurring verbal interactions

The observational evidence for the role of reinforcement in verbal behaviour is to be found in the massive corpus of tape-recorded conversations that has been collected over the past 25 years both by social psychologists and linguists working in the field known as "discourse analysis" and by sociologists in the field known as "conversation analysis". What this corpus shows is that verbal reinforcement of the kind investigated in the three experimental studies I have described are to be found in every naturally-occurring verbal interaction between two people whether on the telephone or face-to-face. But, because of the disrepute into which behaviourism and its technical terminology have fallen over that period, they are not referred to as verbal reinforcers. Discourse analysts call them "back-channels". Conversation analysts call them "response tokens".

```
Penny: it's just this bus'ness of (.) th' party [for the first y:e:ars. . . .01
. . . . . .^[ . . . . . ^
Rose: . . . . . [ye:(s) . . . . . yes= . 02
Penny: =I won't (.) be i:n tomorrow mo:rning. . . . . .03
. . . . . ^
Rose: . . . . . no= . . . . . .04
Penny: =I've left a notice on the board. . . . . 05
. . . . . .^
Rose: . . . . . .yeah.= . . . . . .06
Penny: =and there's a note for them °of the money. . . . . 07
. . . . . .^
Rose: . . . . . .who wants to pick it up?= . 08
. . . . . ^
Penny: =we:ll (.) the:'re on that li:[st. . . . . .09
. . . . . .[ .^
Rose: . . . . . .[oh the're °all on that list.= . . . . .10
. . . . . .^
Rose: = (.) and any-any of these people [can have i, (.) can they.= . . . . .11
. . . . . [ . . . . . ^ . . . . .^
Penny: . . . . . .[yes:: (.) . . . . . 12
Penny: =I do:: know John's girl friend knows about it.= . . . . . .13
. . . . . .^
Penny: =bu(t) she's not free at the same time as them tomorrow.= . . . . . 14
. . . . . ^
Penny: =so:th't lots of people know about it,= . . . . . 15
. . . . . ^
Rose: =anan the:'re goin(g) to get the shoppin(g) ou[t of it. (.) I see= . . 16
. . . . . .^ .[ . . . ^ . . . . .^
Penny: . . . . . .[yes (.) . . . . . 17
```

Table 1 The Party 10/85

Table 1 (Place 1991; 1992; 1997a) is typical of this evidence. The upward pointing arrows show the points where a sentence is completed by the speaker. You will see that in almost every case, each sentence triggers a response from the listener. Most of these are reinforcers. `Yes' where the speaker's sentence is affirmative `No' where it is negative. On two occasions, on lines 8 and 11, the listener's response is a disinforcer in the form of a question requiring the previous speaker to either confirm the listener's interpretation or restate what has just been said.

Do verbal reinforcers reinforce?

Verbal reinforcers are of two kinds (Place 1997a) "*continuers*" whose function is to reassure the speaker that the sentence has been understood and accepted and thus allow her to move on to the next sentence and "*terminators*" whose surface effect is to close a speaker's turn and allow the listener to take over as speaker.[(2)] The implication of describing such events as `reinforcers' is that they have two kinds of effect. On the one hand a reinforcer may have an *intra-episodic* effect whereby the current on-going flow of the speaker's behaviour is maintained. In the reinforcement of verbal behaviour only continuers have this intra-episodic function; and the fact that they have it is clearly demonstrated both by the experimental and by the observational evidence I have cited. Much less easy to demonstrate is the *extra-episodic* effect of both continuers and terminaters, their effect in maintaining and enhancing the speaker's linguistic competence, her propensity to construct and utter sentences in the future which are relevantly similar to those that have been successful in the past, while avoiding constructions and usages that have proved unsuccessful. For the behaviourist the existence of such effects is an article of faith which is seldom questioned. For others, given the virtual impossibility of conducting a controlled experiment on the effect of systematically reinforcing incorrect usage and disinforcing correct usage, the evidence is at best circumstantial. The least circumstantial evidence relates to the initial acquisition of linguistic competence by the child rather than to its subsequent maintenance in the adult. It comes from Ernst Moerk's (1983) re-examination of Roger Brown's (1973) recordings of mother-child interaction in the early stages of language development on which Brown based his claim that children acquire the ability to construct grammatically well-formed sentences despite the fact that the mother never corrects grammatical mistakes. What Moerk's analysis shows is that although it is true that the mother in Brown's data more often supplies positive reinforcement in response to the child's utterances, there is a clear difference in her response between the cases where the utterance is grammatically correct and those where it is grammatically deviant. Not only is the positive reinforcement noticeably more enthusiastic in the former case; it is followed in the deviant cases by the mother's modelling of a corrected version of the child's utterance.

