Contents

Impressum

Evolution and Cognition: ISSN: 0938-2623 **Published by:** Konrad Lorenz Institut für Evolutions- und Kognitionsforschung, Adolf-Lorenz-Gasse 2, A-3422 Altenberg/Donau. Tel.: 0043-2242-32390; Fax: 0043-2242-323904; e-mail: sec@kla.univie.ac.at; World Wide Web: http://www.kli.ac.at/ **Editors:** Rupert Riedl, Manfred Wimmer **Layout:** Alexander Riegler **Aim and Scope:** “Evolution and Cognition” is an interdisciplinary forum devoted to all aspects of research on cognition in animals and humans. The major emphasis of the journal is on evolutionary approaches to cognition, reflecting the fact that the cognitive capacities of organisms result from biological evolution. Empirical and theoretical work from both fields, evolutionary and cognitive science, is accepted, but particular attention is paid to interdisciplinary perspectives on the mutual relationship between evolutionary and cognitive processes. Submissions dealing with the significance of cognitive research for the theories of biological and sociocultural evolution are also welcome. “Evolution and Cognition” publishes both original papers and review articles. **Period of Publication:** Semi-annual. **Price:** Annuals subscription rate (2 issues, postage and handling included): Euro 40, US$ 50, SFr 65, GBP 28. Annual subscriptions are assumed to be continued automatically unless subscription orders are cancelled by written information. **Single issue price** (postage and handling included): Euro 25, US$ 30, SFr 35, GBP 18. **Publishing House:** WUV-Universitätsverlag/Vienna University Press, Berggasse 5, A-1090 Wien, Tel.: 0043/1/3105356-0, Fax: 0043/1/3197050. **Bank:** Erste österreichische Spar-Casse, Acct.No. 073-08191 (Bank Code 20111). **Advertising:** Vienna University Press, Berggasse 5, A-1090 Wien. **Supported by Cultural Office of the City of Vienna, Austrian Federal Ministry of Education, Science, and Culture, and the Section Culture and Science of the Lower Austrian State Government.**

Robert Plutchik

Integration, Differentiation, and Derivatives of Emotion

Abstract

An approach to conceptualizing emotions based upon evolutionary ideas is presented. From this perspective, emotions are seen as fundamental adaptive processes that serve both communicative functions and survival. A theory is outlined that consists of three components, a structural model, a sequential model, and a derivatives model. These models describe systematic connections between emotions, cognitions, personality traits, personality disorders, ego defenses, and other domains. These domains reflect different kinds of languages for describing the same underlying processes. Emotions are proximate methods used to achieve ultimate goals of inclusive fitness. Cognitions developed in order to predict the future more effectively.

Key words

Emotion, evolution, personality, cognitions, theories.

ALMOST EVERYONE agrees that the study of emotion is one of the most confused chapters in psychology. This is supported by a number of facts. First, there are very few university courses that deal primarily with the psychology of emotion. Second, it has been reported that there are over 90 definitions of emotion that had been proposed over the past century or more (KLEINGINA/KLEINGINA 1981). Third, there is much disagreement among contemporary theoreticians concerning the best way to define or conceptualize emotions. Finally, there are a number of different theories of emotion currently in existence whose adherents interpret the role of emotion in life quite differently (PLUTCHIK 1994).

Problems in the Study of Emotion

There are many reasons for this state of affairs. I will briefly list five.

1. Most of us have learned to be cautious about accepting at face value other people's comments about their feelings. One important reason for the distrust we may feel is that we suspect that others censor what they say. This is easy to accept because we are aware of the fact that we often censor our own thoughts and feelings.

2. Another reason for the distrust of verbal reports is the powerful effect that behaviorism has had on the thinking of psychologists. Behaviorists believed that the only truly reliable, objective information obtainable from living creatures was information about their behavior. Emotions were considered to be inner states that could not be reliably observed and were therefore outside the realm of scientific psychology.

3. Language is inherently ambiguous so that the meaning of emotion terms depends to some degree on an individual's particular verbal history and facility with words. Metaphors are often used to express the complexity of emotional feelings. In addition, mixed emotions are common and difficult to describe in an unequivocal way.

4. Psychoanalysts have made us aware of the fact that emotions may be repressed, inhibited or unconscious, and thus be unavailable to introspection.

5. A few major historical figures have stimulated intellectual traditions concerning emotion which still have a powerful effect on the thinking of contemporary researchers. These figures are Charles DARWIN, William JAMES, Walter CANNON, Sigmund FREUD, and Fritz HEIDER. They have established what I call the evolutionary, the psychophysiological, the neurological, the psychodynamic, and the cognitive traditions in the study of emotion. Table 1 lists the key ideas of each tradition.

Despite the different concerns and emphases that characterize these various approaches, it is reasonable to assume that each theory contains some truth. Any adequate integration of emotion theories

The Evolutionary Tradition Key Person: Charles Darwin (1809–1882) Key Idea: Emotions are acts and communications in important life situations that influence interpersonal relations, and function to increase the chances of survival.
The Psychophysiological Tradition Key Person: William James (1842–1910) Key Idea: Emotions are subjective feelings based on the awareness of internal autonomic changes associated with actions.
The Neurological Tradition Key Person: Walter Cannon (1871–1945) Key Idea: Emotions are subjective feelings resulting from hypothalamic arousal, and are usually associated with acts of fight or flight
The Psychoanalytic Tradition Key Person: Sigmund Freud (1856–1939) Key Idea: Emotions are complex states, which may be conscious or unconscious, reflecting conflicts, early experiences, personality traits and defenses. They can only be inferred on the basis of various kinds of indirect evidence.
The Cognitive Tradition Key Person: Fritz Heider (1896–1988) Key Idea: The beliefs of individuals, particularly their goals, their casual attributions, and their expectations all influence their emotions. The presence of emotions may also influence beliefs.

Table 1. Historical traditions in the study of emotions.

should have something to say about the adaptive implications of emotions, about autonomic and brain mechanisms in relation to feeling states, about dynamic sources and complexities of emotional states, and about the role of cognitions.

Emotion as Part of an Evolutionary Process

Most people accept the idea that emotions are not limited only to adults who are capable of verbalizing their feeling states. Clinicians believe that schizophrenics often have intense emotions which they are unable to verbalize. Mentally retarded individuals may show strong emotional reactions that are evident in their behavior. Young children, before they have developed much of a vocabulary of emotions, show emotional behavior. Many studies of infants report that infants express emotions that have an immediate and strong influence on their caretakers. Finally, it is evident from an extensive literature, that researchers believe that animals show emotions. We see them in our pets, and most ethologists are quite comfortable writing about emotions in dogs, cats, rats, elephants, tigers, chimpanzees, baboons, and lower animals. (Buirski/Plutchik 1991; Griffin 1992; Goodall 1987).

These observations raise the reasonable question of how far down the evolutionary scale the concept of emotion applies. Should we arbitrarily stop applying the concept of emotion at the level of fish, for example, or invertebrates, or can the concept of emotion in some way apply even to single celled organisms? In this connection it is interesting to note that Darwin wrote in his 1872 book on emotions "Even insects express anger, terror, jealousy, and love by their stridulations" (1965, p349).

Emotions in Lower Organisms

It is not surprising that those of us who study other animals look for signs of motivation, perception, learning and emotion in them even though we are constantly faced with the specter of anthropomorphism. Animal rights advocates are concerned about the pain and distress caused to research animals by such conditions as social deprivation, isolation, and crowding. However, the farther removed an animal is from ourselves, the less sympathetic we are in ascribing to it those mental states we recognize in ourselves. Although we may be willing to say that someone is as "mad as a hornet", we are reluctant to ascribe emotions to insects. Despite this attitude many investigators had used related terms in describing organisms near the beginning of the phylogenetic scale.

For example, Goodenough (1991, p344) uses the metaphor of warfare in describing infectious disease. "We tend to think of bacteria or viruses as launching an attack, while the immune system puts up a defense… This context of deception and detection takes place at the molecular level. The actors are individual cells and protein molecules". Against the array of immune system defenses the invading pathogens have developed several strategies. One is camouflage. Since the essential discrimination made by the immune system is between self and nonself, the body asks, in effect, "Does this molecule belong to me, or is it part of an invading organism?" This may be an example of implicit cognition without consciousness. One form of camouflage is for the invader to make itself resemble the host molecularly. A strain of E. coli cells uses this trick by covering their surface with long polymers of sialic acid. Another method of camouflage is for the invading bacteria to take a residence inside the host cell as is the case with malarial parasites. Another approach to this biological warfare is mimicry. The rhinovirus, for example, fools the cell into accepting it into the cell where it begins to multiply. These are metaphors, of course, yet they have a strangely persuasive quality.

Microbiologists have made some interesting points in this connection. Every single-celled organism from the blue-green algae to the eukaryotic cells are complete, self-sufficient organisms. They are exposed to daily risks with corresponding needs for adaptations. Such cells need to take in food, excrete waste products, avoid predators, reproduce by exchange of genes, seek safe environments, and explore their microbiological world (RICCI 1990). These adaptations are to the same types of problems faced by higher multicellular organisms.

In detailed studies of five species of ciliates, RICCI (1990) found that they swim forward in a counterclockwise helicoidal path. Every few seconds, the organisms stop and change direction. This has been described as exploratory behavior. Signs of a "mating dance" that leads to the fusion of potential partners into conjugating pairs has been reported in these ciliated bacteria. Preferences of some bacteria for smooth surfaces rather than rough ones has also been observed. All organisms shows start and stop patterns of behavior as well.

What are some implications of these observations? In the early environment of the earth, more than two billion years ago, small bacteria (prokaryotes) appeared. A bacterial cell with 2000 genes could carry out a large number of metabolic activities and could replicate and divide accurately. It could also flourish in diverse environments.

Because no environment is optimum for growth indefinitely, there is a strong selective advantage to moving around. The large eukaryotes became predators that engulfed the smaller bacteria. They evolved nuclei, internal membranes to encase their chromosomes, and special intracellular proteins, all adaptations which are found in many current living cells. In fact, "all eukaryotic organisms, from algae, to trees to elephants, appear to have descended from a single protoeukaryotic cell" (LOOMIS 1988, p161). Evidence for this comes from work in molecular biology. A related idea is that various systems developed millions of years ago that worked so well they have never been improved upon. For example, mammalian sperm now use the same flagellum for locomotion that evolved 600 million years ago to keep algae cells near the surface of water. Similarly, the amino acid sequences for both X and B tubulin (the bases of the microtubules that form flagellae) are more than 70 percent similar in yeast, algae, sea urchins, chickens, rats, pigs, and humans.

Another illustration of the same point is that a small, acidic protein called calmodulin that binds free calcium has an amino acid structure that is almost identical from the amoebae to humans. The same basic histone proteins and nucleosome structure is found in the chromosomes of all eukaryotic cells throughout the plant and animal kingdoms. The gene order on the sex chromosome of mice (the X chromosome) is almost identical with that on the human X chromosome. The development of sexual dimorphism (specialization of males for sperrn production and females for egg production) is also extremely widespread. The advantage of this system is that the variability of genetic potentials increases the chances of the individual successfully dealing with changing or catastrophic environments.

Similarities in evolutionary patterns are also found through studies of development. For example, the similarity of developmental patterns in limb bones is evidence that amphibians, reptiles, and mammals all evolved from a common stock of lobefin fish. More than 100 years ago it had been noted that there are remarkable similarities in the structures that appeared in the early stages of embryogenesis in mammals, amphibians, birds, reptiles, and fish. A small number of developmental genes can radically alter the behavior of cells and change an amoeba into a multicelled organism. LOOMIS (1988) estimates that fewer than 1400 developmental genes may have been sufficient for the evolution of simple cells into fish, and fewer than 2500 developmental genes may be sufficient for the embryogenesis of humans. "The important evolutionary differences between a guppy and a primate probably lie in only a few hundred genes" (LOOMIS 1988, p216).

