

This paper was published in Peter Carruthers and Andrew Chamberlain, eds., *Evolution and the Human Mind: Modularity, Language and Meta-Cognition*, (Cambridge: Cambridge University Press) 2000. Pp. 62-92.

# **DARWIN IN THE MADHOUSE: EVOLUTIONARY PSYCHOLOGY AND THE CLASSIFICATION OF MENTAL DISORDERS**

by

**Dominic Murphy**

**California Institute of Technology**

and

**Stephen Stich**

**Rutgers University**

Recent years have witnessed a ground swell of interest in the application of evolutionary theory to issues in psychopathology (Nesse & Williams 1995, Stevens & Price 1996, McGuire & Troisi 1998). Much of this work has been aimed at finding adaptationist explanations for a variety of mental disorders ranging from phobias to depression to schizophrenia. There has, however, been relatively little discussion of the implications that the theories proposed by evolutionary psychologists might have for the classification of mental disorders. This is the theme we propose to explore. We'll begin, in Section 1, by providing a brief overview of the account of the mind advanced by evolutionary psychologists. In Section 2 we'll explain why issues of taxonomy are important and why the dominant approach to the classification of mental disorders is radically and alarmingly unsatisfactory. We will also indicate why we think an alternative approach, based on theories in evolutionary psychology, is particularly promising. In Section 3 we'll try to illustrate some of the virtues of the evolutionary psychological approach to classification. The discussion in Section 3 will highlight a quite fundamental distinction between those disorders that arise from the malfunction of a component of the mind and those that can be traced to the fact that our minds must now function in environments that are very different from the environments in which they evolved. This mis-match between the current and ancestral environments can, we maintain, give rise to serious mental disorders despite the fact that, in one important sense, there is nothing at all wrong with the people suffering the disorder. Their minds are functioning exactly as Mother Nature intended them to. In Section 4, we'll give a brief overview of some of the ways in which the sorts of malfunctions catalogued in Section 3 might arise, and sketch two rather different strategies for incorporating this etiologically information in a system for classifying mental disorders. Finally, in Section 5, we will explain why an evolutionary approach may lead to a quite radical revision in the classification of certain

conditions. From an evolutionary perspective, we will argue, some of the disorders recognized in standard manuals like DSM-IV may turn out not to be disorders at all. The people who have these conditions don't *have* problems; they just *cause* problems!

## 1 The Evolutionary Psychology Model of the Mind

The model of the mind advanced by evolutionary psychology is built around two central theses which we'll call *The Massive Modularity Hypothesis* and *The Adaptation Hypothesis*. The Massive Modularity Hypothesis maintains that the mind contains a large number of distinct though interconnected information processing systems -- often called "modules" or "mental organs." These modules can be thought of as special purpose or domain specific computational mechanisms. Often a module will have proprietary access to a body of information that is useful in dealing with its domain. The information is "proprietary" in the sense that other modules and non-modular mental mechanisms have no direct access to it.<sup>1</sup> Like other organs, modules are assumed to be innate and (with the possible exception of a few gender specific modules) they are present in all normal members of the species. Some evolutionary psychologists also assume that there is little or no heritable inter-personal variation in properly functioning mental modules and thus that a given type of module will be much the same in all normal people.<sup>2</sup> Paul Griffiths has dubbed this "the doctrine of the monomorphic mind." Both Griffiths and David Sloan Wilson have argued, in our opinion quite persuasively, that this doctrine is very implausible. (Wilson, 1994; Griffiths, 1997, Sec. 5.5. The point goes back to David Hull, 1989.) So, along with Wilson and Griffiths, we will assume that there may be a fair amount of heritable variation in the modules found in the normal population. That assumption will play an important role in Section 6, where we argue that some of the conditions that have been classified as mental disorders are not disorders at all.

Since the appearance, in 1983, of Jerry Fodor's enormously influential book, *The Modularity of Mind*, the term 'module' has become ubiquitous in the cognitive sciences. But the sorts of modules posited by the Massive Modularity Hypothesis differ from Fodorian modules in two crucial respects. First, Fodor sets out a substantial list of features that are characteristic of modules, and to count as a Fodorian module a mental mechanism must have most or all of these features to a significant degree. For Fodor, modules are:

- i. informationally encapsulated
- ii. mandatory
- iii. fast
- iv. shallow

---

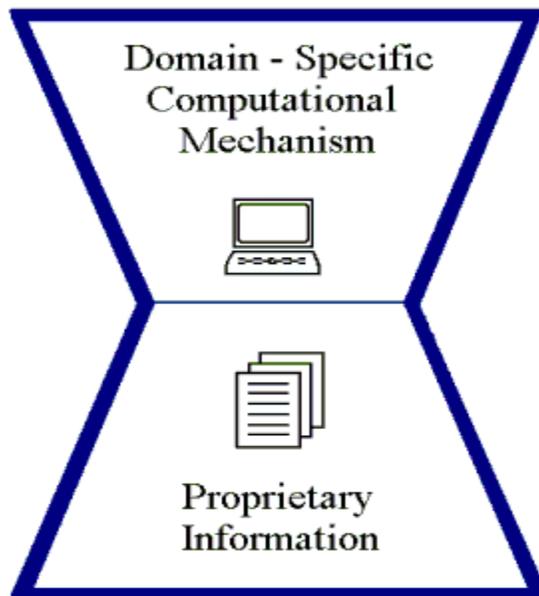
<sup>1</sup> For a much more detailed discussion of the Massive Modularity Hypothesis see Samuels (1998 and this volume).

<sup>2</sup> Tooby and Cosmides, who are among the leading advocates of evolutionary psychology, defend this "psychic unity of mankind" in numerous places including Tooby and Cosmides, 1990a, 1990b and 1992.

- v. neurally localized
- vi. susceptible to characteristic breakdown
- vii. largely inaccessible to other processes.

The notion of a module invoked in the Massive Modularity Hypothesis is much broader and less demanding. Evolutionary psychologists count as a module any domain specific computational device that exhibits (i) and (vii), and occasionally even these restrictions are not imposed. The second important way in which Fodorian modules differ from the sorts of modules envisioned by the Massive Modularity Hypothesis is that, for Fodor, modules only subserve “peripheral” mental processes – those responsible for perception, language processing and the production of bodily movements. Evolutionary psychologists, by contrast, expect to find modules subserving a wide range of other, more “central” cognitive and emotional processes.

The Adaptation Hypothesis, the second central theme in evolutionary psychology, claims that mental modules are *adaptations* -- they were, as Tooby and Cosmides have put it, “invented by natural selection during the species’ evolutionary history to produce adaptive ends in the species’ natural environment.” (Tooby and Cosmides, 1995, p. xiii) To serve as reminder of the fact that the modules posited by evolutionary psychology are adaptations, and to distinguish them from Fodorian modules, we will sometimes call them *Darwinian modules*.



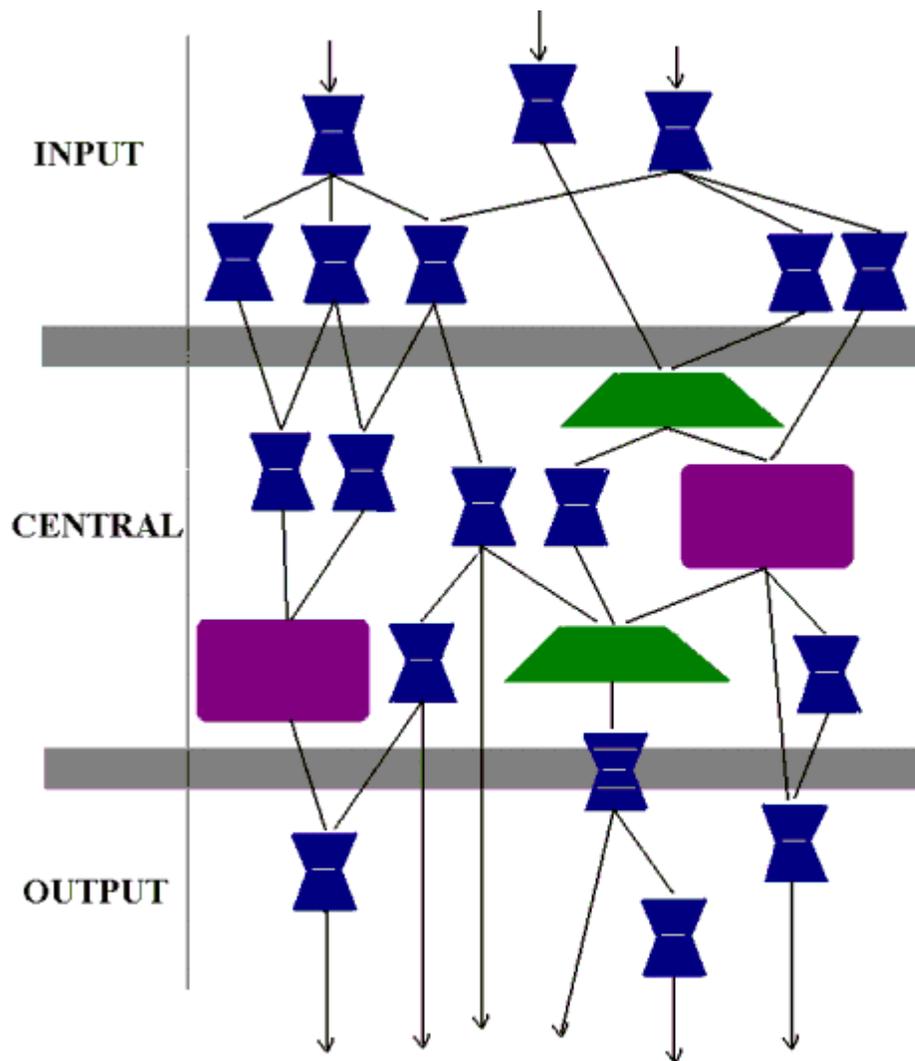
**Figure 1: A Darwinian Module**

**Darwinian Modules are adaptations that can be thought of as special purpose or domain specific computational devices which often have proprietary access to a body of information that is useful in dealing with their domain.**

The picture of the mind that emerges from the conjunction of the Massive Modularity Hypothesis and the Adaptation Hypothesis is nicely captured by Tooby and Cosmides in the following passage:

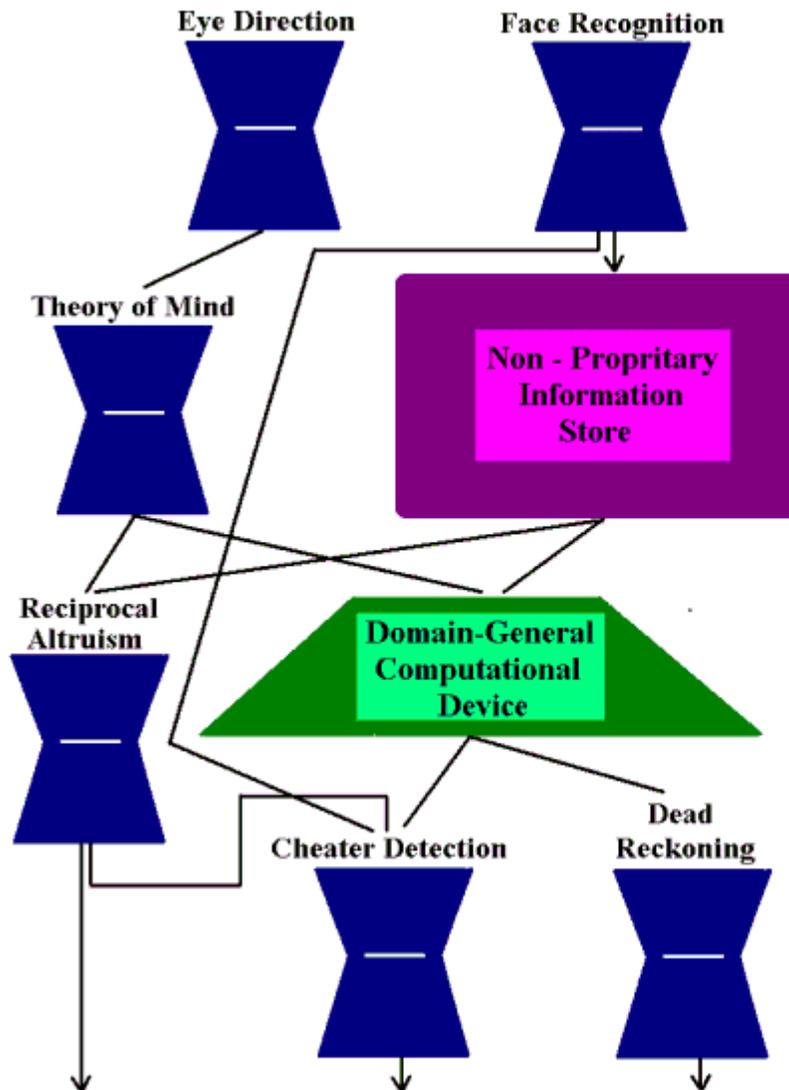
[O]ur cognitive architecture resembles a confederation of hundreds or thousands of functionally dedicated computers (often called modules) designed to solve adaptive problems endemic to our hunter-gatherer ancestors. Each of these devices has its own agenda and imposes its own exotic organization on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

There are two points that we would add to this colorful account. First, these functionally dedicated computers are linked together in complex networks. The output of one module will often serve as the input (or part of the input) for one or more modules that are “downstream.” Second, there is no reason to suppose that *all* of the mechanisms to be found in the mind are plausibly viewed as modular. In addition to the swarm of modules, the evolutionary psychology model of the mind can accommodate computational devices that are not domain specific, stores of information that are not proprietary, and a variety of other sorts of mechanisms. Figure 2 is a sketch of the sort of mental architecture posited by evolutionary psychology. Figure 3 is a close-up of part of the system portrayed in Figure 2.



**Figure 2**

**The mental architecture posited by evolutionary psychology includes networks of Darwinian Modules subserving central cognitive and emotive process as well as peripheral processes.**



**FIGURE 3**

In addition to Darwinian Modules, the evolutionary psychology model of the mind can accommodate computational mechanisms that are not domain specific and stores of information that are not proprietary.

## 2 The Taxonomy Crisis: What's Wrong with the DSM Approach, and Why Taxonomy Matters

In 1964, Carl Hempel thought it very likely “that classifications of mental disorders will increasingly reflect theoretical considerations” (1964; 150). Hempel was a

first-class philosopher but an unreliable prophet; the last thirty years have seen the old psychoanalytically based paradigm replaced by an approach to classification which aims to be “operationalized,” “atheoretical” and “purely descriptive.”<sup>3</sup> Among the most notable products of this approach are DSM-III and its successors DSM-III-R and DSM-IV (American Psychiatric Association, 1980, 1987, 1994). DSM categories are typically specified by providing a list of sufficient conditions (often disjunctive and with an occasional necessary condition thrown in) stated almost exclusively in the language of “clinical phenomenology” which draws heavily on folk psychological concepts and protoscientific clinical concepts (like self-esteem, delusion, anxiety and depressed mood). The classification systems set out in DSM-III and its successors play a central role in guiding research and clinical practice in the United States and, to a lesser extent, in other countries as well. Moreover, as Poland, Von Eckardt and Spaulding (1994, p. 235) note, “DSM categories play pivotal roles in financing mental health care, maintaining hospital and clinical records, administering research funds, and organizing educational materials ... concerned with psychopathology.” Poland and his colleagues go on to claim -- and we agree -- that the DSM approach “is deeply flawed and not doing the work it should be doing” and that as a result, the current situation regarding the classification of mental disorders “involves a crisis.” (255)

According to Poland et al., a classification scheme in psychopathology has two primary purposes. It should enhance the effectiveness of clinical activity, and it should facilitate scientific research concerned with psychopathology and its treatment. The DSM approach, they argue, does neither. Though Poland and his colleagues offer a number of reasons for their deeply skeptical and rather disquieting conclusion, we think that one of these is central. The DSM approach is alleged to be atheoretical, thus allowing clinicians from different theoretical backgrounds to agree on a diagnosis. But in fact the DSM approach is far from atheoretical. Rather, it embraces the highly problematic theory that there exist, in the domain of psychopathology, a substantial number of what Poland et al. call “syndromes with unity.” These are clusters of associated attributes, characterized in the folk psychological and protoscientific language of clinical phenomenology, “that exhibit such dynamic characteristics as typical course, outcome and responsiveness to treatment, and that are related to underlying pathological conditions and etiological factors of development (e.g. genetic and environmental factors).... [T]he operationally defined categories within the DSM system are supposed to be *natural kinds* with a characteristic causal structure (i.e. a core pathology) that underwrites the various lawful regularities characteristic of the disorder (e.g. association of criterial features, dynamic properties of the syndrome).”<sup>4</sup> (241)

---

<sup>3</sup> McCarthy & Gerring, 1994 is a good brief history.

<sup>4</sup> Some advocates of the DSM approach would grant that many currently recognized DSM classifications fail to pick out natural kinds, though they expect that with more research, based on existing assumptions, operationally defined DSM-style diagnoses will converge on natural kinds. The basic “syndromes with unity” theory is assumed to be correct and in need of more empirical elaboration, rather than conceptual overhaul. See, for example, Goodwin & Guze, 1995.

The problem with all of this is very simple: The theory is false. Though there may be a few syndromes with unity in psychopathology, it is unlikely that there are very many. One reason for this is the all but exclusive reliance on the concepts and categories of clinical phenomenology to characterize syndromes. These concepts are notoriously vague, imprecise and unquantified. Moreover, since highly subjective judgments about their application to a particular case must typically be made in emotionally charged settings in which hidden agendas abound, these judgments are often biased. By limiting the data gathered in diagnosis to the salient and easily identifiable signs and symptoms of clinical phenomenology, the DSM scheme fails to attend to a wide range of other data about mental functioning that can be gathered by psychometric techniques and by methods used in cognitive science and neuroscience.

Another, closely related, reason to think that there are relatively few psychopathological syndromes with unity is the fact that the DSM approach to classification is not guided by any theory about the structure and functioning of normal minds and makes no attempt to uncover and use facts about the underlying psychological, biological and environmental mechanisms and processes that give rise to symptoms. Imagine for a moment trying to construct a classification system for malfunctions in some complex and well engineered artifact which is built from numerous carefully designed components – a television set, perhaps, or a computer network. Imagine further that the classification system must be based entirely on clusters of user-salient symptoms, without any inkling of how the mechanism was designed to operate and without any theory about its component parts and their functions. The result, almost certainly, would be set of categories that are *massively heterogeneous* from the point of view of someone who understands how the system works. It would classify together problems which are caused by totally different underlying mechanisms or processes and which require totally different remedies. It would also fail to classify together problems with the same underlying cause (and requiring the same remedy) if they manifest themselves in different ways under slightly different conditions. This, near enough, is just what we should expect in a DSM-style classification of mental disorders. For surely the mind, too, is a complex and well engineered system in which many well designed components interact.<sup>5</sup> Thus, as Poland and his colleagues conclude:

It appears *unlikely* that the domain of psychopathology is best conceived of in terms of syndromes with unity or that natural kinds will be discovered at the level of clinical phenomenology. There is simply no reason to suppose that the features of clinical phenomenology that catch our attention and are the source of

---

<sup>5</sup> This heterogeneity in DSM classifications is magnified by the fact that actual DSM categories often group together very different symptom profiles as manifestations of the same disorder. Thus, for example, according to the DSM criteria, one can qualify as having a Major Depressive Episode even though one does not experience a depressed mood, provided that one does exhibit a markedly diminished interest in daily activities. (DSM-IV, p. 327). And by one reckoning there are 56 different ways to satisfy the criteria for Borderline Personality Disorder! (Clarkin et al., 1983) As one might expect, these heterogeneous categories are poor predictors of the patients' future trajectory or of their response to treatment and thus the vast majority of DSM categories remain "unvalidated".

great human distress are also features upon which a science of psychopathology should directly focus when searching for regularities and natural kinds. Human interests and salencies tend to carve out an unnatural domain from the point of view of nomological structure. Hence the relations between the scientific understanding of psychopathology and clinical responsiveness to it may be less direct than is commonly supposed. In insisting that classification be exclusively focused on clinical phenomenology, DSM not only undermines productive research but also undermines the development of effective relations between clinical practice and scientific understanding. (254)

The remedy that Poland et al. propose is one that we strongly endorse. There is a need for a new approach to the classification of mental disorders that is “based on a more intimate relationship with basic science than is DSM.” (255) In trying to construct this new taxonomy, a natural first step is to ask: *Which science or sciences are the appropriate ones?* We don’t think there is any single right answer to this question. Many sciences can contribute to the construction of a taxonomy that will serve the needs of clinical practice and scientific research into the causes and treatments of mental disorders. But it is our contention that evolutionary psychology has a natural and quite central role to play in this scientifically based reconstruction of the classification system for mental disorders. Evolutionary psychology, as we have seen, seeks to explain how the mind works by characterizing the many computational mechanisms from which it is constructed and attempting to discover the function for which these mechanisms were designed. That sort of account of the mind and its working looks to be just what is needed if we are to take seriously the analogy with a malfunctioning well engineered artifact. Of course evolutionary psychology is not alone in viewing the mind as made up of lots of components that were designed by natural selection. Neuroscience, at various levels of analysis from the molecular to the computational takes much the same view. And while we certainly don’t want to deny that these sciences will be of enormous importance in working toward a new taxonomy, it’s our prediction that, in the short run at least, classifications based on evolutionary psychological theories will be particularly useful for clinicians, since they will be at a level of analysis that meshes comfortably with current clinical practice.