For what it is worth, my own hunch is that the child's verbal interactions with its peers are equally, if not more important than parent-child interactions in the acquisition and maintenance of the child's ability to construct intelligible sentences. The reason for thinking this is the observation that parents, particularly the kind of doting middle class mothers like the one whose behaviour is recorded by Brown, tend to be rather lenient in what they require of a child by way of grammatical structure before supplying reinforcement. Peers, though unconcerned with the niceties of correct literary speech, are much less tolerant than parents of syntactic and semantic deviations which, in the absence of the parent's intimate knowledge of her own child, render an utterance unintelligible to them. One wonders how far the prestige of Chomsky's (1965) nativist theory of the acquisition of linguistic competence may not be due to the delusion on the part of such parents that *they* are the sole source of their child's verbal education. Given this delusion, when a child produces a sentence structure which they know they have never taught the child to produce, it is hardly surprising that this is seen as evidence of the unfolding of an innate syntactic ability.

For the extra-episodic effect of reinforcement and error-correction on the maintenance of the linguistic competence of adults there is circumstantial evidence in the shape of the existence within every verbal community of social conventions backed up by speaker's anger and a serious threat of social ostracism, if they are not observed, whose function is to ensure that every intelligible utterance by the speaker is suitably reinforced, and that where a disinforcing interjection is required, its effect in provoking the speaker's anger is minimised, as in the phenomenon known to conversation analysts as `preference organisation'. If these conventions applied only to the delivery of the continuers that maintain on-going verbal interactions, it would be possible to argue that their function was confined to maintaining verbal interactions for as long as it is in the interests of both parties that they should do so. But that would not explain why the conventions are equally insistent on the supply of terminating reinforcers at the end of a speaker's turn. That only makes sense as an encouragement to the speaker to repeat similar utterances on relevantly similar occasions in the future.

Another convention which only makes sense on the assumption that appropriate reinforcement is essential to the maintenance of linguistic communication within the verbal community is the curious practice which Harry Stopes-Rowe (personal communication) has described as "mutual verbal grooming" in which slight acquaintances greet one another by exchanging remarks about the weather which go no further than what is already patently obvious to both. This practice, I submit, makes sense only on the assumption that its function is to reassure both parties that their ability to construct and construe mutually intelligible sentences describing the current situation is intact and is thus available, should it be needed, for

some more serious purpose later. Were that ability guaranteed by an innate language faculty, such reassurance would not be necessary.

The peculiarities of verbal reinforcement

But while there is good evidence that the acquisition and maintenance of linguistic competence obeys the Law of Effect, the learning situation involved is very different from that encountered in the Skinner box. Four differences stand out:

(i) In the Skinner box the behavioural unit that secures the reinforcement, the key-peck or bar-press, is the same unit that is subsequently repeated. In verbal behaviour the unit that secures reinforcement is the utterance of a complete sentence. But sentences, as Chomsky has often reminded us, are seldom repeated word for word. They are typically constructed anew on each occasion of utterance. The units that *are* repeated, the words, phrases and sentence frames, secure reinforcement only as part of an uttered sentence.

(ii) Whereas the Skinner box provides a *win-stay/fail-shift* contingency, at the level of sentence construction, verbal behaviour is on a *win-shift/fail-stay* contingency (Place 1997a; 1997b). When your sentence succeeds as evidenced by the listener's response, you do not repeat yourself; you go on to the next sentence. You repeat yourself only when the first attempt fails, usually saying the same thing in slightly different words.

(iv) The delivery of food which reinforces responding in the Skinner box strengthens the organism's propensity to emit *any* behaviour which regularly precedes it. Verbal reinforcers are specific to a particular class of behaviours and act as reinforcers only with respect to behaviour of that kind. Thus opinion-stating is reinforced by an expression of agreement, instruction-giving by an expression of comprehension, news-telling by an expression of interest or surprise, troubles-talk (Jefferson 1988) by an expression of sympathy, and joke-telling by laughter.

(iii) Whereas the delivery of a food pellet is an extremely salient stimulus event which cannot fail to occupy the focus of the organism's attention, the speaker is almost totally oblivious of the delivery of the reinforcers supplied by the listener, as is the listener of supplying them. As Verplanck remarks, they attract attention only when they are omitted on an occasion when their delivery is to be expected.

The peculiarities of verbal reinforcement are not unique

Despite these differences there is nothing in the way verbal behaviour is reinforced which does not have its parallels in the operant or instrumental reinforcement of animal behaviour. The dissociation between the response that secures reinforcement and the behaviour that is strengthened as a result has its counterpart in the reinforcement of any non-stereotyped skilled performance. Win-shift/fail-stay contingencies have their counterparts in the foraging behaviour of animals from which the concept derives. The restriction of the reinforcement of particular varieties of verbal behaviour to a particular reinforcer has its counterpart in the secondary reinforcers whose appearance allows the animal to move on to the next component in a behavioural chain (Gollub 1977). The analogy with the reinforcement of verbal behaviour is particularly close in the case of a chain in which two organisms are required to respond in alternation in order to obtain the primary reinforcer at the end of the chain. Note also that secondary reinforcement within a behavioural chain is another case of a win-shift/fail-stay contingency. Finally the reinforcement of behaviour in the absence of awareness of the reinforcer almost certainly has a counterpart in the reinforcement of behaviour by electrical stimulation of the hypothalamus (Olds & Milner 1954), a form of stimulation which we can be tolerably certain from what we know about the effect of stimulating other sub-cortical structures, produces no awareness of the stimulus in human subjects. One may surmise that such stimulation will be accompanied by a vague glow of satisfaction. But so, in my experience, is a well reinforced conversational turn, even though the reinforcers themselves are barely noticed. While it is difficult to be sure that what humans are and are not conscious of corresponds to what animals are and are not conscious of, the recent work of Cowey and Stoerig (1995) showing that monkeys with unilateral lesions of the striate cortex can learn to reach for objects they cannot "see", in exactly the same way that human blindsighted patients can, makes such inferences much more plausible than they used to be.