These various observations, from evolutionary and molecular biology, emphasize the points that DARWIN first made: namely, that evolutionary continuities of structure, function and development imply continuities of behavioral adaptations and mental life.

Based on these kinds of observations and on the basic assumption of conservation of effective systems, the zoologist, SCOTT (1980) has pointed out that the nature of the environment creates certain functional requirements for all organisms if they are to survive. An organism must take in nourishment and eliminate waste products. It must distinguish between prey and predator and between a potential mate and a potential enemy. It must explore its environment and orient its sense organs appropriately as it takes in information about the beneficial an harmful aspects of its immediate world. And in organisms that are relatively helpless at birth and for a while thereafter, there must be ways of indicating the need for care and nurturance. The specific behaviors by which these functions are carried out vary widely throughout the animal kingdom, but the basic prototype functions remain invariant.

SCOTT (1958) has elaborated on this theme. He suggests that there are only a few classes of adaptive behavior found in most species and phylogenetic levels. He describes them in the following terms; ingestive behavior, shelter-seeking behavior, agonistic (fight or flight) behavior, sexual behavior, care-giving behavior, care-soliciting behavior, eliminative behavior, allelomimetic (imitative) behavior, and investigative behavior.

These patterns of behavior may be thought of as prototype adaptations that are relevant to most organisms and that have mental states associated with them. Agonistic behavior, for example, is defined as fight or flight associated with the subjective states of anger and fear.

In my 1980 book (PLUTCHIK 1980), I elaborated on these ideas and suggested that there are eight basic patterns of adaptation to environmental inputs. These are listed in Table 2 along with the hypothesized emotional states for which they are assumed to be prototypes. This does not mean that a prokaryote bacterium feels fear as it is about to be engulfed by a larger bacterium, but it does struggle and attempt to flee. Somewhere in the course of evolution, a subjective feeling of fear develops that acts to amplify the struggle to avoid being eaten. It is important to emphasize that an emotion is more than the subjective feelings we recognize in human adults. Emotions have an evolutionary history as do all mental and bodily states, and, as fundamental adaptive processes related to survival are quite complex.

The basic point being made is that the concept of emotions applies to animals as well as humans, but that in all cases, the judgment that an emotion exists is an inference from certain classes of evidence. This evidence includes: knowledge of stimulating conditions, the effects of behavioral acts, knowledge of typical behavior patterns of the individual and species, choices made when alternatives exist, reactions of other members of one's own group or species, and other considerations. One of the more important reasons that emotional states are difficult to define unequivocally is that more than one emotion can occur at the same time. Any given overt display of emotions can reflect such complex states as approach and avoidance, attack and flight, sex and aggression or fear and pleasure. It appears that the inference of emotions in lower animals can be scientifically valid. A detailed discussion of these issues may be found in my book *Emotion: a psychoevolutionary synthesis* (PLUTCHIK 1980).

Prototype Adaptation	**Hypothesized Emotion**
Protection: Withdrawal, retreat, contraction	Fear, Terror
Destruction: Elimination of barriers to the satisfaction of needs	Anger, Rage
Incorporation: Ingesting nourishment	Acceptance
Rejection: Riddance response to harmful material	Disgust
Reproduction: Approach, contact, genetic exchanges	Joy, Pleasure
Reintegration: Reaction to loss of a nutrient object	Sadness, Grief
Exploration: Investigation of one's environment	Curiosity, Play
Orientation: Reaction to contact with a strange object	Surprise

Table 2. Prototype behavioral adaptations and hypothesized subjective emotional states that appear late in the course of evolution.

The Psychoevolutionary Theory of Emotion

During the past four decades I have been developing a theory of emotion which I refer to as a psychoevolutionary theory. It is based on the DARWINIAN assumption that emotions are modes of adaptation to significant events in an organism's environment, and that they are complex processes having functional value both for purposes of communication and to increase an individual's chances of survival. Emotions represent proximate methods to achieve ultimate goals of inclusive fitness (PLUTCHIK 1962 1980, 1989). The general theory includes three subsidiary models: a structural model, a sequential model, and a derivatives model.

The Structural Model

This aspect of the theory assumes that emotions may be conceptualized in a fashion analogous to colors and that the relations among emotions can be represented by a three-dimensional model shaped like a cone. The vertical dimension represents the intensity of emotions, the circle defines degree of similarity of emotions, and polarity is represented by the opposite emotions on the circle (Figure 1). This postulate also includes the idea that some emotions are primary and others are derivatives or blends in the sense that some colors are primary and others are mixed. A number of studies have been published showing that the language of emotions can be represented by means of a circle or circumplex (PLUTCHIK 1980; CONTE/PLUTCHIK 1981; FISHER et. al. 1985). The circumplex is in essence a cross section of the emotion solid (Figure 2). An empirical circumplex based on a similarities scaling method is shown in Figure 3.

A key aspect of the structural model is the idea that there is a small number of primary emotions and that all others are derived from them. Over the past centuries, from Descarte to the present, philosophers and psychologists have proposed anywhere from three to 11 emotions as primary or basic. All lists include fear, anger and sadness, and most include joy, love, and surprise (KEMPER 1987). There is no unequivocal way to decide on the precise number of primary emotions there are, although factor analytic studies are useful, as are similarity scaling studies. Also relevant are child development studies and cross-cultural studies. But in the final analysis, the number of primary emotions is a theoretical decision to be evaluated in terms of the inferences and insights to which it leads, the research it suggests, and the extent to which empirical data are consistent with it.

Integration of Emotions

From the point of view of integration of emotions from basic ones, an interesting question arises. If there are eight basic emotion dimensions (each with a number of synonyms or related terms), how can we account for the total language of emotions? Various published researches (see PLUTCHIK 1994) imply that the total number of emotion words is a few hundred at most and they tend to fall into families based on similarity. If we follow the pattern used in color theory and research, we can obtain judgments about what results when two or more emotions are combined. When this was done, the results were clear; judges agreed that the mixture of *joy* and *acceptance* produces the mixed emotion of *love*. The blending of *disgust* and *anger* produces the mixed emotional state of *hatred* or *hostility*. Such mixtures have been called primary dyads in the theory. Table 3 shows a number of other examples. By mixing two or more emotions at different intensity levels it is possible to create hundreds of terms representing the language of emotions.

Emotions and Personality

Another important idea stemming from the structural model, is the fact that many of the terms that

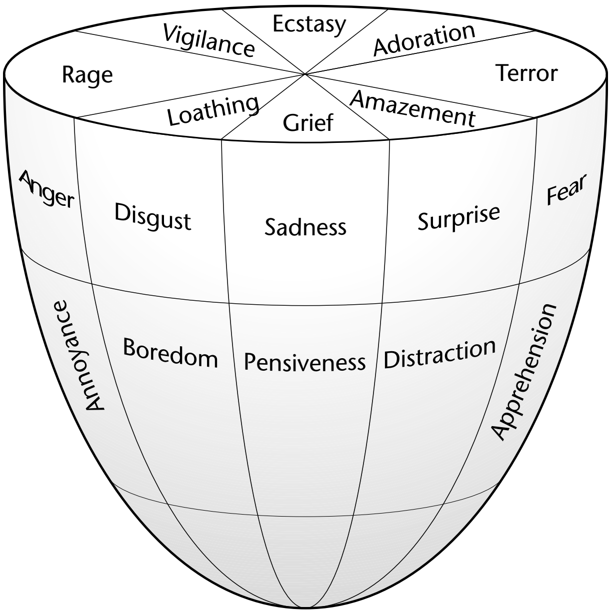

Figure 1. A multi-dimensional model of emotions.

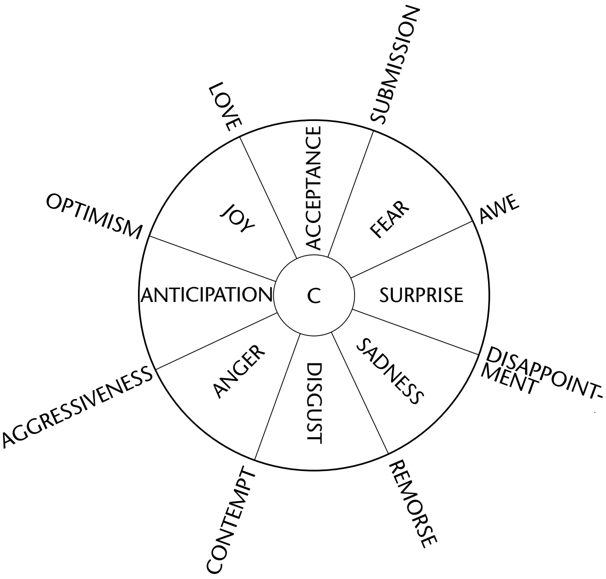

Figure 2. Primary dyads formed by the combination of adjacent pairs of basic emotions.

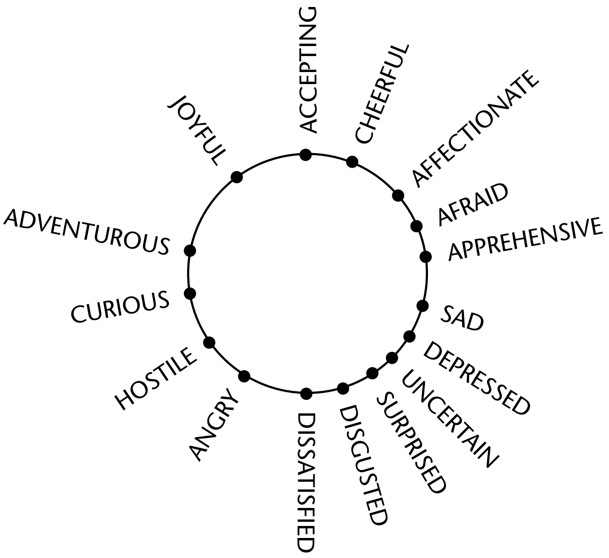

Figure 3. Angular locations of emotions based upon similarity judgments.

Primary Emotion Components		Labels for Mixed Emotions
Joy + Acceptance	=	Love, Friendliness
Fear + Surprise	=	Alarm, Awe
Sadness + Disgust	=	Remorse
Disgust + Anger	=	Contempt, Hatred, Hostility
Joy + Fear	=	Guilt
Anger + Joy	=	Pride
Fear + Disgust	=	Shame, Prudishness
Anticipation + Fear	=	Anxiety, Caution

Table 3. Description of mixed emotions (from Plutchik 1980).

judges used in describing mixtures of the emotions are words that are typically used to describe personality traits. In fact, it is evident that most of the terms used to describe emotions are also used to describe personality traits. For example, words such as *gloomy, resentful, anxious* or *calm* can describe personality traits as well as emotional feelings. The distinction between emotional states and personality traits is largely arbitrary (Allen/Potkay 1981). Often the same adjective checklist can be used to measure both states and traits by a simple change in instructions. If the research participants are asked how they feel *now,* or within the past few days or so, we are asking about emotional states or moods. If, however, they are asked to describe how they *usually* feel, we are asking about personality traits. Thus, from the point of view of this theory, emotions and personality traits are intimately connected, and in fact, personality traits may be considered to be derived from mixtures of emotions. This idea of derivatives of emotion will be elaborated shortly.

The Sequential Model

Psychologists have long been concerned with the sequence of events in emotion. Despite considerable interest in the question, no definitive answers have been found.

A major reason for this lack of closure is the fact that emotions are not simply linear events, but are feedback processes. For example, Weisman (1965) states that the function of affect is to restore the individual to a state of equilibrium, or in other words, the acts associated with emotions tend to reduce the emotions which produced them.