Our goal in the remainder of this paper is to make this prediction plausible. To do this we propose to explore some of the problems that might befall a mind that is structured in the way sketched in Section 1 and that contains some of the mechanisms posited by evolutionary psychologists. In many cases, as we shall see, those problems offer plausible explanations of the sorts of troubling symptoms that manuals like DSM-III and its successors take to indicate the presence of a mental disorder. However, it will often be the case that the classification suggested by evolutionary psychological theories recognizes several distinct disorders where current diagnostic manuals see only one. Thus there is reason to hope that a classification system that takes account of theories in evolutionary psychology will begin to reduce the massive heterogeneity that plagues DSM-style classifications. Another virtue of taxonomizing disorders along the lines suggested by evolutionary psychology is that it pulls apart two very different sorts of disorders: Those in which components of the mind are malfunctioning and

those attributable to a mis-match between the environment in which we live and the environment in which we were designed to live. A third virtue of the evolutionary psychological approach is that it provides a clear theoretical framework in which we can ask one of the most vexing questions that the study of psychopathology must face: What conditions count as disorders at all?<sup>6</sup>

Before setting out our taxonomic proposals we should stress that evolutionary psychology is still very much in its infancy and the theories about mental mechanisms that we will invoke are all both speculative and controversial. We don't pretend to be offering a set of diagnostic categories that mental health professionals might use in preference to those in DSM-IV. Rather, our aim is to begin to explore the ways in which evolutionary psychology can contribute to the elaboration of a taxonomy of the sort that Poland et al. advocate – one that is “based on a more intimate relationship with basic science.”

### 3 A Taxonomy of Disordered Minds

The range of symptoms recognized by modern diagnoses is very broad. To begin with there are cognitive symptoms with highly salient phenomenologies, such as delusions and unwelcome or obsessive thoughts. There are also feelings of “thought disorder,” in which patients report thinking someone else's thoughts or having their own thoughts controlled by another. Other cognitive symptoms include such incapacities as the amnesias, agnosias and aphasias. Then we have behavioral problems, including voluntary patterns of antisocial action and involuntary problems which include drug dependence, motor retardation, sleep disorders and disruptions to the autonomic nervous system like irregular heartbeat. There are also more intuitively qualitative symptoms; some of these are relatively prolonged, such as low affect (“feeling blue”), and others are transitory, such as dizziness, nausea and feelings of anxiety. So there are a great many kinds of symptoms to explain. The exciting thing about evolutionary psychology is the theoretically motivated range of explanatory resources it brings to bear on all this diverse symptomatology.

The evolutionary perspective enables us to make a number of important distinctions among problems that may lead to symptoms of mental disorder. The most theoretically interesting and novel of these is the distinction between problems which are internal to the person and problems which lie in the environment surrounding the person. This marks the first major break in our taxonomy.

---

<sup>6</sup> For reasons that we'll set out in Section 6, we are inclined to think that this is best understood as a *pair* of questions, viz. (i) What conditions count as *mental disorders*? and (ii) What conditions count as *problems that may beset an evolved mind* (or “*E-M problems*” as we'll sometimes say)? The category of mental disorders, we'll argue, is a subset of the category of E-M problems. The taxonomy that we are about to sketch, in Section 4, is should be read as an account of the broader category – the category of E-M problems, though for ease of exposition we propose not to emphasize the distinction until Section 6.

Problems which are internal to the person are what we commonly think of when we envisage mental disorders. The official orthodoxy, enshrined in DSM-IV, views mental illness as Janus-faced, with socially disvalued or disabling symptoms being produced by an underlying *malfunction*. (The extent to which this conception is honored in the discussion of particular disorders is another matter (Wakefield 1997).) However, it is important to recall that the evolutionary perspective on the mind stresses that our psychological mechanisms originated in a past environment, and although those mechanisms may have been adaptive in that past environment, it is entirely possible that the environment has changed enough to render aspects of our cognitive architecture undesirable or obsolete in the modern world. We will discuss this in more detail below. To begin with, though, we'll focus on cases of disorders which are internal to the person.

### 3.1 *Disorders Within the Person*

As we've seen, the evolutionary psychology model recognizes several different sorts of mental structures – modules, stores of non-proprietary information, computational devices that are not domain specific, and pathways along which information can flow from one mechanism to another – and since all of these can break down in various ways, the model will admit of a number of different sorts of disorders. However, since Darwinian modules are the most prominent structures in the evolutionary psychology model it is natural to begin our taxonomy of disordered minds with them. The most obvious sort of difficulty that can beset a mind like the one depicted in Figures 2 and 3 is that one of the modules can behave problematically, producing output which directly or indirectly leads to the symptoms on which diagnoses of mental disorder depend.

There are two very different reasons why a Darwinian module may produce such symptoms, and this distinction generates a first major divide in within-person cases. Sometimes when a Darwinian module generates problematic output the trouble is *internal* to the module -- its special purpose computer is malfunctioning or its proprietary store of information is not what it should be (or both). In other cases the problem will be *external* to the module. In these cases something has gone amiss earlier on in the causal network and “upstream” in the flow of information, with the result that the module which is producing problematic output is being given *problematic input*. In the colorful language of computer programmers, “garbage in, garbage out.”

#### 3.1.1 *Disorders Resulting from Module-Internal Problems: Some Examples*

Perhaps the best known example of a disorder which has been much studied as a case of modular breakdown is autism (Baron-Cohen, Leslie and Frith, 1985; Frith, 1989; Leslie, 1987, 1991; Leslie and Thaiss, 1992; Baron-Cohen, 1995). Recent work has suggested that autism is best explained as a breakdown in the module or system of modules that handle “theory-of-mind,” the capacity of all normal adults to attribute intentional states like beliefs and desires to other people and to explain their behavior in

terms of the causal powers of beliefs and desires. One widely used test of whether a person has a normal adult theory-of-mind-module is the ability to pass the false-belief task, at which autistic children are spectacular failures.<sup>7</sup> They do worse at the false-belief task than do children with Down's Syndrome, even though in general their grasp of causal cognition exceeds the latter's (Baron-Cohen et al., 1986). Some people diagnosed with Asperger's Syndrome -- high functioning autistics whose IQs are normal or higher -- have offered quite moving accounts of their puzzlement when they realized how much more normal people seemed to know about what others were thinking in social situations. One example, made famous by Oliver Sachs, is Temple Grandin's comment that her social experience in adolescence was like being "an anthropologist on Mars." (Sachs, 1995; Frith, 1989; Grandin & Scariano, 1986)

A similar explanation in terms of a broken module occurs in Blair's discussion of psychopathy. Three core features in the characterization of psychopathy are (i) early onset of extremely aggressive behavior, (ii) absence of remorse or guilt, and (iii) callousness and a lack of empathy. Blair (1995) explains psychopathic behavior as due to the absence or malfunctioning of a module which he calls the *violence inhibition mechanism* (VIM). The central idea was borrowed from ethology, where research had long suggested the existence of a mechanism which ended fights in response to a display of submission. A well-known example is the canine tendency to bare the throat when attacked by a stronger conspecific. The assailant then ceases the attack, rather than taking advantage of the opportunity to press it home. Blair hypothesizes that a similar mechanism exists in humans, activated by the perception of distress in others. When the VIM is activated it causes a withdrawal response which people experience as aversive. Following Mandler, Blair suggests that this aversive experience is one of the building blocks for such moral emotions as guilt and remorse. On Blair's account, the VIM acquires new triggers via classical conditioning. Since engaging in aggressive activity will often lead the victim to exhibit distress cues, aggressive activity becomes a conditioned stimulus for the aversive response. Distress cues are also typically paired with the construction a mental representation of the victim's suffering, and as a result these thoughts also become triggers for the VIM. This linkage, Blair maintains, is a crucial step in the development of empathy. Since psychopaths do not have a properly functioning VIM, they do not experience the effects of their violence on others as aversive and this explains why psychopathy is associated with an increase in violent tendencies at an early age. Their deficit does not lead psychopaths to become aggressive, but when they do, they are much less inclined to stop. Blair's model also explains why psychopaths fail to develop the moral emotions and fail to experience any empathic response to the suffering of others. The most intriguing part of Blair's theory is his argument that people lacking a properly functioning VIM would not be able to

---

<sup>7</sup> False belief tasks are intended to evaluate whether or not experimental subjects understand when someone might hold a false belief. One standard version of the task --sometimes called the "Sally-Ann Task"-- involves watching Sally put a piece of chocolate in one place (location A) and later, while Sally is away, Anne moving the chocolate elsewhere (location B). The subject is then asked "Where will Sally look for her chocolate?". In order to answer this question correctly, the subject needs to appreciate that, since Sally was absent when her chocolate was moved from A to B, she will have the false believe that it is at A. (Baron-Cohen, 1995, p.70)

recognize the distinction between moral transgressions which cause other people to suffer and other social transgressions which do not. This prediction was confirmed in a study comparing the moral cognition of psychopathic murderers with the moral cognition of murderers who were not diagnosed as psychopaths.