Operant learning cannot explain the evolution of language

But if the learning process involved in language acquisition is in principle no different from that which we observe in animal learning, it follows that we cannot appeal to the principles of learning in order to explain why it is that only human beings have developed language and why, although apes, can be trained to communicate by means of linguistic symbols, even the most intelligent and symbolically sophisticated cannot progress beyond the linguistic competence of a two-year-old human child. Nor, by itself, is the mutation or set of mutations which have changed the conformation of the human mouth and larynx so as to permit vocal speech sufficient to explain the phenomenon. For the congenitally deaf can learn to communicate in a distinctively linguistic way, even without being taught an official sign-language (Goldin-Meadow & Mylander 1984; 1990). Something has happened to the human brain associated with the development of such structures as the angular gyrus, Wernicke's and Broca's area in the dominant hemisphere of the cerebral cortex which have made it very much easier for human beings to learn the kind of associations and generalisations involved in language than it is for any animal species.

Conclusion: towards a plausible theory of language evolution

Chomsky (1965) asks us to believe that language evolved fully grown with all its syntactic complexity in a single gigantic mutation. Like many others, I find this view frankly incredible. The ability to learn those special associations that are peculiar to language must have evolved along with the development of the human vocal apparatus in a sequence of steps or stages. At each stage a mutation must have occurred which was passed on to the descendants of those in whom it occurred by virtue of its utility in securing the survival and reproduction of the social group and groups constituted by those descendants. But, conceding a role for mutations and the innate learning capacities with which they endow the organism should not be seen as in any way diminishing the importance of learning in the acquisition and maintenance of linguistic competence. On the contrary, what has been acquired in the course of the evolution of our species is the ability to learn the skills of constructing and construing intelligible sentences in whatever natural language or technical code currently prevails within the verbal community of which by virtue of possessing those skills the individual is a member.

References

Azrin, N. H., Holz, W., Ulrich, R. and Goldiamond, I. (1961) The control of the content of conversation through reinforcement. *Journal of the Experimental Analysis of Behaviour*, *4*, 25-30.

Brown, R. (1973) *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Chomsky, N. (1957) *Syntactic structures*. 'S Gravenhage: Mouton.

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cowey, A. and Stoerig, P. (1995) Blindsight in monkeys. *Nature*, *373*, 6511: 247-9.

Goldin-Meadow, S. and Mylander, C. (1984) Gestural communication in deaf children: The effects and non-effects of parental input on early language development. *Monographs of the Society for Research in Child Development*, *49*, 1-121.

Goldin-Meadow, S. and Mylander, C. (1990) Beyond the input given: The child's role in the acquisition of language. *Journal of Child Language*, *17*, 527-563.

Gollub, L. (1977) Conditioned reinforcement: Schedule effects. In W. K. Honig and J. E. R. Staddon (Eds.) *Handbook of operant behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Greenspoon, J. (1954) The effect of two non-verbal stimuli on the frequency of two verbal response classes. *American Psychologist*, *9*, 384. (Abstract)

Greenspoon, J. (1955) The reinforcing effect of two spoken sounds on the frequency of two responses. *American Journal of Psychology*, *68*, 409-416.

Harzem, P. and Miles, T. R. (1978) *Conceptual issues in operant psychology*. Chichester: Wiley.

Jefferson, G. (1988) On the sequential organization of troubles-talk in ordinary conversation. *Social Problems, 35*, 418-441.

Krasner, L. (1958) Studies of the conditioning of verbal behavior. *Psychological Bulletin*, *55*, 148-170.

Lorenz, K. (1935/1957) Der Kumpan in der Umwelt des Vogels; die Artgenoße als auslösendes Moment sozialer Verhaltungsweisen. *Journal of Ornithology*, *83*, 137-213 and 289-413. English translation as `Companionship in bird life: fellow members of the species as releasers of social behavior' in C. H. Schiller (Ed.) *Instinctive Behavior*. New York: International University Press.

McClelland, J. L. and Rumelhart, D. E. (1988) *Explorations in parallel distributed processing*. Cambridge, MA: MIT Press.

Moerk, E. L. (1983) *The mother of Eve - As a first language teacher*. Norwood, NJ: Ablex.

Olds, J. & Milner, P. (1954) Positive reinforcement produced by electrical stimulationof septal area and other regions of the rat brain. *Journal of Comparative and Physiological Psychology*, *47*, 419-427.

Place, U. T. (1991) Conversation analysis and the analysis of verbal behavior. In L. J. Hayes and P. N. Chase (Eds.) *Dialogues on verbal behavior: The first international institute on verbal relations*. Reno, NV: Context Press, Chapter 5, pp. 85-109.