In a related statement, Karasu (1992) points out that each person tries to maintain a certain level of affective equilibrium in everyday life with only moderate fluctuations. Unexpected or unusual events (external or internal) change this affect level and the resulting behaviors try to reestablish the preexisting state.In an emotional reaction, once the goal of aggression or escape, for example, has been achieved, and the individual's relation to the environment has changed, the emotional response declines.The psychoevolutionary theory accepts the idea that emotions are part of complex, circular, feedback systems. It assumes that stimulus events, either external or internal (as in dreams), act as the primary triggers that start the emotion process going. However, events need to be interpreted in order for them to have an effect on the individual. A picture of an American flag may elicit feelings of pride and enthusiasm to an American, and feelings of hate and vengeance for an Iraqi. Sometimes the interpretation is obvious and occasionally less so. For example, individuals sometimes take an instant dislike to someone they have just met. The reason may not be obvious either to an observer or to the individual himself. In such a case, we assume that an interpretation or a cognition has occurred that may be unconscious and so we make an inference about the cognition on the basis of the behavior shown. The psychoevolutionary theory assumes that following the cognition or interpretation, a feeling state occurs as well as a physiological state of arousal, if appropriate.

Feeling states tend to be followed by impulses to action. Such impulses may be expressed by tensions in the muscles, by facial expressions, by clenching of the fists, or by preparations for running, attacking or yelling. Clinicians are well aware of the fact that impulses to action are not always followed by action, often for fear of retaliation or fear of embarrassment. However, action often does occur; the individual runs, attacks, criticizes, cries, compliments, kisses or withdraws.

Such overt behavior is, however, not the end of the emotion process. Such behavior generally has an effect on the stimulus or condition that started the chain of events in the first place. For example, running from a source of threat reduces the threat and tends to reestablish the condition that existed before the threat occurred. Similarly, if a major loss occurs in an individual's life such as the death of a parent, the crying and grieving that results tends to produce supportive and helpful contacts from other members of one's social group and at least in a symbolic way, provides a kind of reintegration with the lost parent. Overall, this process is a kind of homeostatic process, but one that is carried out by behavioral rather than internal changes. I call this process a behavioral homeostatic negative, feedback system. From this point of view an emotion is not simply the feeling state but the entire chain of events including the feedback loops.

Figure 4 depicts this process in general terms. Feedback loops may influence the impulses to action, the feeling states, the cognitions, as well as the initiating stimulus. This process is what leads to the idea that feelings and behaviors can affect cognitions, just as much as cognitions can influence feelings. Also implied by this model is the idea that the term "feelings" is used to represent subjective, reportable states such as joy, or sadness, anger, or disgust, while the word emotion is used in a much broader sense to refer to the entire chain of events that include feelings, but also cognitions, impulses to action, display behaviors, goal-directed acts and the various loops that occur.

Table 4 summarizes the theory's assumptions about some of the key elements involved in the emotion sequence. For each of the eight basic emotions a general description of the stimulus event that triggers it, is described, followed by descriptions of the probable cognitions associated with each of the emotions, the subjective feeling states, the overt behaviors and the effect of the behavior in reducing the disequilibrium.

At the heart of all these descriptions is the idea that emotions have a purpose in the lives of individuals. This idea stems from the evolutionary perspective, is consistent with psychodynamic thinking and is becoming more and more accepted in the writings of contemporary clincians. For example, HAUSER (1996) points out that the primary care that young organisms require is for food, protection and transportation and that crying is a major method for getting such care. SPEZZANO (1993) suggests that we use love or intimidation to keep others invested in out personal agendas. VALLIANT (1994) has discussed the adaptive functions of a number of emotions. Sorrow, for example, increases one's feelings of closeness to others, and listeners often feel compassion and the desire to be helpful. Interest, or anticipation, is often energizing and increases one's involvement with others. Fear protects the self, initiates withdrawl, and allows general functioning to continue. Shame leads to remorse and a decrease in the probability of repetition of the shameful act. All these examples imply that emotions are part of a functional adaptive feedback process.

Some Thoughts on Cognitions and Emotions

In the most basic sense, any organism must predict on the basis of limited information whether there is food, a mate or danger in its environment. Depending on the prediction made, the organism makes a decision to escape, to attack, to eat, or to mate. From this point of view the complex processes of sensory input, evaluation, symbolization, comparison with memory stores and the like—or those processes we call cognitive—are in the service of emotions and biological needs. Predicting the characteristics of environments enables organisms to prepare for those environments.

This idea is not new. For example, LICHTENBERG/NORTON (1970, p53) said that "cognition is the capacity to perceive, evaluate and act in an effective manner in order to cope with one's environment". NEISSER (1963), in his comparison of humans with computing machines, suggests that cognitions are in the service of emotions and biological needs. Information from the environment, he says, is evaluated in terms of its need-satisfying and need-frustrating properties. It is because of the need to learn environmental maps as well as biologically significant events that most primates have such a long childhood. It is a childhood that must be protected in order to allow the child time to learn the major signals needed for survival in a hostile environment. What is particularly significant is that each

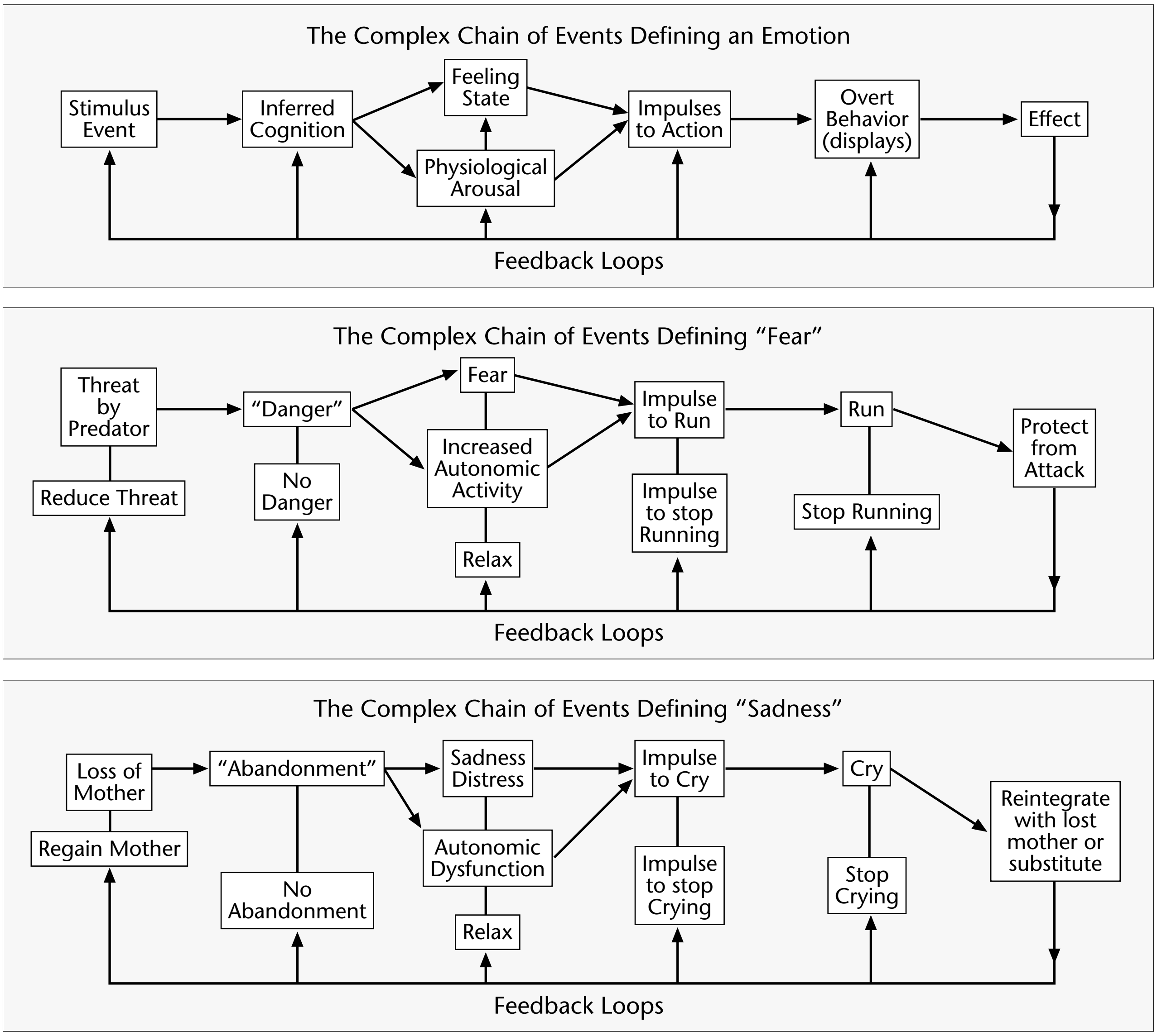

Figure 4. The complex chain of events defining an emotion.

Stimulus Event	**Cognition**	**Feeling State**	**Overt Behavior**	**Effect**
Threat	'Danger'	Fear	Escape	Safety
Obstacle	'Enemy'	Anger	Attack	Destroy Obstacle
Gain of Valued Object	'Possess'	Joy	Retain or Repeat	Gain Resources or New Genes
Loss of Valued Object	'Abandonment'	Sadness	Cry	Reattach with Lost Object
Member of One's Group	'Friend'	Acceptance	Groom	Mutual Support
Unpalatable Object	'Poison'	Disgust	Vomit	Eject Poison
New Territory	'Examine'	Expectation	Map	Knowledge of Territory
Unexpected Event	'What is it?'	Surprise	Stop	Gain Time to Orient

Table 4. Key elements in the emotion sequence.

new cognitive experience that is biologically important is connected with an emotional reaction such as fear, pleasure, pain, disgust, or depression. From the point of view of evolution, cognitions developed in order to predict the future more effectively.

The very first living organism had to emote; that is, fight and flee, eat and expel, reproduce, explore, start and stop. These actions are still part of humans' behavioral repertoire, changed in form, modified in appearance, reactive to different stimuli, but functionally the same. The brain, which evolved as an adaptation to a changing and difficult environment, has now helped create the very environment to which it must continue to adapt.

These views have been elaborated in an early paper of mine (Plutchik 1977). In that paper I presented 10 postulates about the relations between cognitions and emotions. I will briefly summarized some of these postulates here.

1. The existence of any emotion presupposes the occurrence of an evaluation. It is not necessary to assume that such evaluations are conscious or reportable; they may be inferred in the same way that we infer the existence of cognitive maps in lower animals. However, not all evaluations produce emotions.

2. Evaluations that produce emotions are concerned with whether a stimulus is good or bad for survival, beneficial or harmful, productive of pleasure or productive of pain.

3. Most evaluations that individuals make about the events in their lives are multiple; that is, an event may be evaluated as both good and bad, harmful as well as a source of pleasure. Such multiple evaluations lead inevitably to conflicts between emotions.

4. In lower animals, certain evaluations occur without prior learning. Examples would be the distress vocalizations of puppies when isolated, and fighting behavior of Siamese fish.

5. Evaluations may be in error. An individual may evaluate a stimulus as dangerous (as in a bird phobia) when it actually is not. However, on the average, most evaluations must be reasonably accurate if an individual is to survive.

In summary, cognitions are generally found near the beginning of the chain of events we call an emotion. They can be influenced, however, by events that appear later in the chain (such as states of arousal, or ego defenses) through a circular feedback process. From this point of view, the question of which comes first the cognition or the emotion is a pseudoproblem that might best be ignored.