Since the publication of Robert Trivers' seminal paper on reciprocal altruism (Trivers, 1971) the capacity to engage in reciprocal exchanges has played an important role in the thinking of sociobiologists and evolutionary psychologists. More recently a number of theorists, including Cosmides, Tooby and Gigerenzer have argued that this capacity is subserved by a module or a cluster of modules designed to compute what is and is not required in reciprocal exchange arrangements and to detect "cheaters" who fail to reciprocate. (Cosmides, 1989; Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992) If the module that computes what is required in reciprocal altruism malfunctions, the likely result will be that the module's owner will systematically misunderstand what is expected in cooperative behavior and reciprocal exchanges. Such a person might regularly over-estimate the value or importance of his own contribution in a reciprocal relationship and/or regularly under-estimate the value or importance of the other party's contribution.<sup>8</sup> From the point of view of the person with a malfunctioning reciprocal altruism module (though not from the point of those he interacts with) he would be regularly exploited or cheated in social exchanges, and this might well lead him to avoid social interaction and to be in a depressed mood for extended periods.

In an important series of publications, McGuire and his colleagues have argued that this sort of malfunction may be a central factor in many individuals who fit the DSM criteria for dysthymia which is an affective disorder characterized by persistent depressed mood for over two years, but without major depressive or manic episodes. In one study, McGuire and his colleagues found that dysthymic patients had a notable deficit in their ability to achieve social goals and carry out simple social tasks. They tended to blame others for their dissatisfactions, rather than considering their own behavior (as did a matched control group.) Dysthymic patients were also less likely than controls to interact socially with others. Perhaps the most striking finding of the study was that dysthymic subjects "believed that they helped others *significantly more* than they were helped by others. Thus, by their own reckoning, they were cooperators." However, "a detailed analysis of their social interactions, which involved collecting data from siblings or friends, strongly suggested otherwise." Subjects with dysthymic disorder "not only tended to exaggerate their helpfulness to others, but they also downplayed the value of others' help.... In addition, they were skeptical of others' intentions to help as well as to reciprocate helping that [they] might provide. For the majority of [dysthymic subjects], these views began *prior* to adolescence...." (p. 317)

A defective module (or "algorithm") for computing what is expected or required in reciprocal relations is not the only sort of defect that might lead dysthymic persons to

---

<sup>8</sup> Of course, the mere fact that the reciprocal altruism module malfunctions does not entail that a person will over-value his own contribution and under-value the other party's. Various other patterns are possible. And if the first pattern is typical, some further explanation is needed for this fact.

exaggerate their own helpfulness and downplay the helpfulness of others. Though the basic principles of reciprocal exchange may be universal, the value of specific acts varies enormously from culture to culture. In our culture giving your neighbor a hot tip on a stock counts as a valuable favor, while paying a shaman to chant secret prayers for his child who is down with the flu does not. In other cultures this pattern is reversed. A person who had failed to master the local culture's value system might well end up thinking that he helped others vastly more than they helped him. It is plausible that information about the value that one's culture assigns to various actions is not proprietary to any given module, but is stored in a location to which many mental mechanisms have access. If that is right, then dysthymia may be a heterogeneous category since the tendency of dysthymic people to misunderstand reciprocity relationships might have two quite different causes. This suggests an intriguing hypothesis. Suppose that some people diagnosed as dysthymic have defective reciprocal altruism modules while others have normal modules and have simply failed to master the prevailing principles of social value. If so, it might well be the case that this latter group, but not the former, could be treated effectively by a regimen of cognitive psychotherapy that sought to inculcate the social codes they have failed to internalize.

In the preceding cases we have focused on modules for whose existence we have some independent empirical or theoretical support. Some of the symptoms that characterize the disorders we've considered are those we would expect when these modules malfunction. Indeed, in the case of autism the clinical data have been taken to provide important additional support for the hypothesis that a theory of mind module exists. Especially noteworthy in this connection is the double-dissociation evidence provided by studies of Williams' Syndrome patients. (Karmiloff-Smith, et al., 1995; Bellugi, Klima, & Wang, 1997) For the reasons set out in Section 3, it is probably unwise to expect the fit between hypothetical mechanisms and currently recognized symptomatology to be too exact. In some cases -- autism is one -- the hypothesized modular deficit does not generate all the symptoms which current clinical thinking takes to characterize the disorder. In these cases there are at least the following two possibilities when it comes to mapping the diagnosis onto the architecture. In many cases the full suite of recognized symptoms is not necessary for the diagnosis at all - a subset of the recognized symptoms will do. When it comes to relating the current diagnosis to our mental architecture, we can isolate a broken module which might explain a subset of the symptomatology which is sufficient for diagnosis. Further clusters of symptoms could be due to other causes. In such cases it is possible that what we are dealing with is actually several disorders, represented by the different sets of symptoms which are currently thought to be variant forms of one disorder. This is the first possibility we have in mind. These different conditions might co-occur due to a common cause which disrupts several mechanisms. The second possibility is that this co-occurrence is more coincidental, and that the different disorders sometimes just happen to occur together. Since the DSM categories are not validated, it probably happens quite often that DSM picks out symptom clusters that are not in fact all that reliably linked with one another. It is even possible that the cause of some of the unexplained symptoms may not be a disorder at all. They may, for example, just represent the stress of being in treatment for a different condition, or be responses to

what are termed “problems in living”. Evolutionary psychology offers a model of the mind which allows us to disentangle one set of symptoms from the wider collection and recognize it as a distinct condition.

Though the modules that have played a role in our discussion thus far are ones which we have non-clinical reasons for recognizing, there have been cases, especially in the neuropsychological literature, in which the discovery of particular deficits has led investigators to argue for the presence of specialized systems or modules in the mental architecture. For example, dorsal simultanagnosics can recognize the spatial relations among parts of an object but are unable to compute the spatial relations between objects. This suggests that there are separate systems underlying these two forms of spatial perception (Farah, 1990). It is noteworthy that in this case the symptoms that led to an hypothesis about the underlying mechanism are not among the standard items of clinical phenomenology that loom so large in DSM-III and its successors. Indeed, for a variety of historical and practical reasons, the agnosias, amnesias and aphasias are not even in DSM-IV as conditions, although some of their characteristic symptoms are.

### *3.1.2 Disorders Resulting from Upstream Problems in the Cognitive System*

As we noted earlier, when a module behaves problematically, there can be two very different sorts of reasons. In some cases, the module itself is to blame. In other cases, the trouble is further upstream. Many modules receive input from other modules, so it will often be the case that if an upstream module is malfunctioning one or more of the modules to which it is supposed to provide information will also produce output that yields symptoms of mental disorder. If the broken upstream module provides information to several separate downstream systems, an upstream problem can result in several quite different clusters of symptoms. (Figure 4) The possibility that a single malfunctioning module may cause several other modules to produce problematic outputs may provide a partial explanation for the very high rate of comorbidity that is found in psychiatric patients. A lot of people have more than one disorder at the same time. The National Comorbidity Survey concluded that “more than half of all lifetime disorders occurred in the 14% of the population who had a history of three or more comorbid disorders.” (Kessler et al. 1994)<sup>9</sup>

---

<sup>9</sup> It is worth stressing that this is only one sort of explanation for comorbidity, even at the architectural level. For example, if there are domain-general systems then if these systems are damaged we might get a general reduction of functioning which causes problems in several areas.

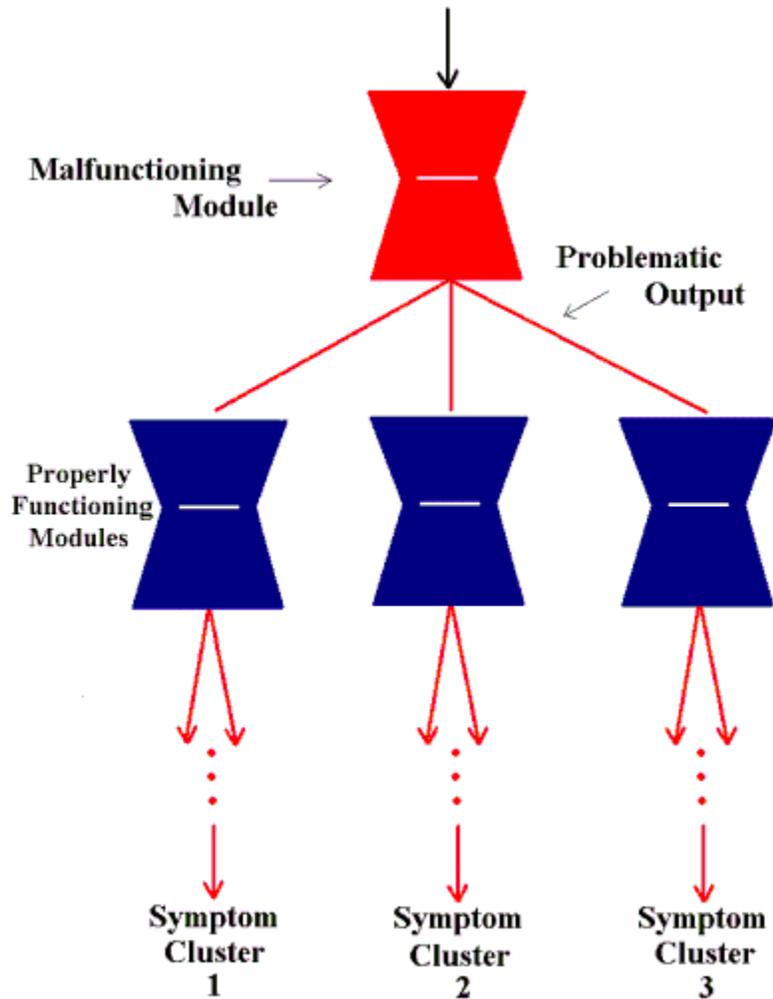


Figure 4

**When a malfunctioning module provides information to several separate downstream systems, the upstream problem can result in several quite distinct clusters of symptoms.**

Our picture does not mandate that if one module feeds information to another it must always be the case that the second will produce problematic output as a result of a breakdown in the first. There are typically several ways in which a single module can malfunction, and each of these is a matter of degree. So it may happen that a malfunction upstream produces problems in a second module in some people, while in other people it does not. All of this can be iterated for modules which are further downstream. The result is that the profile for a specific patient is likely to be quite complex.