Place, U. T. (1992) Eliminative connectionism and its implications for a return to an empiricist/behaviorist linguistics. *Behavior and Philosophy*, *20*, 21-35.

Place, U. T. (1997a) Contingency analysis applied to the pragmatics and semantics of naturally occurring verbal interactions. In J. L. Owen (Ed.) *Context and communication behavior*. Reno, NV: Context Press, Chapter 18, pp. 369-385.

Place, U. T. (1997b) Rescuing the science of human behaviour from the ashes of socialism. *Psychological Record47*, 649-659.

Rosenblatt, F. (1959) Two theorems of statistical separability in the perceptron. In *Mechanisation of Thought Processes: Proceedings of a Symposium held at the National Physical Laboratory, November 1958. Vol. 1*. London: HM Stationery Office, pp. 421-456.

Skinner, B. F. (1981/1984) Selection by consequences. *Science*, *213*, 501-504. Reprinted with peer commentary in A. C. Catania and S. Harnad (Eds.) Canonical papers of B. F. Skinner. *The Behavioral and Brain Sciences*, *7*, 477-481.

Spielberger, C. D. and Levin, S. M. (1962) What is learned in verbal conditioning? *Journal of Verbal Learning and Verbal Behavior*, *1*, 125-132.

Thorndike, E. L. (1911) *Animal intelligence*. New York: Macmillan.

Tinbergen, N. (1951) *A study of instinct*. Oxford: Clarendon Press.

Verplanck, W. S. (1955) The control of the content of conversation: reinforcement of statements of opinion. *Journal of Abnormal and Social Psychology*, *51*, 668-676.

Widrow, G. and Hoff, M. E. (1960) Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4*, 96-104.

Notes

1. For the use of this term see Harzem & Miles 1978.

2. As a boy, my father, so he used to tell me, had the curious nickname `Ayemanayebygo'. This was a reflection of his invariable practice in those days (c. 1885) of responding to information supplied by another boy with the two continuers *Aye man!* and *Aye!* followed by the terminator *By Go!* (a euphemistic curtailment of *By God!* used as an expression of surprise).

Reviews

Terrence Deacon, *The Symbolic Species*

1997, Allen Lane: The Penguin Press: London. ISBN: 0 -713-99188-7

Mary Strate
Evolutionary Psychology Research Group
Dept. of Biological Sciences
University of Liverpool

The evolution of language is a difficult topic to grasp. There are many issues surrounding hominid evolution which can be, and have been, used to account for the cognitive gulf which separates modern humans from other nonhuman primates. It has proved difficult to isolate any single underlying factor which served as the impetus for our lineage's drastic divergence away from our apelike ancestors and Deacon sets himself a heavy task. In his preface, the author identifies some of the key questions:

- How is language different from other forms of communication?
- Why do other species encounter virtually intractable difficulties when it comes to learning even a simple language?
- How have human brain structures evolved to overcome these difficulties?
- What forces or conditions initiated and steered us along this unprecedented evolutionary course?

The Symbolic Species is divided into three relatively fluid parts: language, brain, and the coevolution of the two; and the book discusses what makes humans unique in light of cognitive, linguistic, and neuroanatomical evidence.

Language

Deacon initially chastises the reader for our tacit assumption that nonhumans merely have `simple languages', because although nonhuman communication can be subtly complex, it is fundamentally different from human language. Following this line of argument, Deacon does not see how comparative neurological and ecological studies will illuminate which variable(s) influenced cortical expansion and increasing linguistic complexity in humans (but see Dunbar, 1993; Hauser, 1996). For Deacon, the linguistic rubicon is *symbolic reference*, humans have a social convention that consists of an established code linking objects with words -outwith of any physical similarity. This, Deacon argues, differs from iconic reference, where the calls are a resemblance of something we notice, and indexical reference, where the signal is causally linked to the sign that it depicts. Our ability to spontaneously refer to events symbolically through language is contrasted to other species' communicative abilities (*e.g.,* ASL ape research from Savage-Rumbaugh's laboratory and Hoover the seal) and uniquely human dysfunctions (*e.g.,* Williams Syndrome and autism).

One of the now-classic studies which Deacon examines is the vervet alarm call work by Cheney, Seyfarth and colleagues (*e.g.,* Seyfarth *et al,* 1980). Vervets classify different types of predators acoustically, and this has been hailed as evidence for symbolic referential capabilities in nonhuman primates, yet the more parsimonious interpretation of indexical reference cannot be overlooked. However difficult it is to understand how intention is necessary for symbolic reference, experts remain cautious about ascribing volition to nonhuman primate vocalisations (e.g., Hauser, 1996). For example, Deacon cites J. Goodall's observation of a chimpanzee muffling a food call (using his hand over his mouth) as the absence of volition and suggests that human brains are unusual for having cortical motor control over sound production when closely related species have none.