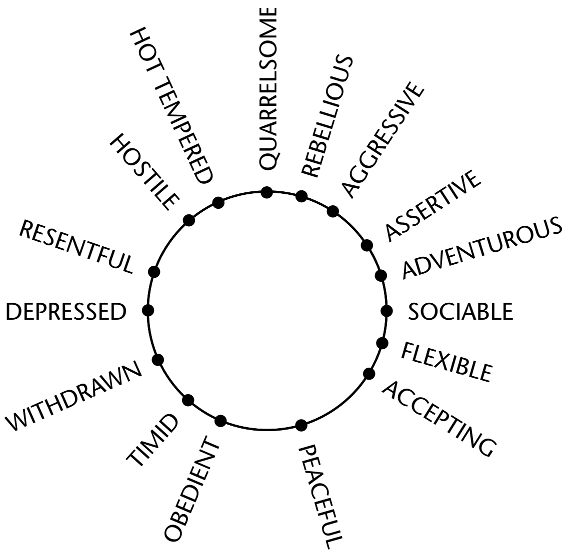

Figure 5. Angular locations of personality traits based upon similarity scaling.

The Derivatives Model

One of the most important ideas of the psychoevolutionary theory is the concept of derivatives. This term is used to express the idea that certain conceptual domains are derivatives of other more basic concepts.

Take, for example, the domain of personality. The subject is usually taught in universities as if it had little or nothing to do with emotions. Yet the language of emotions and the language of personality are remarkably similar. An individual can feel depressed or be a depressed person; can feel nervous or be a nervous person; or can feel joyful or be a joyful person.

In addition to the overlap in the language of the domains of emotion and personality, there is also the fact that both domains can be represented by means of a circumplex model. This has been illustrated by Conte/Plutchik (1981) who used two independent methods to establish a circumplex structure for personality traits; one was a direct similarity scaling method and the other was bsed upon the use of the semantic differential followed by factor analysis of the data. Angular locations of a sample of 40 personality traits obtained by the two methods correlated +.98. Some of these locations are shown in Figure 5.

This circumplex for personality traits, using a different methodology, has been very closely replicated (Fisher/Heise/Bohrnstedt/Lucke 1985; Fisher 1997). Circumplex models have also been described for

Subjective Language	Behavioral Language	Functional Language	Trait Language	Diagnostic Language	Ego-Defense Language	Coping-Style Language
Fear	Escape	Protection	Timid	Avoidant, Dependent	Repression	Avoidance
Anger	Attack	Destruction	Quarrelsome	Antisocial	Displacement	Substitution
Joy	Mate	Reproduction	Sociable	Hypomanic	Reaction-Formation	Reversal
Sadness	Cry	Reattachment	Gloomy	Dysthymic	Compensation	Replacement
Acceptance	Groom	Incorporation	Trusting	Histrionic	Denial	Minimization
Disgust	Vomit	Rejection	Hostile	Paranoid–Narcissistic	Projection–Intellectualization	Mapping
Expectation	Map	Exploration	Demanding	Obsessive–Compulsive	Intellectualization	Mapping
Surprise	Stop	Orientation	Indecisive	Borderline	Regression	Help Seeking

Table 5. Emotions and their derivatives.

other interpersonal domains including facial expressions (MYLLYNIEMI 1997), vocational interests (TRACEY/ROUNDS 1997), parent-child relations (SCHAEFER 1997), social acuity (empathy) (GURTMAN 1997), and social support (WIGGINS/TROBST 1997), in short, anything related to interpersonal relations.

The idea of derivatives can be extended further. Diagnostic terms such as "depressed", "manic", and "paranoid" can be conceived as extreme expressions of such basic emotions as sadness, joy, and disgust. Several studies have also revealed that the language of personality diagnoses shares a circumplex structure with emotions (PLUTCHIK/PLATMAN 1977; PLUTCHIK/CONTE 1985). These ideas are illustrated in Table 5.

A Note on Differentiation

Most emotions that occur in life are mixed emotions. They contain elements of approach and withdrawal, sexuality and aggression, or surprise and curiosity. When judges are asked to identify the components of a given mixed emotion (or personality trait) they are able to do this for most terms quite reliably. This ability to analyze the components of a mixed state is particularly useful in a clinical psychotherapeutic context. Here is a clinical example.

The psychoevolutionary theory proposes that all emotions are either one of the eight basic ones, or mixed states. It further assumes that the basic emotions seldom, if ever, occur in pure form, and if they do, only transiently. Most emotions, therefore, are mixed emotions or blends. A further assumption is that the blending of emotions always produces some level of conflict. Several personality tests based on these assumptions have been developed and empirical research supports the idea that different personality traits indirectly express different levels of conflict (PLUTCHIK/KELLERMAN 1974; CONTE/PLUTCHIK 1986). The therapeutic principle this leads to is this: Most emotional states are mixtures implying conflicts. To understand the nature of the conflict we need to identify the components.

Example: A client said she felt guilty about leaving her husband and getting her own apartment. Previous research suggests that guilt is a mixture of fear and pleasure. It was possible to explore these components of her guilt. She was in conflict over her fear of not being able to make it on her own (i.e. continued dependence),and her pleasure at the thought of making it (i.e. being independent). She was in conflict over the fear of breaking up her family versus the joy of consolidating or remaking her family. She was in conflict over the fear that her husband would interfere and stop her from leaving and her pleasure at the thought of saying "No" to him. Examining these components separately enabled her to evaluate the relative importance of each one and to then make a reasoned decision for her life.

In a book based on the psychoevolutionary theory, 16 clinical implications are given for psychotherapeutic strategies which are designed to identify and uncover emotions (PLUTCHIK 2000).

Carrying the notion of derivatives still further, our research has shown that the language of ego defenses can also be conceptualized as being related to emotions. For example, displacement can be conceptualized as an unconscious way to deal with anger that cannot be directly expressed without punishment. Similarly, projection can be conceptualized as an unconscious way to deal with a feeling of disgust for (or rejection of) oneself by attributing this feeling to outsiders. Parallels of this sort have been drawn for each of the primary emotions and are described in detail in KELLERMAN (1979), PLUTCHIK/KELLERMAN/CONTE (I979), and PLUTCHIK (1995). The concept of derivatives is illustrated more fully in Table 5, where the conceptual links between affects, behavior, functions, personality traits, diagnoses, and ego defenses are shown. Also added is the domain of coping styles, which can be hypothesized as the conscious derivatives of the unconscious ego defenses. Thus, faultfinding corresponds to projection, reversal to reaction formation, and mapping to intellectualization. Other derivative domains have also been proposed (PLUTCHIK 1984, 1989).

Implications of the Psychoevolutionary Theory

The theory states that emotions represent fundamental adaptive mechanisms related to interorganismic communication and individual as well as genetic survival. Emotions have a genetic substrate, and can be identified in functional terms at all phylogenetic levels. This theory is parsimonious in that the same set of assumptions has relevance and explanatory value for a number of conceptual domains (affects, personality, defenses, diagnoses, coping styles). It has predicted some new observations that have been empirically confirmed (the circumplex structure of affects, personality traits, diagnoses, and defenses). It has also provided some new insights into specific issues such as the relations between emotions and cognitions (PLUTCHIK 1977), emotions and imagery (PLUTCHIK 1984), emotions and empathy (PLUTCHIK 1987), emotions and nightmares (KELLERMAN 1987), emotions and primary processes (KELLERMAN 1989), and emotions and projective tests (KELLERMAN 1989). Also of great importance, it has provided a theoretical rationale for the construction of a number of new test instruments designed to measure affects (PLUTCHIK 1966, 1989), personality (PLUTCHIK/KELLERMAN 1974) ego defenses (PLUTCHIK/KELLERMAN/CONTE 1979), and coping styles (BUCKLEY et. al. 1984, WILDER/PLUTCHIK 1982). The theory thus provides a guide for research, shows connectedness between diverse domains, and predicts some relationships that have been empirically confirmed.

Author's address

Robert Plutchik, Albert Einstein College of Medicine and University of South Florida, 4505 Deer Creek Boulevard, Sarasota, Florida 34238, USA. Email: proban@home.com

References

Allen, B./Potkay, C. R. (1981) On the arbitrary distinction between states and traits. Journal of Personality and Social Psychology 4:916–928.

Buckley, R./Conte, H. R./Plutchik, R./Wild, K./Karasu, T. B. (1984) Psychodynamic variables as predictors of psychotherapy outcome. American Journal of Psychiatry 141:742–747.

Buirski, P./Plutchik, R. (1991) Measurement of deviant behavior in a Gombe chimpanzee: Relations to later behavior. Primates 32:207–211.

Conte, H. R./Plutchik, R. (1981) A circumplex model for interpersonal traits. Journal of Personality and Social Psychology 2:823–830.

Darwin, C. (1965) The expression of the emotions in man and animals. Chicago: University of Chicago Press. Originally published in 1872.

Fisher, G. A. (1997) Theoretical and methodological elaborations of the circumplex model of personality traits and emotions. In: Plutchik, R./Conte, H. R. (eds) Circumplex models of personality and emotions. American Psychological Association Press: Washington DC, pp. 245–269.

Fisher, G. A./Heise, D. R./Bohrnstedt, G. W./Lucke, J. F. (1985) Evidence for extending the circumplex model of personality trait language to self-reported moods. Journal of Personality and Social Psychology 49:233–242.

Goodall, J. (1987) The chimpanzees of Gombe: Patterns of behavior. Belknap: Cambridge MA.

Goodenough, U. W. (1991) Deception by pathogens. American Scientist 79:344–355.

Griffin, D. R. (1992) Animal minds. University of Chicago Press: Chicago.

Gurtman, M. B. (1997) Studying personality traits: The circular way. In: Plutchik, R./Conte, H. R. (eds) (1997) Circumplex models of personality and emotions. American Psychological Association Press: Washington DC, pp. 81–102.

Hauser, M. D. (1996) The evolution of communication. Bradford Press: Cambridge MA.

Karasu, T. B. (1992) Wisdom in the practice of psychotherapy. Basic Books: New York.

Kellerman, H. (1979) Group therapy and personality: Intersecting structures. Grune & Stratton: New York.

Kellerman, H. (ed) (1987) The nightmare: Psychological and biological foundation. New York: Columbia Univ. Press.

Kellerman, H. (1989) Projective measures of emotion. In: Plutchik, R./Kellerman, H. (eds) The measurement of emotions, Volume 4. Academic Press: New York, pp. 187–204.

Kemper, T. D. (1987) How many emotions are there? Wedding the social and the autonomic components. American Journal of Sociology 93:263–289.

Kleinginna. R. R./Kleinginna, A. M. (1981) A categorized list of emotion definitions, with suggestions for a consensual definition. Motivation and Emotion 5:345–379.

Lichtenberg, P./Norton, D. G. (1970) Cognitive and mental development in the first five years of life. PHS Publ. No. 2057-NIMH.

Loomis, W. F. (1988) Four billion years: An essay on the evolution of genes and organisms. Sinauer Associates: Sunderland MA.

Myllyniemi, R. (1997) The interpersonal circle and the emotional undercurrents of human sociability. In: Plutchik, R./Conte, H. R. (eds) Circumplex models of personality and emotions. American Psychological Association Press: Washington DC, pp. 271–298.

Neisser, U. (1963) The imitation of man by machine. Science 139:193–197.

Plutchik, R. (1962) The emotions: Facts, theories, and a new model. Random House: New York.

Plutchik, R. (1966) Multiple rating scales for the measurement of affective states. Journal of Clinical Psychology 22:423–425.

Plutchik, R. (1977) Cognitions in the service of emotions: An evolutionary perspective. In: Candland, D. K./Fell, J. P./Keen, E./Leshner, A. I./Plutchik, R./Tarpy, R. M. (eds) Emotion. Brooks/Cole: Monterey CA, pp. 189–212.

Plutchik, R. (1980) Emotions: A psychoevolutionary synthesis. Harper & Row: New York.

Plutchik, R. (1984) Emotions: A general psychoevolutionary theory. In: Scherer, K. R./Ekman, P. (eds) Approaches to emotion. Erlbaum: Hillsdale NJ, pp. 197–220.