This suggests a way of classifying mental disorders which are internal to the subject. The idea would be that such disorders are to be identified with a chain or network of modules each of which is producing problematic outputs. Those outputs in turn are responsible for a suite of symptoms that is characteristic of the disorder. The canonical specification of a disorder would also include, for each module in the network, an indication of whether it is itself broken or whether it is receiving tainted input from elsewhere in the network. This cuts things up rather finely, but it does allow for some important theoretical distinctions. Suppose that two modules are each delivering problematic outputs, but that only the first is actually malfunctioning. The second is producing problematic output because it is downstream from the first and is being provided with problematic input. In this case the solution to the problem lies in repairing the upstream module that is the source of the trouble. However, if both modules are malfunctioning then both will have to be repaired if the disorder is to be dealt with. Merely noting that in each case the two modules form the network underlying the disorder is insufficient to direct therapeutic interventions. (Figure 5)

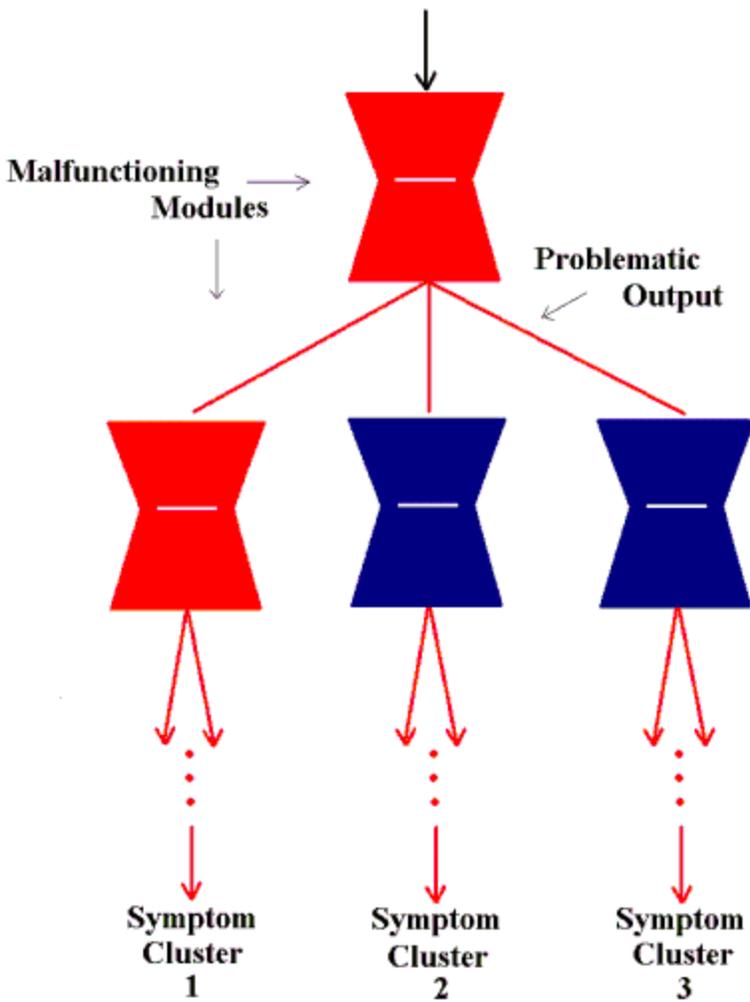


Figure 5

An example of a disorder which may be caused by problems upstream in the flow of information is the Capgras delusion. Patients with Capgras believe that someone close to them - typically a spouse - has been replaced by an exact replica. Recent work suggests that part of the explanation for the delusion is that the face recognition system appears to have the structure of an and-gate; it requires two sorts of input. The first sort is the input which is absent in prosopagnosics, who are unable to recognize the faces of close relatives, or even their own face in a mirror. It has been suggested that the mechanism which produces this input is either a template-matching system or a constraint-satisfaction network (Farah 1990). However, it appears that there is also an affective response needed to underwrite face recognition, a neural pathway that gives the face you see its emotional significance. Only if both these sorts of inputs feed into the face recognition gate does full recognition take place. Several authors have suggested that it is the system subserving this affective response that is disrupted in Capgras patients. As a result, these patients have an experience analogous to seeing an identical twin of one's best beloved. The visual match is there, but not the emotional response.

This cannot be the whole story, however. Most of us would think that the trouble lay within ourselves if we started having such an experience. It appears that in Capgras the facial recognition system, getting only one sort of appropriate information, sends this on to more central systems which are also in trouble. Stone and Young (1997) argue that in addition to their agnosia Capgras' patients have a belief system which is too heavily weighted towards observational beliefs at the expense of background knowledge. Another possibility is that normal subjects have a mechanism which sets an upper bound on the weirdness of permissible beliefs, and that in Capgras subjects this is absent, or at least very permissive. Stone and Young argue that similar combinations of disordered affective response and belief-formation problems may underlie Cotard's Delusion, in which one forms the belief that one is dead. If these speculations are correct, then there is, in Capgras' Delusion, a part of the visual system that is working as designed, but receiving only one of the two sorts of input that it needs and passing its own output on to a central system which is also, perhaps, abnormal.

Since the Darwinian modules posited by evolutionary psychology can utilize input from non-proprietary stores of information, it can also happen that a module produces a problem-generating output because the non-proprietary information it needs is incorrect. A clear example of this is one of the hypotheses we have already mentioned - the idea that some dysthymic people may have a normally functioning reciprocal altruism module which is being fed with inappropriate data about the values of various acts in the patient's culture.

### *3.2 Disorders that Result from an Environment Different from What Mother Nature Intended*

Natural selection has no foresight; it is concerned only with what works in the here and now. A central tenet of evolutionary psychology is that the human mind is

designed to work in our ancestral, hunter-gatherer environment. Natural selection did not design it for the contemporary world. But, of course, a system may function admirably in one environment and work rather poorly in another. So it is entirely possible that the mind contains modules or other sorts of systems which were highly adaptive in the ancestral environment but which do not lead to functional behavior in our novel modern environments.

For example, the social competition theory of depression (Price et al., 1994; Nesse and Williams, 1995) is based on the idea that depression is an evolved response to loss of status, or to an unsuccessful attempt to gain status. In response to such a loss, it might be adaptive to abandon the strategy you were previously using in your attempts at status enhancement. Similarly, perhaps you should change behaviors if your previous behaviors were tied to reproductive potential you have now lost. The social competition theory claims that depression provides an introspectable marker which indicates when switching strategies to seek another niche is in order. If you are living in a small group, as our ancestors typically did, switching strategies might well result in considerably greater success. Depressed mood is nature's way of telling you to accept that your current behavior will not improve your reproductive lot and motivating you to try behaving differently. In the circumstances, you should evaluate your behavior thoroughly, dwelling on the negative.<sup>10</sup> In addition, you might try to stay out of social situations altogether if you think you lack the resources to do well in them, and indeed we find that "depressed individuals report being uncomfortable in interactions with others, often perceiving these interactions as unhelpful, or even as unpleasant or negative."(Gotlib 1992, p.151).

The social competition hypothesis sees our ancestral communities as miniature ecosystems in which individuals strive to find niches where they can excel and make a good living. In modern societies, though, your chance of excelling -- of being the best at anything, or indeed anywhere near the best -- are remote. If we have inherited a mechanism which is triggered when we believe ourselves to be outcompeted then that mechanism will fire frequently as we are inundated with information about accomplished people. But, of course, in the modern world it is far more likely that the mechanism will fail to achieve the goal it was selected to attain. If the mechanism is set off by the realization that one is not even close to being the best at anything in the global village of the information age then getting depressed is not likely to be an effective reaction. For it is typically the case that there is no other strategy to adopt -- no other niche one could fill -- which would do significantly better than the present one in that global competition. Moreover, the mechanism will frequently be set off even though its owner

---

<sup>10</sup> There seems to be good evidence that this happens in depression. In a review article, Pyszczynski & Greenberg (1987) found support for the idea that depressed individuals have elevated levels of self-focus and that self-focus increases following a loss in the personal, social or employment spheres. In addition, depressives have elevated levels of negative self-complexity and lowered levels of positive self-complexity (Woolfolk et al., 1995). That is, depressives tend to think of themselves unfavorably in many different ways, and are quite sophisticated at drawing distinctions among different ways of not being terribly good.

is actually doing very well in the *local* environment. You can be the most respected and admired real estate developer in Sioux Falls without being Donald Trump.<sup>11</sup>

The social competition hypothesis is not the only explanation for depression which sees it as a formerly adaptive trait which causes problems in our current environment. The defection hypothesis, proposed by Watson and Andrews (1998), Hagen (1998, MS) and others maintains that in the ancestral environment postpartum depression was an adaptive response which led women to limit their investment in the new child when, because of social, biological or environmental factors, a major investment in the infant would be likely to reduce the total number of offspring produced by that woman during her lifetime who would reach reproductive age and reproduce successfully. Among the social conditions in the ancestral environment that would have been good cues for triggering a sharply reduced maternal investment would be insufficient investment from the father and / or other appropriate kin. Biological cues would include problems with the pregnancy or birth, or other visible indications that the infant was not likely to be viable and healthy. Environmental cues would include harsh winters, famine conditions and other indications that material resources would be inadequate. In modern societies with elaborate support systems provided by the state and other organizations, it may be much less likely that these cues are reliable indicators that a mother who “defects” and sharply reduces her investment in her baby will increase her own reproductive fitness. But there is a growing body of evidence suggesting that these situations are indeed significantly correlated with postpartum depression. (Hagen, MS) So it may be that postpartum depression is yet another example of a condition produced by an adaptive mechanism that is functioning just as it was designed to function, though in an environment that is quite different from the one in which it evolved.<sup>12</sup>

---

<sup>11</sup> One question often asked about the social competition hypothesis is why it does not entail that just about everyone in modern societies should be depressed, since almost all of us are aware that there are lots of people who are better than us in just about anything that we do. Part of the answer, we think, is that different individuals will have different levels of sensitivity to the cues that trigger this sort of depression. We will say a bit more about individual differences in “trigger” sensitivity in our discussion of panic disorders at the end of this section. Another factor that might be relevant is that in some important ways modern societies may not be all that different from ancestral communities. For as Dunbar (this volume) has shown the social networks that individuals maintain in contemporary societies are similar in size to the social networks of individuals in surviving hunter-gatherer communities.