Brain

One of the first steps when analysing brain size differences over evolutionary and taxonomic categories is to account for allometric variation. Deacon neatly summarises the problems of scaling when looking at brain/body relationships and throws the odd teaser out, `What makes selection on brain size a more

attractive hypothesis than selection on body size' (p.168), which can provoke one to continue reading to see how he deals with the question.

Deacon applies cogent arguments to illustrate how the human vocal repertoire differs from that of other animals. Amongst his exemplars are babbling (the spontaneous and frequent phonemic sampling which human babies exhibit), motherese (an infant-directed form of speech employed with exaggerated prosody and rhythmicity), laughter, and sobbing - all of which are intriguing in their own right. For example, babbling (thus far a uniquely human vocalisation) concurs with the onset of cortical motor tract myelinization which may be related to self-directed speech. Laughter is aspirated in contrast to the `play chuckle' of nonhuman primates. Deacon suggests that the production of even such `basic' calls set humans apart.

The framework of *The Symbolic Species* is set around asking what it is about our cognition that sets us apart from other species and what neuroanatomical evidence lies behind these assumptions. Many of the `requisites' for human language (*e.g.,* thoracic enlargement, laryngeal descent, finer hypoglossal control, etc.,) may be consequential vocal abilities which were selected for over evolutionary time, once the foundation for human language was set down. Humans have undergone significant cortical expansion when compared to other nonhuman species; for example, the prefrontal cortex has undergone at least a two-fold increase over that of nonhuman primates. Deacon discusses brain elaboration using empirical evidence ranging from homeobox gene homologues to xenotransplants of quail mesencephalons to illustrate that human cortical expansion is not simply extra neurones- rather connectivity and neural organisation. For Deacon, the rubicon of language is symbolic reference, and the enlargement of the prefrontal cortex with its increased projections to other brain structures and interconnectivity have laid the groundwork.

Coevolution

Deacon then uses Baldwinian evolution, where behavioural flexibility can channel genetic predispositions, to dispel tacit acceptance of theories of universal grammar and the language instinct (*c.f.,* Chomsky, 1976; Pinker, 1994). There has been scant opportunity for innate grammatical abilities to develop although certain structural features of language could have already become internalised over the course of time and he creates a mythical species, *Homo symbolicus,* to illustrate his point.

H. Symbolicus is a nöospecies which, hypothetically, could be the first hominid to habitually use symbolic communication; as information transmission is independent of genetic descent, its morphology cuts across the fossil record determining relationships potentially independent of phylogenetic descent. This example is ambitious, as it must allow for a myriad of behavioural genetic modifications, but the fundamental question which he raises with *H.symbolicus* is poignant: Is long term selection on brain size and other internal organs critical to symbolic learning?

Deacon re-examines brain size trends and suggests that Baldwinian evolution can explain the predicted trend in increasing prefrontalisation as bearing a corresponding shift in learning predispositions. Symbolic reference is a diagnostic trait of *H.symbolicus*, and `...[the] origin of `humanness' can be defined as that point in our evolution where these tools became the principle source of selection on our bodies and brains.' (p.345) Many authors concur that a positive feedback between the socio-ecological demands and the need for cognitive flexibility of early hominids could have existed, but Deacon interprets the data in reverse: the ability for symbolic reference precedes any ecological or social influence on encephalisation because this novel tool with which to manipulate information created the need for a new adaptive complex.

Deacon asks why symbolic communication arose as it carries the costs of cognitive effort, time, social organisation, and reduced efficiency when compared to other forms of social communication. The socio-ecology of a social carnivore and intersexual competition are introduced in an attempt to resolve this because constraints such as the risk of cuckoldry are major threats when males must provision nursing females and their immature offspring with meat. To avoid the breakdown of sexual exclusivity, hominids developed social contracts, promises of fidelity, which can only be expressed symbolically. The economic and sexual pair bond of humans is a symbolic agreement, and is not found in other species as they do not have symbolic abilities.

Symbolic culture was a response to a reproductive problem that only symbols could solve: the imperative of representing a social contract (p.401)

Deacon's major tenet is that neurological predispositions and socio-ecological constraints sponsored the development of language, and that the feedback system which naturally arose built an ever more complex coevolution of the two.

In conclusion, Deacon's book is informative and well-written. Deacon often arrives at different interpretations from what might be called accepted theories in the field, perhaps due to the fact that Deacon does not centre his arguments around a socioecological perspective, but the book itself is rewarding as it forces the reader to see how the same evidence can be used for contrasting arguments. The only major criticism one might have is the book's lack of an socio-ecological perspective. *The Symbolic Species'* examination of the evolution of language and the adaptive complex between it and the brain has neatly tied together cognitive, philosophical, neuroanatomical, and fossil evidence.

References

Chomsky, N. (1976). 'On the nature of language.' In: *Origins and Evolution of Language and Speech*. H.B. Steklis, S.R. Harnad, and J. Lancaster (eds.); New York Academy of Sciences, New York.

Dunbar, R.I.M. (1993). 'Coevolution of neocortical size, group size and language in humans.' *Behavioral and Brain Sciences*. 16: 681-735.

Hauser, M.D. (1996). *The Evolution of Communication*. Bradford/MIT Press. 760 pp.