Plutchik, R. (1987) Evolutionary bases of empathy. In: N. Eisenberg, N./Strayer, J. (eds) Empathy and its development. Cambridge University Press: New York, pp. 38–46.

Plutchik, R. (1989) Measuring emotions and their derivatives. In: Plutchik, R./Kellerman, H. (eds) Emotion: Theory, research, and experience, Volume 4: The measurement of emotions. Academic Press: San Diego CA, pp. 1–35.

Plutchik, R. (1994) The psychology and biology of emotion. Harper Collins: New York.

Plutchik, R. (1995) A theory of ego defenses. In: Conte, H. R./Plutchik, R. (eds) Ego defenses: Theory and measurement. Wiley: New York, pp. 13–37.

Plutchik, R. (2000) Emotions in the practice of psychotherapy: Clinical implications of affect theories. The American Psychological Association Press: Washington DC.

Plutchik, R./Conte, H. R. (1985) Quantitative assessment of personality disorders. In: Nickols, R./Cavenar Jr., J. O./Brodie, H. K. H. (eds) Psychiatry, Volume 7. J. B. Lippincott: Philadelphia, pp. 1–13.

Plutchik, R./Kellerman, H. (1974) Manual of the Emotions Profile Index. Western Psychological Services: Los Angeles.

Plutchik, R./Kellerman, H./Conte, H. R. (1979) A structural theory of ego defenses and emotions. In: Izard, C. E. (ed) Emotions in personality and psychopathology. Plenum Press: New York, pp. 229–257.

Plutchik, R./Platman, S. R. (1977) Personality connotations of psychiatric diagnoses. Journal of Nervous and Mental Disease 165:418–422.

Plutchik, R./Van Praag, H. M. (1987) Interconvertibility of five self-report measures of depression. Psychiatry Research 22:243–256.

Ricci, N. (1990) The behavior of ciliated protozoa. Animal Behavior 40:1048–1069.

Schaefer, E. S. (1997) Integration of configurational and factorial models for family relationships and child behavior. In: Plutchik, R./Conte, H. R. (eds) Circumplex models of personality and emotions. American Psychological Association Press: Washington DC, pp. 133–154.

Scott, J. P. (1958) Animal behavior. University of Chicago Press: Chicago.

Scott, J. P. (1980) The function of emotions in behavioral systems: A systems theory analysis. In: Plutchik, R./Kellerman, H. (eds) Theories of emotion, Volume 1. New York: Academic Press.

Spezzano, C. (1993) Affect in psychoanalysis: A clinical synthesis. Analytic Press: Hillsdale NJ.

Tracey, T. J. G./Rounds, J. B. (1997) Circular structure of vocational interests. In: Plutchik, R./Conte, H. R. (eds) Circumplex models of personality and emotions. American Psychological Association Press: Washington DC, pp. 183–204.

Vaillant, L. M. (1997) Changing character: Short-term, anxiety-regulating psychotherapy for restructuring defenses, affects and attachment. Basic Books: New York.

Weisman, A. D. (1965) The Existential Core of Psychoanalysis: Reality Sense and Responsibility. Little, Brown: Boston.

Wiggins, J. S./Trobst, K. K. (1997) When is a circumplex an "interpersonal circumplex"? The case of supportive actions. In: Plutchik, R./Conte, H. R. (eds) Circumplex models of personality and emotions. American Psychological Association Press: Washington DC, pp. 57–80.

Wilder, J. S./Plutchik, R. (1982) Preparing the professional: Building prevention of burnout into professional training. In: Payne, W. (ed) Job Stress and burnout. Sage: Beverly Hills CA.

Jesse M. Bering

Are Chimpanzees 'Mere' Existentialists?

A Phylogenetic Approach to Religious Origins

Abstract

The evolutionary emergence of mental representational abilities is of paramount concern to investigators interested in tracking the biological bases of religion. Both explicit cultural religious agent concepts, such as God, ancestral spirits, deities, and demons as well as cross-cultural, implicit nonmaterial concepts, such as luck, fate, and immanent justice, involve reasoning about mental causality of natural events. The author explores the possibility of religion as a specialized capacity dedicated to finding intention in events rather than actions. Viewing the cognitive capacities of Pan troglodytes *as important indices of the minds of early hominids, a new phylogenetic approach to religious origins is advanced by reviewing chimpanzees' understanding of intentional causation and placing our knowledge of chimpanzees' "theory of mind" into a discussion of religion's natural foundations.*

Key words

Religion, evolution, theism, chimpanzees, cognitive evolution, theory of mind, intentionality, causality.

The neural architecture of the human brain, the organ that has more than tripled in size in the last 6 million years as a result of presumably selective forces (Povinelli/Preuss 1995), is responsible for, among perhaps other cognitive specializations, the ubiquity of religion in the species. Surprisingly, religion and its possible cognitive precursors have been highly neglected topics in the cognitive sciences, despite the importance of this category of thought to human behavior and cognition. This may be because cross-cultural comparisons of religious behaviors have generally concentrated on the differences between institutionalized beliefs, and, in so doing, have overlooked critical similarities that may be tied together by isomorphic cognitive demands. Yet capturing these similarities can enable scholars to track the evolution of a central system involved in religious reasoning and explication. The term *religion* is used throughout this paper not in the sociological sense as a cultural construct, but in the biological sense in which non-natural[1] intentional causality (e.g., theism, luck, fate, immanent justice, deontic code, etc.) of naturally occurring events is inferred in all human societies.

To claim that any cognitive system is specialized to a single species demands that investigators take a comparative approach in order to determine the extent to which the parts that play a role in its construction are present or absent in other closely related species (Byrne 1995). By charting the phyletic history of modular-like mechanisms, researchers are better able to see the timeline and unfolding of emergent properties lending uniqueness to species. The supposition, of course, is that the brains of extant species were tinkered with through selective processes, and that the common ancestors of related species were not as specialized to later ecological and social niches as their offspring species. Only through selective forces, either natural or artificial, do modifications to the basic system take place.

Because it has largely faced the same ecological pressures as the common ancestor of African apes and modern humans, there is reason to believe that *Pan troglodytes*, but not necessarily *Pan paniscus*, is a *conservative* species (Zihlman 1996). That is, in all likelihood, it has retained the same basic body and social structure of the progenitor species as a result of relatively stable environmental demands. Shifts in these demands might have led the chimpanzee lineage to evolve more derived physical or behavioral traits, but there was little need, it seems, for the line to have undergone much recent specialization. In contrast, while it is unclear what first prompted the cascading changes in the brains and bodies of those *derived* species of organisms that became human, dramatic changes nevertheless occurred. With each new species that erupted in the human lineage, special adaptations, both physical and behavioral (and correspondingly brain-based), made humans strikingly distinct from their comparatively meager beginnings, such that a mere 5–7 million years later the significant shared heritage between humans and chimpanzees is scarcely recognizable behind a gross body morphology and a basic behavioral and social repertoire.

What Chimpanzees Can Tell Us about Religious Origins

By adhering to the reductionist definition of religion presented above, chimpanzees' understanding of intentions is critical to our understanding of the natural history of spirituality (see BERING 2001; MASER/GALLUP 1990). The combination of religion and evolution, in this sense, and its focus on chimpanzee cognition will face some acrimonious contention, to be sure, but outright dismissal of its plausibility would be without just *scientific* reason. Chimpanzees have been an important model in comparative psychology, and have informed researchers on such important topics as the origins of language (e.g., JENSVOLD/GARDNER 2000), tool use (e.g., STANFORD/GAMBANEZA/NKURUNUNGI/GOLDSMITH 2000), teaching (e.g., BOESCH 1991), cooperation (e.g., MITANI/MERRIWETHER/ZHANG 2000), gestural communication (e.g., TOMASELLO/CAMAIONI 1997), culture (e.g., TOMASELLO 1999), deceit (e.g., BYRNE/WHITEN 1991), imitation (e.g., BERING/BJORKLUND/RAGAN 2000), folk physics (e.g., POVINELLI 2000), mathematics (e.g., BOYSEN 1993), art (e.g., LENAIN 1997), and self-conception (e.g., ANDERSON/GALLUP 1999), to name but a few areas of recent inquiry. Why has religion, as a biologically engrained cognitive mechanism, not similarly received its due? Arguably as central to being human as any of these things, it would appear as if religion has stood apart from this list because it has been overwhelmingly viewed from the sociological standpoint; religion has been portrayed as something that is entirely "learned" in the context of culture, and thus coming from "outside the head". The role of intuitive structures dedicated to religion has been explicitly avoided by many evolutionary theorists (e.g., BOYER/WALKER 2000; SPERBER/HIRSCHFELD 1999), who have instead paid increasing attention to the transmission of cultural religious concepts. Such investigations, however, fail to recognize the recurrent aspects of religiously related behavior, such as rituals (see LAWSON/MCCAULEY 1990) and the construction of creation myths (including Judeo–Christian conceptions), as defining patterns of the way human brains have evolved. Whatever its reason, in order to gain any real insight into the phylogenetic emergence of religious behavior, the neglect of intuitive religion from an evolutionary perspective must be seriously dealt with, and chimpanzee cognition may be one of the first starting points to do just this.

Religious Representation and Causal Agency

But this, the skeptic will assert, is easier said than done. After all, we cannot ask chimpanzees whether they believe in God. Or can we? If we are concerned with the cognitive mechanisms that are primary players in the instantiation of religious beliefs, and if these cognitive mechanisms—which may have functions outside religious matters as well—can be probed in the minds of another species, then in many ways we certainly can. The question is not, do chimpanzees possess religious beliefs, but rather, *can chimpanzees, given their evolved cognitive architecture, harbor religious representations?*

The ontogeny of symbolic representation has been well documented in the cognitive developmental literature (see BJORKLUND/PELLEGRINI in press). PIAGET argued that young children's symbolic abilities are evident in their deferred imitation (shown by MELTZOFF (1990) to be present even in young infants), and pretend play, which is first demonstrated at about 18–24 months of age (LESLIE 1987). Although there may be differences between the capacities of chimpanzees and children in these areas (see BERING 2001), apes demonstrate some measure of proficiency in these domains as well. For a discussion of religion and phylogeny, however, such mental symbolism, which involves the storage and retrieval of conceptual imagery, is less important than skills centering on inductive inference involving perceptual displays serving as symbols, such as concrete objects, manifest behavior, and natural events. When we see a broken key lying next to a door, for instance, we infer that someone has unsuccessfully attempted to open it; when we see a woman scratching at hives on her neck, we infer that she is incredibly itchy; and when we hear a loud explosion as we ride down the expressway, we (woefully) infer that our car has just gotten a flat tire. If we ask ourselves, "what does this mean?" after experiencing these symbols, we are led to logical presumptions. In this sense, meaning is reserved to the propositional relationships between symbolic perceptual experiences (seeing a key, watching an actor, hearing a noise) and these symbols' logically embedded properties. Such properties may be hidden, as in the perceptual state of "itchiness", or observable, as in a flat tire.