<sup>12</sup> Some theorists have proposed generalizing the defection hypothesis to cover many more (perhaps *all*) cases of depression. On this account, the adaptive function of depression is to negotiate a greater investment from other people with whom one is engaged in collective activities when one’s own investment seems unlikely to have a positive payoff. Depression, on this view, functions a bit like a labor strike. The depressed person withdraws his or her services in an effort to get a better deal in some cooperative enterprise. It is our view – though at this point it is little more than a guess – that it is counterproductive to seek a single account of the adaptive function of depression. Rather, we suspect, there may be several quite different kinds of depression, each with its own set of triggers and its own characteristic symptomology. Thus, for example, we noted earlier that Woolfolk et al. (1995) found that people who are depressed

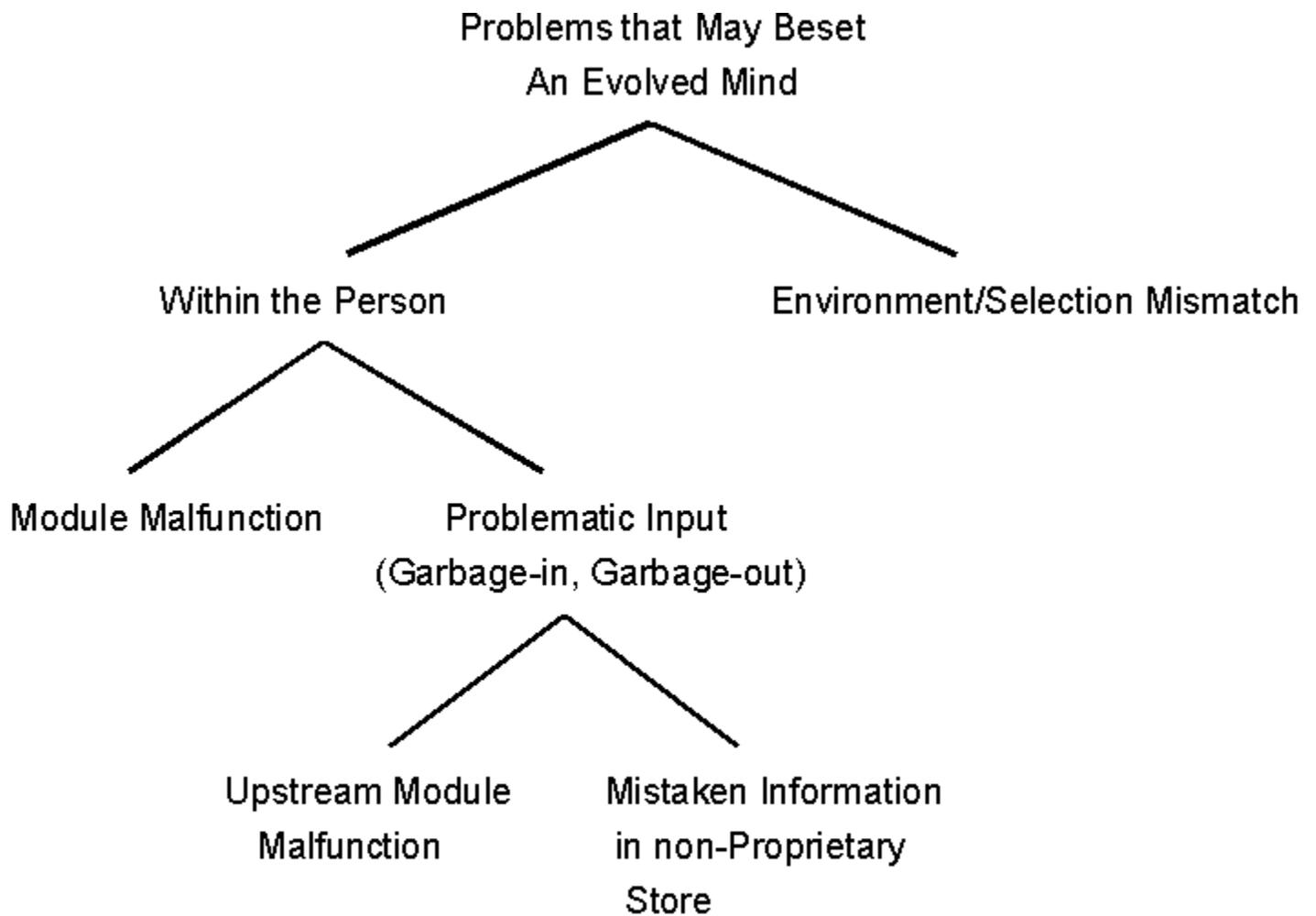
One of the morals to be drawn from these two hypotheses about depression is quite general. The environment in which selection pressures acted so as to leave us with our current mental endowment is not the one we live in now. This means that any mental mechanism producing harmful behavior in the modern world *may* be fulfilling its design specifications to the letter, but in an environment it was not designed for. In the disorders that result there is nothing in the mind which is malfunctioning.

Some anxiety disorders provide another possible example of disorders that result from a mismatch between the contemporary environment and the environment in which our minds evolved. Marks and Nesse (1994) note that in the ancestral environment fear of public places and fear of being far from home might well have been adaptive responses "that guard against the many dangers encountered outside the home range of any territorial species." (251) Similarly, a fear of heights accompanied by "freezing instead of wild flight" (251) would have had obvious adaptive value to our hunter-gatherer forebears. Moreover these trait, like most traits, could be expected to show considerable phenotypic variation even in a population of individuals who are genotypically identical with respect to the relevant genes. Individuals who are toward the sensitive end of these distributions -- those who become anxious more readily when far from home or when they find themselves in high places -- might well have functioned quite normally in ancestral environments. In a modern urban environment, however, people who become extremely anxious when they are away from home or when they are in public places will find it all but impossible to lead a normal life. And people who become extremely anxious in high places will find it difficult or impossible to travel in airplanes, ride in glass enclosed elevators or work on the higher floors of modern buildings. Thus, because the modern environment is so different from the ancestral environment, people who are toward the sensitive end of the distribution of phenotypic variation may be incapable of coping with many ordinary situations despite the fact that all of their mental mechanisms are functioning in just the way that natural selection designed them to function.

Figure 6 indicates the main categories of the in our proposed taxonomy of problems that may beset an evolved mind.

---

have elevated levels of "negative self-complexity." This is a symptom that makes perfect sense if the episode of depression is triggered by the sort of perceived failure or loss of status that plays a central role in the social competition account of depression. People in that situation need to think hard about what they are doing wrong. But the symptom makes less sense if the episode is triggered by a situation in which an individual's reproductive interests require renegotiation of the expected levels of investment in a collective activity. It would be very interesting indeed to know whether women suffering from postpartum depression exhibit negative self-complexity. On the pluralistic account of depression that we favor, it would be predicted that episodes of postpartum depression triggered by inadequate paternal and family investment are not marked by high levels of negative self-complexity.



**Figure 6**

**E-M Problems. A taxonomy of the main categories of problems that may beset an evolved mind.**

#### **4 The Causes of Disorder**

In the previous section we saw how the evolutionary psychology model of the mind suggests a theoretically motivated strategy for classifying mental disorders. We also noted some of the ways in which this taxonomy might prove useful in planning

therapeutic interventions. In our taxonomy, the notion of a module malfunction played a central role. In this section we want to consider some of the factors that can *cause* a module to malfunction. The discussion will illustrate some of the ways in which an evolutionary motivated taxonomy of disorders can integrate with a variety of scientifically promising approaches to mental disorder.

When considering the causes of module malfunction, it is, we think, both convenient and theoretically well motivated to distinguish two importantly different kinds of case: (i) those in which a fully developed, normally functioning module begins to malfunction and (ii) those in which, as the result of some problem in the course of development, the module in question never functions in the way that it was designed to. We'll consider these two kinds of cases in turn.

#### 4.1 *How can a normal module become pathological?*

Modules are computers and hence require some kind of physical substrate; in human beings, this substrate is the brain. Psychological disturbances may be a result of damage to the brain caused by strokes or injuries or by other kinds of physical trauma. Apperceptive agnosia, for instance, has often been noted in subjects who have undergone carbon monoxide poisoning (Farah 1990). Various physical disorders can also cause fully developed minds to malfunction. Metabolic disorders which interfere with the synthesis of neurotransmitters are an obvious example. Autoimmune responses of the sort believed to be responsible for Multiple Sclerosis can lead to the demyelination of nerve tissue which slows up the transmission of impulses. Late onset genetic disorders are still another example. Huntington's disease (a DSM-IV diagnosis) is a late-onset neurodegenerative disorder caused by an abnormally long CAG trinucleotide repeat in a dominant gene close to the tip of chromosome 4.

Earlier we urged adopting the idea that a disorder should be identified with a network of problematic modules. The physiological causes we have just discussed might have implications for more than one network. Indeed cognitive neuropsychology is bedeviled by the problem of understanding how to apportion to their proper diagnoses the variety of symptoms which typically follow one injury to the brain. How should our taxonomy take these causal factors into account? We propose a *two dimensional* classification of disorders that arise from problems within the person. One dimension is the network of problematic modules, the other dimension is the etiology of the malfunction - physiological, developmental, and so on. It may seem that the two-dimensional approach generates disorders beyond necessity. However, etiology is always important in medicine, which tends to regard information about causal history as vital to accurate diagnosis. The other possibility is to opt for a *one dimensional* classification that identifies disorders with a network of problematic modules (and other sorts of mental mechanisms) and to note in addition that some disorders can be caused in quite distinct ways. The two dimensional picture is richer and probably more in keeping with medical practice generally, but the question whether to adopt a one or two

dimensional picture should be decided on grounds of utility. Whichever emerges as the more useful approach should be adopted.

#### 4.2 *Problems that can prevent a module from developing properly*

In the last section we briefly considered some of the causes that might lead to a malfunction in a fully developed Darwinian module that was functioning normally to begin with. All of the factors we mentioned -- brain damage, stroke, physical trauma, physical disorders and genetic disorders -- can also occur in infancy or childhood. However, when a module functions improperly in the course of development, it can lead to a suite of problems that are quite different from the sorts of problems that would arise if the same module were damaged in an adult. One reason for this is that modules may require appropriate input to develop normally. Thus if a downstream module is supposed to be getting input from another module further upstream, and if the upstream module is damaged and fails to provide appropriate input, the downstream module may never develop properly. By contrast, damage to the upstream module in an adult might leave the downstream module unscathed.

In a series of recent publications, Baron-Cohen has proposed a theory of the origin of autism that fits this pattern. (Baron-Cohen, 1995; Baron-Cohen & Swettenham, 1996) On Baron-Cohen's account, normal development of the theory of mind module (ToMM) requires that three upstream systems be in place: an Intentionality Detector, an Eye Direction Detector, and a Shared-Attention Mechanism. In autism, a variety of biological hazards, most notably perinatal problems, damage the Shared Attention Mechanism. As a result, the downstream ToMM is deprived of the input it needs to develop properly.