Leiberman, P. (1995). 'What primate calls can tell us about human evolution.' In: *Current Topics In Primate Vocal Communication*. E. Zimmermann, J.D. Newman, U. Jurgens (eds.); Plenum Press, London; pp. 273-282.

Pinker, S. (1994). *The Language Instinct*. Penguin Book Ltd., London, 494 pp.

Seyfarth, R.M.; Cheney, D.L.; Marler, P. (1980). 'Vervet monkey alarm calls: semantic communication in a free-ranging primate.' *Animal Behaviour*. 28:1070-1094.

Michael S. Gazzaniga, *The Mind's Past*

Berkeley and Los Angeles: University of California Press, 1998, xv+201pp, U$ 22 ,50
ISBN 0-520-21320-3 (hbk).

Joao Teixeira
Center for Cognitive Studies
Tufts University

Nothing could produce a more disheartening feeling than the idea that we are just puppets controlled by our brains - brains so smart that they could even produce the illusion that we control our own thoughts and actions. This is the leit-motiv of this book by one of the founders of cognitive neuroscience where a defense of this most outrageous thesis is presented. Gazzaniga's main endeavor in his new book is to present a thorough assault on the notion of "self" and to argue that such a notion can no longer survive the impact of contemporary brain science. In seven elegantly written chapters the author provides the educated layman with an overview of several themes addressed by the new and promising field of cognitive neuroscience. These range from a discussion of the nature of the "self", brain architecture, the relationship between information-processing structures of the brain and experience, to the nature of perception, action, memory and consciousness.

The first chapter is striking for its very suggestive title: "The Fictional Self". It is devoted to a presentation of what Gazzaniga takes to be one of his major findings in neuroscience, namely, the existence of an "interpreter" on the left hemisphere of the brain. Such an interpreter is not a "self" nor "part of a self" but - as the author points out - a brain device that accounts for a reconstruction of our past experiences, thus "weaving its story in order to convince itself and you that it is in full control" (p.25). Furthermore, by providing us with some kind of personal story or an experience of an ongoing narrative the interpreter or "what amounts to a spin doctor in the left brain" (p.26) gives us the sensation that the "self" exists, detached from the brain. Such a "detached self" is, nonetheless, illusory - a sheer by-product of brain activity attempting to gather the multifarious output of cortically based automatic systems working outside of conscious awareness.

The second chapter focuses on "brain construction" and it aims to show that most of the development of brain structure and functioning is due to genetic pre-programming. The brain is not a tabula rasa nor is it mostly shaped by the environment. Less emphasis should be assigned to environmental factors

in the development of our mental devices. This is surely one of the most controversial ideas presented in this book - an idea that clashes with some current trends of contemporary neuroscience. By assigning excessive importance to the role of both environment and brain plasticity, experimental psychologists have mistakenly urged the conclusion that "genetic specification plays little or no role in the development of our mental devices" (p.13). In accordance with such a view, "the brain is idling in neutral until it experiences the world"(p.38).In the nature/nurture debate that still pervades neuroscience, Gazzaniga argues in favor of nature, by emphasizing that nurture alone is not enough to shape brain function. He advocates a prominent role for modularity, adaptive specialization and genetically driven mechanisms in the ontogenetic/phylogenetic development of the brain. In several passages of Chapter II pro-nurture arguments are acrimoniously ridiculed, either as an article of popular science or as an hypothesis at odds with more accurate empirical evidence.

The third chapter returns to the main thesis of the book. The task faced is that of showing that our brain - and our brain mechanisms - control our cognition and behavior and not vice-versa. The contention is presented in the opening sentence of this chapter: "By the time we think we know something - it is part of our conscious experience - the brain has already done its work. It is old news to the brain, but fresh to us" (p.63). We are not the masters of our own mental processes and whatever we "decide" to do next our brain has already decided for us a few milliseconds ago. The "self" as a result of our mastering of our cognitive processes is nothing over and above a delusive aspect of our own cognition - a delusion ultimately produced by the working of our brains.

One of the brightest chapters of Gazzaniga's work bears on the nature of memory (Chapter 6). By dismantling the current conception of memory as the immense archive of recollections - a conception still largely inherited by many cognitive scientists - Gazzaniga achieves one more step in his assault to the notion of "self". Most of our memories are re-invention i.e., a reconstruction to fill out gaps in the narratives we produce about ourselves. So viewed, most of the puzzle of memory storage and retrieval that puzzles psychologists, neuroscientists and cognitive scientists is just a false riddle, whose roots are to be found in the popular conception of the memory as an immense storage of recollections dutifully organized.

Gazzaniga's sober scientific prose could mislead the reader to the idea that no philosophical controversy lurks beneath his approach to the main themes of cognitive neuroscience. No doubt one could contend that there is not very much novelty in this book except for a suitable presentation of his ideas for a broader audience. Nevertheless, the philosophical gist of Gazzaniga's theses crops up once we consider that his book raises a mostly controversial and disquieting issue for philosophers and cognitive scientists: Will the success of cognitive neuroscience in the explanation of mental/cognitive phenomena entail that the notions of "mind" and "self" are doomed to disappear? Will the ideas of mind and of self become obsolete scientific concepts in the same way that the discovery of oxygen led to the superseding and obsolescence of the notion of "phlogiston"?