This quality of *intentionality*—inferring the "aboutness" of a perceptual display—need not have anything to do with inferring the intentions of mental agents (DENNETT 1987). The woman intentionally scratched her neck to relieve her itching, it is assumed, and in this case her scratching behavior was "about" her discomfort, but for the other two scenarios there is no reason to posit any intentional actions to make the likely inference. Broken keys are "about" failed attempts at door-opening, and loud explosive noises on the expressway are often "about" flat tires. These things symbolize, then, *observable* properties embedded in propositional relationships. However, all three perceptual displays can also, given certain important contextual features, constitute events that lead to the instantiation of secondary representation of non-natural agency intentions. For the individual who has just returned home from work to find that her key has been stolen from its secret hiding spot beneath the potted plant by her front door, the broken key might symbolize not only the fact that someone unsuccessfully attempted to break into her house, but possibly that "someone was looking out for her" that day—namely some non-natural agent such as God–and intentionally caused the key to break apart. Likewise, the woman scratching at hives not only symbolizes her state of itchiness, but for the seventh grader who did not study for the day's geometry test and enters the classroom to see his teacher applying calamine lotion to her neck and writing "class cancelled" on the blackboard, such an event might also symbolize his deceased grandfather's intentions to stave off the exam for one more day in order to give him a second chance to study. Finally, the loud explosive noise on the freeway might, for the unfaithful spouse, symbolize not only the fact that he or she has a flat tire, but also the admonishing wrath of a judicial non-natural agent.

Thus, we can speak of *levels of intentionality*, whereby "aboutness" can be nested hierarchically among the aggregated layers of a life event and can appear at either a 'lower' level (e.g., what is the broken key about in relation to the door?) or a 'higher' level (e.g., what is the broken key about in the context of my life?). There can be many shades of meaning within the hierarchy, holding physical and social implications, but the highest level of meaning represents a search for 'ultimate' causality and approaches the religious impulse. *Religious reasoning occurs when there is an encroachment of intentionality in the domain of event perception such that events high in self-relevancy symbolize the intentions of a non-natural mental agent.* Because it operates at the molar level, religion serves an intentional causal explanatory function by explaining other, lower levels of natural causal explanation (WEEKS/LUPFER 2000).

In humans, then, the causal explanatory system typically reserved for reasoning about the underlying causes of manifest behavior in the self and other organisms—mental state attribution—is also evident when natural events are explained—implicitly or explicitly—as being intentionally driven by a non-natural agent. This "existential theory of mind" occurs such that certain important life events are perceived as possessing meaning independent of the self's imposition of meaning upon them. To interpret the effects of the natural world in this way is not a trivial cognitive feat, yet so far as I can tell, it has been either glossed over in the comparative religious literature or simply taken to be self-evident and not discussed at all (but see MITHEN 1996 for a related argument). No ontogenetic or phylogenetic models have been erected to chart its developmental pathways, nor empirical work done to determine how it comes into being, so speculation must reign absolute until precise mechanisms are made clear through experimental means.

Behavior– and Event–Reading: Shared Assumptions, Different Domains?

However, given that religion, from the cognitive biological stance, shares its central features with the comparatively mundane cognitive components underlying mental state attribution in the social world of other natural organisms, the processes that involve finding meaning in events lend themselves relatively well to such speculation. Currently, the field of cognitive developmental psychology is pregnant with studies on the nature and trajectory of children's understanding of mind (for recent reviews, see CADINU/KIESNER 2000; FLAVELL 1999; WELLMAN/CROSS/WATSON 2001), and the past several decades have seen tremendous advancements in knowledge in this area. While it is not yet entirely clear how young infants understand the overt actions of agents, nor how culture institutes sequential changes throughout development, it *is* clear that, by early childhood, humans regularly reason about the hidden causes of behavior, including their own. With this capacity, a focal behavior (i.e., an action that occupies online attention) is automatically perceived as symbolic of agent intentions, and thus becomes automatically enriched with meaning. Cross-cultural analyses show that this form of social cognition, referred to by various authors as 'mindreading' (WHITEN 1998), 'mentalizing' (HAPPÉ 1995), 'mental state attribution' (HEYES 1994), 'second order representation' (POVINELLI/BERING/GIAMBRONE 2000), 'metarepresentation' (LESLIE 1987), 'intuitive psychology' (CAREY 1995), 'folk psychology' (HARRIS 1992), 'MACHIAVELLIAN intelligence' (BYRNE/WHITEN 1991), and, most broadly, 'theory of mind' (PREMACK/WOODRUFF 1978; WELLMAN 1990; WIMMER/PERNER 1983), is fundamental to the species.[2] For purposes of the current paper, *behavior–reading* seems the most apt term, as all of these things (usually) deal with reasoning about overt action and making sense of behavior. In contrast, *event–reading* should therefore refer to the cognitive processes involved in interpreting, explaining, and predicting events wherein events are envisioned to arise through intentional causation.

Because theism involves reasoning about the mind of a non-natural agent, some theorists (e.g., BERING 2001; BOYER 2000; KIRKPATRICK 1999; MASER/GALLUP 1990) have posited that religion is a cognitive byproduct of the basic theory of mind module, which is widely envisioned as a response to complex social group living. However, actions and events are not synonymous. According to ZACKS/TVERSKY (2001, p3), an event is "a segment of time at a given location that is conceived by an observer to have a beginning and an end". Events thus possess *boundedness* as an inherent feature, and are therefore able to be divided into particulate units as are general action schemes (see VALLACHER/WEGNER 1987). Yet "many of the events people observe are actions, but many are not. Actions are performed intentionally by actors, so they are less general than events. Second, actions occur *objectively* in the world, whereas, for our purposes as psychologists, events arise in the *perception* of observers" (ZACKS/TVERSKY 2001, p4, italics added). While behavior and life event might be entwined together such that they tap the same causal explanatory system of intentional agency, it should not be taken as a matter of fact that their interpretive processes are products of the same evolutionary or developmental history. Developmentally, one (e.g., behavior) might be subjected to modularization as a product of crucial ontogenetic experiences (KARMILOFF-SMITH 1992), whereas the other (e.g., event) may encounter relatively innate patterns of activation and is less sensitive to extrauterine environmental contingencies. Evolutionarily, finding meaning in one (e.g., event) might hinge upon the presence of the autobiographical self, while the other (e.g., behavior) requires only a rough episodic memory to invoke a search for meaning, or vice versa. The self-memory system probably has its own complex phyletic history (POVINELLI/CANT 1995; SUDDENDORF/CORBALLIS 1997). It would be premature, to say the least, to comment further on possible differences between the two categories of phenomena. What is important, at this early stage of thinking, is to show that they share the same basic explanatory system, but perhaps operate according to different principles of activation. The generality of the mechanism questions the validity of arguing for the reserved rights of a dedicated modular system trading exclusively in the domain of behavior.

Chimpanzees' Understanding of Intentionality

Chimpanzees, it seems, are competent in the general symbolic domain. Does this mean that they, too, perceive certain self-relevant natural events, such as the death of their offspring, or the fortuitous discovery of already opened *Coula* nuts, as meaningful? Apes' use of gestures shows that they can use communicative symbolic codes to affect behavioral, if not intentional, change in others (e.g., JENSVOLD/GARDNER 2000); they are capable of representing numerosity through abstract symbolic means (see BOYSEN 1993); they are able to engage in inductive inference when confronted with scale model, hidden reward, tasks (KUHLMEIER/BOYSEN/MUKOBI 1999); and they are competent users of ideographic language systems based on mostly arbitrary lexigrams (see SAVAGE-RUMBAUGH 1986).

However, all of these competencies, grounded in basic symbolic abilities, are shared with human toddlers and hinge upon an observable connection between symbols and their referents, either perceptual iconicity (in the case of the scale model tasks) or goal-behavioral connectedness (in the case of chimpanzee language use). In this sense, they might be more appropriately referred to as *signal skills*, as they have more to do with capitalizing on physical cues and manifest behavior than hidden states. Chimpanzee American Sign Language, for example, more often than not involves the rigid implementation of communicative symbols to achieve a predictable, and desirable, behavioral end in human experimenters in the apes' pursuit of hedonic goals (e.g., chase, drink, berry, etc.) (SAVAGE-RUMBAUGH 1986; TERRACE/BEVER 1980). Yet after decades of research, and not for a lack of trying, there is still very little evidence that signing apes are capable of symbolic representation of non-ostensive states. Mental state terms, for instance, are seemingly Greek to these animals, quite possibly because they have no means to explicitly represent cognitive experiences such as intentions, desires, beliefs, and knowledge. (How can one teach an organism without metarepresentational abilities the referents of words like "think" or "feel"?) We will be waiting indefinitely, it seems, for language-trained chimpanzees to inform us of what it is really like to be a chimpanzee; in a very real sense, even they may not know.

It is when inductive inference involves hidden properties embedded in propositional relationships that apes, it seems, face difficulties in deciphering the "stands for" meanings of symbolic perceptual displays. Yet how can this be, when reports of intentional deceit, teaching, imitation, and cultural traditions in wild chimpanzee communities are now well documented (see WHITEN et al. 1999)? How can a 'mindblind' organism, for instance, intentionally instruct a naïve conspecific on how to perform a culturally encapsulated ritual if it has no cognitive access to the unobservable properties of "knowledge" and "ignorance", or the semiotic correlates of communicative behavior? To be sure, I am not the first to ponder these questions; the issue of whether chimpanzees have a theory of mind has been the subject of much rancor and division among contributors to the field (HEYES 1998; POVINELLI et al. 2000; SUDDENDORF/WHITEN 2001). Controlled experiments conducted in laboratories have almost unanimously provided either unintelligible or null results, and while the perpetual argument that one cannot infer evidence of absence on the basis of absence of evidence will keep most critics of these experiments satisfied, the scientific provision of ample opportunities to display incontrovertible evidence of mental state attribution (e.g., POVINELLI/EDDY 1996), and chimpanzees' recurring failure to provide such evidence in the laboratory, raises serious doubts that they are proficient at inferring the unobservable causes of, at least, *human* behavior.

Evidence from exquisitely controlled research on chimpanzees' understanding of other chimpanzee minds, however, indicates that they may have some understanding of unobservable causation when the stakes are high and the conditions express those of the environment of evolutionary adaptedness. Recently, HARE and his colleagues (HARE/CALL/AGNETTA/TOMASELLO 2000; HARE/CALL/TOMASELLO 2001) have published a series of papers on the "competitive conspecific paradigm"—a convincing methodological approach to primate social cognition that offers a promising glimmer of *some* theory of mind capacities in chimpanzees (HARE 2001). This paradigm pits chimpanzee *against* chimpanzee in competition for a desirable, monopolizable food item, an experimental situation with high ecological validity and one quite different from the strikingly unnatural experimental designs involving humans *cooperating with* chimpanzees to obtain such items (cf. CALL/HARE/TOMASELLO 1998; POVINELLI/EDDY 1996; POVINELLI, REAUX, BIERSCHWALE/ALLAIN/SIMON 1997). Under these more naturalistic—but still tightly controlled—conditions, chimpanzees have demonstrated firm evidence of at least an implicit understanding of the perceptual state of seeing in other apes, opting for food choices that are out of sight (and therefore out of mind!) of dominant conspecifics over ones in full view of such foes.

Findings such as these call into question the logic behind asking whether or not chimpanzees have a "theory of mind" (TOMASELLO/CALL 1997). There may be precursor systems in the brains of our closest genetic relatives that incorporate some of the basic principles of mental state attribution but that do not necessarily involve the explicit level of secondary representational abilities present in 4-year-old humans. Indeed, there is some evidence that a modified version of the competitive conspecific paradigm yields positive evidence of theory of mind skills in 2–3-year-old children (CHANDLER/FRITZ/HALA 1989; for review, see WELLMAN et al. 2001). In any event, it is too soon to tell whether other species are capable of representing mental states, and it would be premature to assert with any justifiable confidence that we have a complete picture of chimpanzee minds. Consciousness (perhaps) aside, there is little reason to assume that chimpanzee brains are any less complex and ecologically adapted than those of humans, and we should therefore not believe *a priori* that there are no such things as individual differences, age effects, or environmental contingencies related to acquisition of adaptive mental competencies, issues that are not typically addressed in primate laboratory studies but which comprise the empirical foundation of human cognitive science.