Blair's theory of psychopathy provides another illustration of the way in which a malfunction of one mechanism early in development can prevent other mechanisms from developing properly. In Blair's theory, the Violence Inhibition Mechanism (VIM) generates an aversive affective response to signs of pain or distress, and this response is required if the systems responsible for empathy and for moral emotions such as guilt and remorse are to develop properly. Blair claims that in psychopaths the VIM is absent, due to a physiological deficit or poor socialization early in development, and as a result young psychopaths never acquire a capacity for empathy and are never able to experience the moral emotions.

There is another reason why module malfunction early in development can lead to problems that are quite different from those that result when a functional adult module is damaged. To see this, we must first explain a distinction between two importantly different sorts of modules. All of the modules that we've considered so far have, as their main function, subserving some capacity which, once in place, will typically remain intact for the rest of a person's life. Following Segal (1996) we will call these *synchronic modules*. However, theorists who advocate a highly modular account of mental architecture also posit a quite different sort of module which, again following Segal, we

will call *diachronic modules*. The computational program and proprietary information embedded in some synchronic modules may be largely insensitive to environmental variation. Synchronic modules of this sort will end up with the same program and information in any normal environment. In many other cases, however, natural selection has found methods of exploiting information available in the environment to fine tune the workings of a synchronic module in a way that would be adaptive in the environment in which our minds evolved. Thus some synchronic modules can develop in a variety of different ways resulting in adult modules that compute different functions or use different proprietary information. Diachronic modules are one of the mechanisms responsible for this process of fine tuning. Many diachronic modules function only in development, though others may be operative throughout life. Their job is to monitor the environment and to set the switches and dials of developing synchronic modules in appropriate ways. Perhaps the best known example of a diachronic module is the Language Acquisition Device (LAD) posited by Chomsky and his followers. Its job is to set the parameters of the Language Competence System thereby determining which of the large number of languages compatible with Universal Grammar the child will come to know. If a diachronic module malfunctions in the course of development, the synchronic module that it services may be set improperly or, in extreme cases, it may not function at all. Though there is still much research to be done, Specific Language Impairment might well turn out to be a developmental disorder that results from a malfunctioning diachronic module. (Gopnik, 1990a, 1990b; Gopnik & Crago, 1991)

In a well known series of experiments, Mineka et al. (1980, 1984, 1989) showed that young rhesus monkeys who have never seen snakes are not afraid of them, though they develop an enduring fear of snakes after only a few observations of another rhesus reacting fearfully to a snake. Rhesus monkeys do not, however, develop fear of flowers when they see another rhesus reacting fearfully to flowers. This suggests that rhesus may have a diachronic module (a Fear Acquisition Device, if you will) whose function it is to determine which of the switches on an innately prepared fear system get turned on. It is entirely possible that humans have a similar Fear Acquisition Device (FAD). If we do, then it may well be the case that some phobias (or some characteristic symptoms of phobias) are caused by a malfunction in the device which toggles an enduring terror of snakes or spiders, say, despite the fact that the person with the phobia has never seen anyone injured or frightened by snakes or spiders. In other cases phobias may arise when a properly functioning FAD is triggered inappropriately. If, for example, a child sees his parent reacting fearfully in response to something that looks like a snake or a spider, he may acquire a phobia even if the parent's fear was feigned or provoked by something else entirely. Still another intriguing possibility is that there are people whose FAD is defective in the other direction; instead of being too active, it is not active enough. These people would fail to develop fears or anxieties that they ought to develop (cf. Marks and Nesse, 1994). They would be unlikely to come to the attention of psychiatrists or clinical psychologists, though they might be more likely to come to the attention of coroners since the disorder may have a negative impact on their life expectancy.

## 5. Disorders that may not be

We noted above that the environment, especially the social environment, may change in ways which render well-designed systems pathological. However, an important possibility is that certain forms of what we currently take to be pathology are in fact straightforwardly adaptive in the current environment, just as they were in the ancestral environment in which our minds evolved. To put the point starkly, some people may be designed to be anti-social.

Personality disorders are patterns of experience and behavior which are culturally very deviant, persistent, inflexible, arise in adolescent or early adulthood, and lead to distress or impairment. However, it is not clear that antisocial behavior of this sort is always bad for the individual who commits it, rather than the people who are on the receiving end. McGuire et al. (1994) suggest that two personality disorders in particular may represent adaptive deviant behavioral strategies. The first, antisocial personality disorder, is characterized by a disregard for the wishes, rights or feelings of others. Subjects with this disorder are impulsive, aggressive and neglect their responsibilities. "They are frequently deceitful and manipulative in order to gain personal profit or pleasure (e.g. to obtain money, sex or power)". Typically, they show complete indifference to the harmful consequences of their actions and "believe that everyone is out to 'help number one'" (DSM-IV, 646).

The second disorder McGuire and his colleagues discuss is histrionic personality disorder. Subjects diagnosed as having this disorder are attention-seeking prima donnas. Often lively and dramatic, they do whatever is necessary to draw attention to themselves. Their behavior is often sexually provocative or seductive in a wide variety of inappropriate situations or relationships (DSM-IV, 655). They demand immediate satisfaction and are intolerant of or frustrated by situations which delay gratification. They may resort to threats of suicide to get attention and coerce better caregiving.(DSM-IV, 656) Both anti-social and histrionic personality disorders are characterized by manipulateness, although antisocial subjects manipulate others in the pursuit of material gratification and histrionics manipulate to gain nurture.

Now, on the face of it you might think that being able to manipulate other people so that they nurture you or further your material ends would be quite a useful trait to have, moral qualms aside. And of course one of the more annoying facts about such people is that they don't have moral qualms about their behavior. That makes it easier for them to commit the sorts of acts which occasionally lead to their arrest or undoing. To be classified as suffering from the relevant personality disorder, people must manifest a pattern of behavior that involves these undesirable social acts, though to satisfy the diagnostic criteria set out in DSM-IV their behavior must also "lead to clinically significant distress or impairment in social, occupational, or other important areas of functioning." (DSM-IV, 633) To put the point more colloquially, their behavior has to get them in trouble. However, it is quite likely that there are many people who are just as unsavory and manipulative but who do not get in trouble or suffer adverse consequences. It is estimated, for instance, that fewer than 25% of those who commit

nonviolent crimes are apprehended (McGuire et al. 1994). Such folk may cheat, deceive and manipulate but be good enough at reading social cues and understanding the structure of reciprocal exchange that they can exploit the social system successfully.

The natural way for philosophers to understand the function of a psychological mechanism, according to the conception of the mind we have presented, is in causal-historical terms (Millikan 1993, Neander 1991). This influential view construes the function of a psychological unit as the effect it has in virtue of which it is copied in successive generations. Now if it is indeed true that the disorders we have been considering are adaptive strategies, then we can give precisely this causal-historical explanation of the existence of the mechanisms which generate the antisocial behavior of sociopaths. Antisocial behavior is the proper function of these mechanisms. That pattern of behavior enables enough sociopaths to make a good enough living to ensure that antisocial mechanisms are copied in subsequent generations. On the standard causal-historical view of functions, then, the antisocial behaviors of the sociopathic are produced as the proper, selected functions of their peculiar psychological mechanisms. So these people are, in this respect, functioning as they should; they do not have a broken module or any other sort of malfunctioning mental mechanism. Nor is there reason to believe that the environment has changed in relevant ways since the time when the system was selected. The relevant environment in this case is social, and the current social environment, like the ancestral one, offers many opportunities to cheat and exploit one's fellows.

It is true that some sociopathic individuals spend their best reproductive years incarcerated. However, statistically it may be that other things being equal (general intelligence, normal childhood environments and so on) sociopathic behavior is quite adaptive -- it is an effective way getting one's genes into the next generation. Indeed, a population with a minority of sociopaths may be in an evolutionarily stable state. Skyrms has shown how this is mathematically possible for apparently bizarre strategies such as "Mad Dog", which rejects a fair division of resources but accepts a grossly unfair one (1996, pp.29-31); that is, Mad Dogs punish those who play fair. It is not hard to imagine the survival of more complex strategies which unfairly manipulate others.

These strategies will be useful provided two conditions apply. First, the subjects must often be able to disguise their cheating and deception, perhaps by exploiting and mimicking the signals which others use to convey cooperativeness (Frank, 1988). Second, the antisocial behaviors must be maintained at a comparatively low level in the population. If there are too many people who refuse to co-operate and deal fairly with others, then refusing to co-operate will gain one no dividends. We can expect an arms-race as sociopathic cheaters evolve to be even better at exploiting others and the others evolve to become better at detecting cheaters and avoiding them<sup>13</sup>.

---

<sup>13</sup> McGuire et al. are not the only theorists to have tried explaining a disorder in this way. Mealey (1995) thinks that primary sociopathy or psychopathy is an adaptive strategy. Her account is straightforwardly sociobiological, and, in contrast with Blair's theory, it neglects to go into any detail on cognitive mechanisms. (Blair & Morton, 1995). However, the two are not entirely incompatible. Blair thinks that the VIM is missing due to neurological impairment or poor early

As we suggested earlier,<sup>14</sup> we think there is an important distinction to be drawn between mental disorders and what, for want of a better term, we call *E-M problems* (*problems that may beset an evolved mind*). To count as an E-M problem, a condition must be located somewhere in the taxonomic structure sketched in Figure 6. And having an E-M problem is, we maintain, a *necessary* condition for having a mental disorder. But not all cases of E-M problems are or ought to be counted as mental disorders. Mental disorder is a partly normative notion; to count as a mental disorder a condition must cause problems for the people who have it or for those around them. A brain lesion that disrupts the normal function of some mental mechanism but whose only enduring result is that those with the lesion develop an intense interest in gourmet food would produce an E-M problem, but not be a mental disorder.<sup>15</sup> In other cases, E-M problems do not count as mental disorders for a variety of historical, social or practical reasons. Thus, as we mentioned earlier, Huntington's disease is a disorder included in DSM-IV, but Multiple Sclerosis is not, nor are various agnosias and aphasias. If, as McGuire and others have suggested, the mechanisms underlying various sorts of personality disorders are adaptations that evolved in environments which were relevantly similar to the modern environment, then people with these conditions do not have E-M problems, and thus, we maintain, *they do not have mental disorders*.