Recent developments of contemporary brain science may incline us to believe that some traditional philosophical problems may be ultimately reduced to scientific ones i.e., amenable to the tools of empirical science. An illustration of such a change of conceptual status in the history of science would be provided by the notion of vacuum. In the seventeenth century discussions concerning the nature of vacuum were a matter of philosophical dispute until modern physics could approach it as a scientific, empirical issue. According to such a view, the notion of vacuum was initially infested with philosophical/ontological presumptions concerning the conceivability of "nothingness". Such philosophical presumptions prevented the clarification of the notion of vacuum by thwarting its conception as an empirical entity. Once such presumptions were shunned as secondary or parasitic, the conundrum was solved: The notion of vacuum became a scientific, empirically tractable problem. The same movement would be pursued by contemporary brain science by seeking to show that concepts such as consciousness, representation, self, etc., can undertake the same change of conceptual status once we find their neural correlates.

Does Gazzaniga's approach to the notion of the self entails such an eliminativist character? There is a huge difference between elimination and revision. As a first approximation one would be inclined to derive from Gazzaniga's work the claim that the major problems of philosophy of mind could ultimately boil down to elimination. So viewed the problem of the nature of the self would be solved if it could be rephrased as empirical problems addressed by brain science. The problem of the self would become nothing over and above the problem of finding its neural correlates. However, this is a hasty interpretation of Gazzaniga's enterprise. Indeed, he proclaims that "Psychology itself is dead" (p. xi) and that "the grand questions originally asked by those trained in classical psychology have evolved into matters other

scientists can address" (p.xii). But is it all he is proclaiming? A few lines below he carefully asserts that "we human beings have a centric view of the world. We think our personal selves are directing the show most of the time" (p. xiii). It is not the self per se that must be dissolved, but the centric view of it. Such a centric view of the self is to be superseded if we are committed to the development of a serious scientific account of mentality. The real self is a brain device - a very sophisticated one in so far as it has the capability of engendering such a thing as "the centric self". The "centric self" lures us to first-person certainty about our own nature, leading to a mismatch between what our brain does and how we experience such an activity. The centric self is likely to be revealed a delusion once we find its neural correlates and realize that there is nothing over and above a brain device that unifies thought and action by weaving a fictional story. We are misled to the idea of the centric self as an autonomous originator and to the systematic illusion that we are in full control of the activity of a multitude of automatic systems responsive to internal and external stimuli. The illusion of the existence of such a centric self as well of its proclivities is what is to be shunned. But not the story it weaves about itself, for no matter how fictional it may be it plays a major role in our cognitive capabilities, such, as for instance, reasoning and several others that "enabled us to become psychologically interesting to ourselves as a species" (p.152).

So viewed, Gazzaniga's conception of the nature of the self does not succumb to any strict eliminativist program. There would be more to the notion of the self than the finding of its neural correlates. The target does not seem an elimination of conscious experience but, rather, that of finding from whence comes the mismatch between the conscious experience of the self and its possible empirical description at the sub-personal level.

Gazzaniga's strictly neurological approach to the nature of the self has no explicit philosophical agenda. Nonetheless, his view of the brain as a cluster of specialized circuits and of consciousness as emerging from the feeling of them challenges traditional philosophical assumptions. In his conception of the self there is no room left for what Dennett would pejoratively label "The Cartesian Theatre" or the centre of the mind/brain where a central controlling unit is located - the very arena where consciousness happens. Both Gazzaniga and Dennett would rather bet on the hypothesis of an orchestra without a conductor - a hypothesis that allows them to dispense with further assumptions concerning the existence of any underlying unifying element for our conscious experiences.

To what extent can Gazzaniga's left-brain interpreter provide a full-fledged account of the nature of the self, including the generation of the notion of a "centric self"? Or, in other words: Can a brain device per se account for the generation of our habitual centric self? One issue not addressed by Gazzaniga is how the instantaneous reconstruction of the mind's past can lead to the illusion of the centric self - an illusion inherited and cherished by traditional philosophy of mind. For whence comes the feeling that we are endowed with some inner initiating cause of thought and action? For one thing, such a feeling seems to emerge in so far as some preliminary assumption of incorrigibility of the mental comes into play. It is hard to conceive that such an assumption could emerge without the contribution of culture and language - some kind of language that forges a preliminary idea of "I" allowing us to speak and think about ourselves as centric selves endowed with the power to produce autonomous action. Would a culture without a word for "I" develop some idea of a centric self? To what extent the idea of a social construction of the self would mark an essential dissimilarity between Gazzaniga's and Dennett's conception of the self?