Implications and Ethological Interpretations

The importance of the question of whether chimpanzees are able to "metarepresent"—represent representations—should not be underestimated; when firmly in place, such a capacity might be deeply woven into the social fabric of more ancient behavioral systems such that it permits critical hallmarks of human cognition, such as language and culture (POVINELLI et al. 2000; TOMASELLO 1999). What does this mean for a discussion of the phylogenesis of religion? By postulating a central, shared mechanism for metarepresentation, the current uncertainty as to whether chimpanzees understand behavior in terms of underlying mental states, or whether they operate on the grounds of implicit representation of such states, is tremendously important for investigations into the biological bases of religion. If such a mechanism is present in chimpanzee minds, it suggests that an explanatory system built upon non-natural agency might, in fact, exist in the minds of other species (MASER/GALLUP 1990). While there is just as much reason, if not more, to suggest that this is not the case (see BERING 2001), in light of recent—but limited—positive findings of an understanding of unobservable causality in chimpanzees, it still remains to be seen. If the requisite cognitive architecture is lacking, it is difficult to imagine how other species would be able to acquire, transmit, or represent religious concepts—as such concepts, more often than not, hinge upon the basic ability to represent agent intentions. In the case of religion, those agents, if made culturally explicit, are of gods, ancestral spirits, and demons (BOYER 2000). If not fleshed out by culture, those agents remain fuzzy-faced and implicit, but nevertheless are perceived to act through such non-material belief structures as fate, immanent justice, and luck.

Several scholars have already ventured to explain the behaviors of wild chimpanzees as possible evidence of religion in the species. GOODALL (1975), for instance, first reported 'rain-dancing'[3] behavior—piloerection and general threat behavior in the face of rainstorms—in male chimpanzees under the rubric of a 'proto-religion' and was confident that it could shed light on the emergence of spiritual concerns in the human species. It has been claimed, for instance, that such activity might be evidence of animism, wherein the chimpanzees are attributing intentions to the storms and attempting to drive them off through elaborate threat displays (GUTHRIE 1993). Yet such an interpretation is not necessary to explain the behavior. Accompanying the storms is a range of environmental effects, such as loud noises, shaking branches, and apparently self-generated movement, which also occurs in the presence of predators or rival conspecifics. Given that it is only adult males that demonstrate 'rain-dancing' (see WHITEN et al. 1999), and given that only adult males engage in threat displays in normal, dominance asserting social interactions, it is possible that the physical effects of the storms on the environment simply activate canalized responses to specific environmental input (BERING 2001). The behavior is still poorly understood, however; convergent ethnographies of several different chimpanzee communities have since shown that 'raindancing' regularly occurs across a wide geographical range, but, interestingly, argue for cultural components in that it is not observed in all communities (WHITEN et al. 1999).

Concluding Remarks

Research suggesting that chimpanzees do possess, however limited, some conception of mind is controversial but slowly mounting, yet this is an area rich in relevant data for those interested in the biological foundations of the cognitive underpinnings of religious belief and behavior. In addition, paradigms testing more directly the central thesis of event–reading can be developed (although not necessarily easily) for use with apes. A possible dissociation in competency between behavior– and event–reading in species so closely related to humans would be especially significant, as it would appear to answer the question of which informational domain (behavior or event) first exploited an intentional causal explanatory system, and would militate against assumptions that the two are one and the same. In summary, it may be toward the chimpanzee's understanding of causality that will lead the scientist, if not necessarily the theologian, in the direction of 'truth.'

Author's address

Jesse M. Bering, Department of Psychology, Florida Atlantic University, Boca Raton, FL 33431-0991. Phone: 561-297-3360. Fax: 561-297-2160. Email: jber4317@fau.edu

Acknowledgements

I am grateful to David BJORKLUND and Benjamin BEIT-HALLAHMI for careful reading and editing suggestions on prior versions of this manuscript. Todd SHACKELFORD, Bennett SCHWARTZ, Jennifer YUNGER, and Robin VALLACHER also provided important commentary.

Notes

1 The use of the term 'non-natural' is used throughout this paper to describe those entities in religious belief systems that act as intentional agents. BOYER (1994) provides almost an encyclopedic listing of the qualities that make such agents "special"—but, most notably, shows that such entities violate natural laws (e.g., a tree that can hear, a mental agent that can attend to all things at once, a person who is invisible and thus does not conform to folk physics principles, etc.). Hence, non-natural implies that such entities cannot exist in any known natural ontological categories.

2 Collapsing these terms in such a manner may, of course, obfuscate subtle differences in meaning which distinguish them from one another. While the semantic nuances allowing each term its independence may be usefully highlighted in some theoretical applications, what is important in the current context is that the similarities are much more global than are the differences. All of these terms are used to describe a form of social cognition in which an individual ascribes mental states to the self and others in order to predict, interpret, or explain overt behavior.

3 The term 'raindancing' conjures up images of the human ritualistic activity bearing the same name. However, this connotation seems wildly misleading. The human ritual category involves the encoding of communicative intentions in the form of specific action displays that are learned and transmitted through cultural channels. Unlike chimpanzee 'raindancing,' which appears to be primarily induced by external stimuli and is a response to events, human raindancing is an attempt to affect change in events by deferring to an agent mediator.

References

Anderson, J. R./Gallup, G. G. (1999) Self-recognition in nonhuman primates: Past and future challenges. In: Haug, M./Whalen, R. E. (eds) Animal models of human emotion and cognition. APA: Washington DC, pp. 175–194.

Baron-Cohen, S. (1995) Mindblindness: An essay on autism and theory of mind. MIT Press: Cambridge MA.

Bering, J. M. (2001) Theistic percepts in other species: Can chimpanzees represent the minds of non-natural agents? The Journal of Cognition and Culture.

Bering, J. M./Bjorklund, D. F./Ragan, P. (2000) Deferred imitation of object-related actions in human-reared, juvenile great apes. Developmental Psychobiology 36:218–232.

Bjorklund, D. F./Pellegrini, A. D. (in press) Evolution and development: The emergence of evolutionary developmental psychology. APA: Washington DC.

Boesch, C. (1991) Teaching among wild chimpanzees. Animal Behaviour 41:530–532.

Boyer, P. (1994) The naturalness of religious ideas: A cognitive theory of religion. University of California Press: Berkeley CA.

Boyer, P. (2000) Functional Origins of Religious Concepts: Conceptual and Strategic Selection in Evolved Minds [Malinowski Lecture 1999], Journal of the Royal Anthropological Institute, 6: 195–214.

Boyer, P./Walker, S. (2000) Intuitive ontology and cultural input in the acquisition of religious concepts. In: Rosengran, K. S./Johnson, C. N./Harris, P. L. (eds) Imagining the impossible: Magical, scientific and religious thinking in children. Cambridge University Press: New York, pp. 130–156.

Boysen, S. T. (1993) Counting in chimpanzees: Nonhuman principles and emergent properties of number. In: Boysen, S. T./Capaldi, E. J. (eds) The Development of Numerical Competence: Animal and Human models. Lawrence Erlbaum: Hillsdale NJ, pp. 39–59.

Byrne, R. W. (1995) The thinking ape: Evolutionary origins of intelligence. Oxford University Press: Oxford.

Byrne, R. W./Whiten, A. (1991) Computation and mindreading in primate tactical deception. In: Whiten, A. (ed) Natural theories of mind. Basil Blackwell: Oxford, pp. 127–141.

Cadinu, M. R./Kiesner, J. (2000) Children's development of a theory of mind. European Journal of Psychology and Education 15:93–111.

Call, J./Hare, B./Tomasello, M. (1998) Chimpanzees gaze following in an object choice task. Animal Cognition 1:89–99.

Carey, S. (1995) On the origin of causal understanding. In: Sperber, D./Premack, D. (eds) Causal cognition: A multidisciplinary debate. Clarendon Press: New York, pp. 268–308.

Chandler, M./Fritz, A. S./Hala, S. (1989) Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' theories of mind. Child Development 60:1263–1277.

Dennett, D. C. (1987) The intentional stance. The MIT Press: Cambridge MA.

Flavell, J. H. (1999) Cognitive development: Children's knowledge about the mind. Annual Review of Psychology 50:21–45.

Goodall, J. (1975) The chimpanzee. In: Goodall, V. (ed) The Quest for Man. Praeger: New York, pp. 131–170.

Guthrie, S. (1993) Faces in the clouds: A new theory of religion. Oxford University Press: New York.

Happé, F. (1995) Understanding minds and metaphors: Insights from the study of figurative language in autism. Metaphor & Symbol 10:275–295.

Hare, B. (2001) Can competitive paradigms increase the validity of experiments on primate social cognition? Animal Cognition 4:269–280.

Hare, B./Call, J./Agnetta, B./Tomasello, M. (2000) Chimpanzees know what conspecifics do and do not see. Animal Behaviour 59:771–785.

Hare, B./Call, J./Tomasello, M. (2001) Do chimpanzees know what conspecifics know? Animal Behaviour 61:139–151.

Harris, P. L. (1992) From simulation to folk psychology: The case for development. Mind and Language 7:120–144.

Heyes, C. M. (1994) Anecdotes, training, trapping and triangulating: Do animals attribute mental states? Animal Behaviour 46:177–188.

Heyes, C. M. (1998) Theory of mind in nonhuman primates. Behavioral and Brain Sciences 21:101–134.

Kirkpatrick, L. (1999) Toward an evolutionary psychology of religion and personality. Journal of Personality 67:921–952.

Kuhlmeier, V. A./Boysen, S. T./Mukobi, K. L. (1999) Scale-model comprehension by chimpanzees (Pan troglodytes). Journal of Comparative Psychology 113:396–402.

Lawson, E. T./McCauley, R. N. (1990) Rethinking religion: Connecting religion and culture. Cambridge University Press: Cambridge.

Lenain, T. (1997) Monkey Painting. Reaktion Books Ltd: London.

Leslie, A. M. (1987) Pretense and representation: The origins of "theory of mind". Psychological Review 94:412–426.

Karmiloff-Smith, A. (1992) Beyond modularity: A developmental perspective on cognitive science. MIT Press: Cambridge MA.

Maser, J. D./Gallup, G. G. (1990) Theism as a by-product of natural selection. Journal of Religion 90:515–532.

Mitani, J. C./Merriwether, D. A./Zhang, C. (2000) Male affiliation, cooperation and kinship in wild chimpanzees. Animal Behaviour 59:889–893.

Mithen, S. (1996) The prehistory of the mind: The cognitive origins of art, religion and science. Thames and Hudson: London.

Povinelli, D. (2000) Folk physics for apes: The chimpanzee's theory of how the world works. Oxford University Press: Oxford.

Povinelli, D. J./Bering, J./Giambrone, S. (2000) Toward a science of other minds: Escaping the argument by analogy. Cognitive Science 24:509–541.

Povinelli, D. J./Cant, J. G. H. (1995) Arboreal clambering and the evolution of self-conception. Quarterly Review of Biology 70:393–421.

Povinelli, D. J./Eddy, T. J. (1996) What young chimpanzees know about seeing. Monographs of the Society for Research in Child Development 61:3.

Povinelli, D. J./Preuss, T. M. (1995) Theory of mind: evolutionary history of a cognitive specialization. Trends in Neurosciences 18:418–424.

Povinelli, D. J./Reaux, J. E./Bierschwale, D. T./Allain, A. D./Simon, B. B. (1997) Exploitation of pointing as a referential gesture in young children, but not adolescent chimpanzees. Cognitive Development 12:423–461.

Premack, D./Woodruff, G. (1978) Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences 4:515–526.

Savage-Rumbaugh, E. S. (1986) Ape language: From conditioned response to symbol. Columbia University Press: New York.

Sperber, D./Hirschfeld, L. (1999) Culture, cognition, and evolution. In: Wison, R. A./Keil, F. C. (eds) The MIT encyclopedia of the cognitive sciences. Cambridge MA: MIT Press, pp. cxi-cxxxii.