These people are problems, of course. But they are problems to us, and so are lots of other people who do not receive diagnoses of psychopathology. We might perhaps be able to drug them into submission, but that is best viewed as punishment or preemptive social control, not therapy. Similarly, if we could devise ways of restructuring their motivational system, it would be inappropriate to call the process "therapy." Rather, we should simply recognize that we are trying to manipulate behavior in the interests of social harmony. Unless we want to medicalize all deviant behavior, we must acknowledge the possibility that apparently disordered behavior, which receives a DSM diagnosis, can be produced by a psychological endowment functioning exactly as it was designed to, in just the environment it was picked to work in. One of the virtues of the evolutionary approach to psychopathology is that, in some cases at least, it provides a principled way of drawing the distinction between mental disorders and patterns of anti-social behavior produced by people whose evolved minds are beset by no problems at all.

---

socialization. Mealey can be read as offering an alternative reason for the absence of the VIM; some people are just not designed to have one. The developmental consequences might then unfold as Blair envisages. We could treat Mealey as giving the "ultimate" explanation and Blair the "proximate" one (Mayr, 1976).

<sup>14</sup> See fn. 6.

<sup>15</sup> This is not merely a hypothetical case. See Regard & Landis, (1997).

## ACKNOWLEDGEMENTS

We are grateful to James Blair, Peter Carruthers, Andrew Chamberlain, Rachel Cooper, Fiona Cowie, Brian Loar, Richard Samuels, David Sloan Wilson, Terry Wilson and Robert Woolfolk for helpful feedback on the ideas developed in this paper. Earlier versions of the paper were presented to audiences at the California Institute of Technology, Central Michigan University, the University of California, Santa Cruz, the University of Sheffield, the University of Utah and Washington University. Comments and criticisms from these audiences have proved helpful in many ways.

## REFERENCES

- American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3<sup>rd</sup> ed. Washington, D.C.: American Psychiatric Association. (DSM-III).
- American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders*, 3<sup>rd</sup> ed., revised. Washington, D.C.: American Psychiatric Association. (DSM-III-R).
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4<sup>th</sup> ed. Washington, D.C.: American Psychiatric Association. (DSM-IV).
- Baron-Cohen, S, Leslie, A & Frith, U (1985). Does the autistic child have a theory of mind ? *Cognition*, 21, 37-46.
- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.
- Baron-Cohen, S. & Swettenham, J. (1996). The relationship between SAM and ToMM; two hypotheses. In P. Carruthers & P. Smith (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
- Bellugi, U. Klima, E. & Wang, P. (1997). Cognitive and neural development: Clues from genetically based syndromes. In Magnusson, D. et al. (eds.), *The lifespan development of individuals: Behavioral, neurobiological, and psychosocial perspectives: A synthesis*. New York: Cambridge University Press. Pp. 223-243.
- Blair, R. & Morton, J. (1995). Putting cognition into sociopathy. *Behavioral and Brain Sciences*, 18, 548.
- Blair, R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, 57, 1-29.
- Clarkin, J., Widiger, A., Francis, S., Hurt & Gilmore, M. (1983). Prototypic typology and the borderline personality disorder. *Journal of Abnormal Psychology*, 92, 263-275.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with Wason Selection Task. *Cognition*, 31, 187-276.
- Cosmides, L. and Tooby, J. (1992). Cognitive adaptations for social exchange. In Barkow, J., Cosmides, L., and Tooby, J. (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press. 163-228.
- Dunbar, R. (this volume). On the origin of the human mind.
- Farah, M. (1990) *Visual Agnosia*. Cambridge, MA: MIT Press.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

- Frank, R (1988) *Passions Within Reason*. New York: W.W. Norton.
- Frith, U. (1989). *Autism: Explaining the Enigma*. Oxford: Blackwell.
- Gigerenzer, G. and Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition*, 43, 127-171.
- Goodwin, D. & Guze, S. (1995). *Psychiatric Diagnosis*, 5th ed. New York: Oxford University Press.
- Gopnik, M. & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition*, 39, 1-50.
- Gopnik, M. (1990a). Dysphasia in an extended family. *Nature*, 344, 715.
- Gopnik, M. (1990b). Feature blindness: A case study. *Language Acquisition*, 1, 139-164.
- Gotlib, I. (1992) Interpersonal and cognitive aspects of depression. *Current Directions in Psychological Science*, 1, 149-154.
- Grandin, T. & Scariano, M. (1986). *Emergence Labelled Autistic*. Tunbridge Wells: Costello.
- Griffiths, P. (1997). *What Emotions Really Are*. Chicago: University of Chicago Press.
- Hagen, E. (1998). The functions of postpartum depression and the implications for general depression. Paper presented at the Tenth Annual Meeting of the Human Behavior and Evolution Society, Davis, CA, July 8-12, 1998.
- Hagen, E. (Undated MS). The functions of postpartum depression.
- Hempel, C (1965). Fundamentals of taxonomy. In C. Hempel, *Aspects of Scientific Explanation*. New York: The Free Press.
- Hull, D (1989). On human nature. In D. Hull, *The Metaphysics of Evolution*. Albany: SUNY Press.
- Karmiloff-Smith, A., Klima, E., Bellugi, U., Grant, J. et al. (1995). Is there a social module? Language, face processing, and theory of mind in individuals with Williams syndrome. *Journal of Cognitive Neuroscience*, 7, 2, 196-208.
- Kessler, R., McGonagle, K., Zhao, S., Nelson, C., Hughes, M., Eshleman., W. & Kendler, K (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorder in the United States. *Archives of General Psychiatry*, 51, 8-19.
- Leslie, A & Thaiss, L (1992). Domain specificity in cognitive development; neuropsychological evidence from autism. *Cognition*, 43, 225-251.

Leslie, A (1987) Pretense and representation: the origins of "theory of mind". *Psychological Review*, 94, 412-426.

Leslie, A (1991) The theory of mind impairment in autism: evidence for a modular mechanism of development ? In Whiten, A (ed) *Natural Theories of Mind*. Oxford: Blackwell.

Marks, I. M & Nesse, R. M (1994). Fear And fitness: An evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, 15, 247-261.

Mayr, E (1976). Cause and effect in biology. In E. Mayr, *Evolution and the Diversity of Life: Selected Essays*. Cambridge, MA: Harvard University Press.

Mayr, E (1982) *The Growth of Biological Thought*. Cambridge, MA; Harvard University Press.

McCarthy, L & Gerring J (1994) Revising psychiatry's charter document DSM-IV. *Written Communication*, 11, 147-192.

McGuire, M & Troisi, A (1998) *Darwinian Psychiatry*. New York; Oxford University Press.

McGuire, M., Fawzy, F, Spar, J., Weigel, R. & Troisi, A. (1994) Altruism and mental disorders. *Ethology and Sociobiology*, 15, 299-321.

Mealey, L (1995) The sociobiology of sociopathy; an integrated evolutionary model (with commentary). *Behavioral and Brain Sciences*, 18, 523-599.

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

Mineka, S., Keir, R. & Price, V. (1980). Fear of snakes in wild and laboratory-reared rhesus monkeys, *Animal Learning and Behavior*, 8, 653-663.

Mineka, S., Davidson, M., Cook, M. & Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. *Journal of Abnormal Psychology*, 93, 355-372.

Mineka, S. & Tomarken, A. (1989). The role of cognitive biases in the origins and maintenance of fear and anxiety disorders. In L. Nilsson & T. Archer (eds.), *Aversion, avoidance, and anxiety: Perspectives on Aversely Motivated Behavior*. Hillsdale, NJ: Erlbaum.

Neander, K (1991) Functions as selected effects; the conceptual analysts defense. *Philosophy of Science*, 58, 168-184.

Nesse, R.M & Williams, G.C (1995). *Why We Get Sick*. New York: Times Books.

Poland, J., Von Eckardt, B. & Spaulding, W. (1994). Problems with the DSM approach to classifying psychopathology. In G. Graham & G. L. Stephens (eds.), *Philosophical Psychopathology*. Cambridge, MA: MIT Press.

Price, J, Sloman, L, Gardner, R, Gilbert, P & Rohde P (1994). The Social Competition Hypothesis of Depression. *British Journal of Psychiatry*, 164, 309-315

Pyszczynski, T & J. Greenberg (1987). Self-regulatory perseveration and the depressive self-focussing style: a self-awareness theory of reactive depression. *Psychological Bulletin*, 102, 122-138.

Regard, M. & Landis, T. (1997). " 'Gourmand Syndrome': Eating passion associated with right anterior lesions." *Neurology*, May 1997.

Sachs, O. (1995). *An Anthropologist on Mars: Seven Paradoxical Tales*. New York: Knopf.

Samuels, R. (1998). Evolutionary Psychology and The Massive Modularity Hypothesis. *British Journal for the Philosophy of Science*, 49, 575-602.

Samuels, R. (this volume). Massively modular minds: the evolutionary psychological account of cognitive architecture.

Segal, Gabriel (1996) The modularity of theory of mind. In P. Carruthers & P. Smith (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Skyrms, B (1996). *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Stevens, A. & Price, J. (1996). *Evolutionary Psychiatry: A New Beginning*. London: Routledge.

Stone, T. & Young A. (1997). Delusions & brain injury: The philosophy and psychology of belief. *Mind and Language* 12, 327-64.

Tooby, J. and Cosmides, L. (1990a). 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. and Cosmides, L. (1990b). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, 11, 375-424.

Tooby, J. and Cosmides, L. (1992). The psychological foundations of culture. In Barkow, J., Cosmides, L., and Tooby, J. (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press. 19-136.

Tooby, J. and Cosmides, L. (1995). Foreword. In Baron-Cohen (1995), xi-xviii.

Trivers, R (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-57.

Wakefield, J. (1997). Diagnosing DSM IV, part 1: DSM and the concept of disorder. *Behavior Research and Therapy*, 35, 633-49.

Watson, P. & Andrews, P., An evolutionary theory of major depression. Paper presented at the Tenth Annual Meeting of the Human Behavior and Evolution Society, Davis, CA, July 8-12, 1998.

Williams, G. (1977). *The Pony Fish's Glow*. New York: Basic Books.

Wilson, D. (1994). Adaptive genetic variation and human evolutionary psychology. *Ethology and Sociobiology*, 15, 219-235.

Woolfolk, R., Novalany, J., Gara, M. Allen, L. and Polino, M. (1995). Self-complexity, self-evaluation and depression: An examination of form and content within the self-schema. *Journal of Personality and Social Psychology*, 68, 1108-1120.