Surely a conception of the formation of a centric self such as Dennett's differs from Gazzaniga's in so far as the former would assign much more weight to "linguistic memes" in the production of a social image of the self - the centric self deeply inculcated by language and perpetuated by social roles. Still, from a philosophical viewpoint there is more convergence than dissent between Gazzaniga's and Dennett's approaches to several other topics. Dennett construes the stream of consciousness as resulting from a virtual serial machine installed on the massive parallel information processing provided by the brain. Dennett endorses multiple realizability whereas for Gazzaniga there is more emphasis on the specificity of brain devices in the production of experience. Nonetheless, both would agree with some idea of a narrator. And both would agree with the idea that the stream of consciousness is not what our inner experience reveals, although by pointing to different reasons. Furthermore, both would hold the view that there is much less to mind and memory then what our current experience mistakenly leads us to suppose.

All in all, Gazzaniga's book provides enjoyable, enlightening and provocative reading. It is a book whose ultimate goal is to rescue the science of mind from misleading propositions by showing that there is no need to explain thought and action as the outcome of an inscrutable self encapsulated in a shell.

Acknowledgements
The author thanks John Symons for helpful suggestions and FAPESP for financial support (grant # 97-03518-6).

Steven Pinker, *How the Mind Works*
1997, Allen Lane: The Penguin Press, ISBN: 0 -713-991305 pp. 660

Jill Boucher
Dept. of Human Communication Sciences
University of Sheffield

At the outset of this book Pinker is concerned to establish that the mind operates as a modular computational mechanism, and that it is symbol-based (mentalese) information processing which makes humans intelligent. He is also concerned to establish the credentials of Darwinism, and to see off the various criticisms which have been voiced about it. Having established the two main planks supporting his view of how the mind works, Pinker spends the rest of the book arguing that what natural selection wrote into the genes of our remote ancestors can be detected in almost all aspects of human behaviour - visual, cognitive (via a legacy of intuitive theories about the world), emotional, social, and cultural. He also considers sentience, morality and free will, but - somewhat unexpectedly - blinks and draws back, describing them as mysteries beyond our present computational capacities to understand. Thus, the book is not about how the mind evolved, but rather about how the genetic products of natural selection determine the way the mind works, and the evidence for this in contemporary human behaviour.

Like *The Language Instinct* the book is designed to appeal to non-specialist as well as to specialist readers, and the same methods of making some quite difficult material palateable are used. For example, there are no dated references in the text, Pinker's sources being given in 20 pages of notes at the end of the book, prior to the actual references. The book is written in a style designed to be informal and entertaining, and includes colloquialisms, anecdotes, and jokey chapter/section headings.

Despite the sugar coating - partly because of it, in fact - I found the book irksome to read. There were a number of reasons for this, apart from the sugariness. In particular, I dislike the author's preferred method of setting up his own arguments by ridiculing or belittling other people's theories or work. When no real targets are around, he sets up straw men in order to knock them down, resuscitating hoary theories or popular myths in order to turn his scorn on them. With this goes a lack of academic objectivity, which is replaced by a seeming desire to score points. The book is also hard going because of its length, which is excessive, and also because of the lack of explicit structuring of the text. The jokey chapter and section headings give little indication of the content which follows, and there is not much explicit signposting or summarising within the text itself. There are unheaded sub-sections within the headed sections, and some of these subsections have only tangential relationships with what has gone before and what comes after. There are no short cuts to extracting the meat of the book: you either read it or you don't, and that is a disincentive for those people who might be interested in what Pinker has to say but don't have unlimited time for reading. The quality of the contents is also uneven, the chapters on computation and vision being detailed and informative, whereas the chapter on social psychology, for example, was much less convincing. Nevertheless, there is a lot of good scholarly stuff in the book, and alongside his arrogant self-assertivenes Pinker displays a disarming moral seriousness which helps to keep the reader plodding on.

At the end, was it worth it? Well, the research which has gone into the book is clearly considerable, it is well indexed and referenced, giving it some value as a source book. And those parts of the book which are factual and informative are very good indeed. If you are comfortable with the Standard Cognitive Science Model of the mind, and with Tooby and Cosmides' incorporation of this model into evolutionary psychology, you may even think that this is a great book. However, for readers like me who feel uneasy with the model and who seek a rapprochement between a strictly Fodorian view of the mind and a view of the mind as emergent and dynamic, the central arguments are stale and say nothing new. Some of the critical arguments, such as the argument for nonlinguistic symbols, are difficult to follow and seem to me to be incoherent. This is frustrating because Pinker is undeniably capable of moving ideas on

and does so elsewhere. He did not seem concerned to do that here, no doubt because he was more concerned to write a popular book in which certain controversial hypotheses could be purveyed forcefully and entertainingly to a perhaps not too critical readership.

You have to be a convinced Darwinist, too, in order to feel that the main endeavour of the book is worthwhile - (namely to link contemporary human behaviour back to traits selected in prehistory). I found Pinker's defence of Darwinism persuasive and illuminating, unlike his defence of the symbolic mind, but then I don't need much persuading here.

My overall conclusion is, therefore, that if you share Pinker's two central premises, you will tolerate the weaker sections and thoroughly enjoy the meaty bits of this book: you may even enjoy the rubbishing of other academics, with the smug feeling that you are on the side of the angels. If you don't share Pinker's premises you may feel alienated and irritated, as I did - not so much because you disagree with what is being said but because of the way in which it is said.