Stanford, C. B./Gambaneza, C./Nkurunungi, J. B./Goldsmith, M. L. (2000) Chimpanzees in Bwindi-Impenetrable National Park, Uganda, use different tools to obtain different types of honey. Primates 41:337–341.

Suddendorf, T./Corballis, M. C. (1997) Mental time travel and the evolution of the human mind. Genetic, Social, and General Psychology Monographs 123:133–167.

Suddendorf, T./Whiten, A. (2001) Mental evolution and development: Evidence for secondary representation in children, great apes and other animals. Psychological Bulletin 127:629–650.

Terrace, H./Bever, T. G. (1980) What might be learned from studying language in the chimpanzee? The importance of symbolizing oneself. In: Sebeok, T./Sebeok, J. (eds) Speaking of apes: A critical anthology of two-way communication with man. Plenum Press: New York, pp. 179–189.

Tomasello, M./Call, J. (1997) Primate cognition. Oxford University Press: New York.

Tomasello, M./Camaioni, L. (1997) A comparison of the gestural communication of apes and human infants. Human Development 40:7–24.

Vallacher, R. R./Wegner, D. M. (1987) What do people think they're doing? Action identification and human behavior. Psychological Review 94:3–15.

Weeks, M./Lupfer, M. B. (2001) Religious attributions and proximity of influence: An investigation of direct interventions and distal explanations. Journal for the Scientific Study of Religion 39:348–362.

Wellman, H. M. (1990) The child's theory of mind. MIT Press: Cambridge MA.

Whiten, A. (1998) Evolutionary and developmental origins of the mindreading system. In: Langer, J./Killen, M. (eds) Piaget, evolution, and development. Lawrence Erlbaum: Mahwah NJ, pp. 73–99.

Whiten, A./Goodall, J./McGrew, W. C./Nishida, T./Reynolds, V./Sugiyama, Y./Tutin, C. E. G./Wrangham, R. W./Boesch, C. (1999) Cultures in chimpanzees. Nature 399:682–685.

Wimmer, H./Perner, J. (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13:103–28.

Zacks, J. M./Tversky, B. (2001) Event structure in perception and cognition. Psychological Bulletin 127:3–21.

Zihlman, A. (1996) Reconstructions reconsidered: Chimpanzee models and human evolution. In: McGrew, W. C./Marchant, L. F./Nishida, T. (eds) Great ape societies. Cambridge University Press: Cambridge, pp. 293–304.

Jaak Panksepp

On the Subcortical Sources of Basic Human Emotions and the Primacy of Emotional-Affective (Action-Perception) Processes in Human Consciousness

An Introduction to the Following Reprinted Article

THE FOLLOWING REPRINTED ARTICLE was originally written as a contribution to the Conference on 'Emotion, Cognition and Evolution' held at the Konrad LORENZ Institute at Altenberg, Austria (17th–19th June 1999). Because of delays and ambiguities in publication, it was re-crafted as a submission for the inaugural issue of *Consciousness & Emotion* edited by Ralph ELLIS and Natika NEWTON (PANKSEPP 2000a). A related contribution, translated into German, will appear in the forthcoming proceedings of the above meeting (see WIMMER/CIOMPI 2002). In short, this reprinted paper is how the English version of that manuscript finally evolved. When the request was made to republish that article in the Konrad LORENZ Institute journal *Evolution and Cognition*, I wished to refresh some arguments, and that is the aim of this short introductory piece.

The area of emotion research is moving at a rapid pace. Although a coherent consensus has yet to emerge in the literature, the aim of my article was to steer thinking toward what I believe is the most empirically defensible position. I believe the analysis of higher emotional issues in humans is best constructed upon an understanding of *basic* emotional systems—ones we share with other animals, which evolved in a much deeper evolutionary time than is usually considered in most conceptualizations. Obviously the emotional systems we share with other animals are expressed very differently because of the comparatively recent emergence of our massive cognitive abilities. For instance, our capacity to regulate our behaviors, cognitions and emotions is more sophisticated than that achievable by other species. However, at a deep affective level, our raw emotional feelings may be remarkably similar to those of other mammals; the more substantial differences may be in the way basic emotional systems interact with the cognitive abilities of each species. Some of us are convinced that such perspectives can help generate a new evolutionary psychology built upon a sound cross-species foundation as opposed to the remarkably anthropocentric ones that presently prevail (PANKSEPP/PANKSEPP 2000, 2001). There were three major points I wished to make in the reprinted article, which I will expand upon briefly in this introductory piece.

First, the task of a coherent synthesis would be much easier if we accepted that other mammals not only have emotional responses but also basic affective experiences, both of which emerge substantially from evolutionarily homologous action systems concentrated in subcortical regions of the brain. These systems operate largely through evolutionarily ancient *instinctual action-to-perception* principles, which I believe constitute foundational mind/brain functions that regulate *perception-to-more-voluntary-action* processes that have arisen evolutionarily from more recent abilities of highly encephalized cortico-cognitive regions of the brain/mind. I advocate the position that through the detailed neuro-behavioral study of these more ancient emotional action–perception systems (see also PANKSEPP 2000b) the most penetrating and useful knowledge concerning emotions will emerge—for instance, the discovery of new medications to treat excessive affective dysregulations that characterize many psychiatric problems.

Second, I argue for the likelihood that the classic distinction between affective and cognitive processes makes evolutionary sense, and that it also allows us to focus empirical efforts on that most important and most neglected aspect of emotional psychology, the fundamentally affective nature of primary-process consciousness. In my estimation, clarification of the fundamental neural nature of affect is essential for major progress in the field. A distinction between evolutionarily ancient emotional–affective subcortical processes (largely constructed around *action–perception* principles) and more recently evolved cortico-cognitive capacities (constructed largely around *perception–action* principles) allows us to focus on the deep nature of affective processes rather than simply focussing our efforts on here-and-now experiential cognitive issues where the light seems brightest. By respecting that distinction, we may also develop better and more efficient strategies to evoke and study the types of emotion–cognition interactions that emerged in the human brain–mind during ensuing eras of neocortical expansion (CIOMPI 1997; PANKSEPP 1999). It seems likely that the generation of cortico-cognitive strategies to cope with certain types of emotional arousal have their own consistent patterns (epigenetically derived 'affect-logics') that deserve to be studied harmoniously in the context of the pre-existing ancestral processes we share with other animals. In other words, basic emotion and motivation systems approaches to mind provide natural categories (see BUCK 1999; PANKSEPP 1998) for potentially distinct expressions of affect-logic that are organized around and energized by different feeling states.

Third, in this massively computational age, it is important to consider that the most fundamental truths about emotions must be derived through organically-based research programs rather than the mathematical–informational modeling approaches that continue to increase in popularity because of the availability of powerful personal computers. I see little reason to believe that affective processes emerge from computational complexities rather than from organic ones. For instance, will it really be possible to compute *the feeling* of sadness, etc? In my estimation, the cognitively focused computational myth, by draining resources from the necessary brain–behavior work, may be retarding substantive scientific work in the field. I would briefly elaborate on these issues:

1. Affect Is Largely a Subcortical Process

Many investigators and theoreticians remain skeptical about the fundamental role of subcortical systems in the elaboration of affect, partly because they feel consciousness is only a characteristic of the human species. In addition, modern brain imaging studies, which contrary to the evidence long provided by animal brain research, have highlighted so many higher cortical regions in the emotional processes aroused by exteroceptive stimuli (for summaries, see LANE/NADEL 2000). I believe many of those studies are yielding deceptive findings, at least for understanding the nature of affect: Investigators, by using perceptually driven methodologies, are typically visualizing the cognitive components of emotional processing rather than core affective processes. Fortunately, an increasing number of experiments published during the past year have been more concordant with the animal data. I would briefly share my favorites:

Perhaps the most compelling evidence comes from DAMASIO, et al. (2000), who asked individuals to achieve deep, existentially experienced feeling states of anger, fear, sadness and happiness via personal reminiscences. When subjects truly experienced those feelings, radioactive water was infused and PET images were constructed. The results affirmed abundant arousals in subcortical brain regions, accompanied by substantial reductions of blood flow in many higher brain areas, suggesting a narrowing of information processing in neocortical systems during intense emotional states.

Various other studies have also highlighted the importance of subcortical regions in human affective experiences such as air hunger (LIOTTI et al. 2001), the taste of chocolate (SMALL et al. 2001), the appetite for various rewards including winning money (KNUTSON et al. 2001), the sex-specific appeal of pretty faces (AHARON et al. 2001), the pleasure of musical peak experiences (BLOOD et al. 2001), male sexual arousal (REDOUTÉ et al. 2000), as well as the sexually differentiated experience of rectal distention (KERN et al. 2001). All of these studies report arousals of subcortical brain areas implicated in the generation of affect by animal research, as well as those ancient archicortical zones, especially orbitofrontal, anterior cingulate and insular cortices that MACLEAN (1990) originally highlighted in his Limbic System concept (which has been increasingly attacked by a growing number of cognitive neuroscientists more accustomed to working on the higher informational functions of the brain). In short, the extended limbic system, descending deep into the medial diencephalon and upper brainstem does comprise the fundamental neuro-geography of affective experience in humans. Accordingly, we should be devoting much more effort to studying the details of those systems in appropriate animal models, since the homologous systems are largely inaccessible for *causal* human research.

In this vein, we should also recall that emotional feelings have typically been much easier to activate in humans through stimulation of subcortical circuits that mediate the instinctual emotional behaviors of our fellow animals, than through higher brain stimulation (for reviews see HEATH 1996; PANKSEPP 1985). Perhaps the most recent striking effects have been BEJJANI et al.'s (1999) observation of sudden onset of depression by stimulating midline diencephalic structures near the subthalamic nuclei. It could easily be argued that affective experience emerges only when these ancient systems interact with higher cortico-cognitive structures, such as working-memory fields of the dorso-lateral prefrontal cortices (e.g., LEDOUX 1996) or uniquely human linguistic-symbolic abilities (ROLLS 1999). However, those viewpoints seem to put a new and almost dualistic onus on higher neo-cortical structures to create affect out of practically nothing but individual human, rather than cross-species, evolutionarily inherited potentials for certain experiences. I strongly urge us to consider, with our traditional respect for parsimony, to resurrect the long-neglected possibility that affect is an organically-embodied part of subcortical instinctual–emotional systems that arouse basic *action-to-perception* processes. Such evolutionary views can serve as a solid platform for conceptualizing the developmental emergence of more recently evolved cognitive *perception-to-action* processes. We should remember that the capacity to have experiences is certainly not the same as the capacity to talk about them.

A compelling set of data for a subcortical locus of control for affect in human neuropsychology comes from the famous 'Patient B' (ADOLPHS/TRANEL/DAMASIO 2002). Although Patient B has severe damage to most of the structures of the higher limbic system (including orbitofrontal cortex, anterior cingulate and insula), he can still identify most emotions from *dynamic*, albeit not static, visual stimuli (showing only a severe deficit for disgust). To the extent that one can evaluate internal experiences in severely amnesic, hippocampally-impaired individuals, Patient B still feels affective states (personal communications with Ralph ADOLPHS). Although that may reflect readout of 'as if' bodily processes within Patient B's intact somatosensory cortices (DAMASIO 1994), I would submit that it is more likely to reflect the fact that his deeper subcortical emotional systems remain essentially intact.

2. Cognitive–Affective Distinction

Obviously emotions and cognitions interact, but if the perspective advocated in the previous section is on the right track, then the following conclusion is unavoidable: Emotional action-to-perception processes, and their accompanying affects, are more ancient in brain evolution than most of the perceptual, thinking and planning processes that are typically subsumed under the concept of cognition. If we accept the transitive properties of evolution across species, then we may be able to usefully distinguish affective and cognitive neural processes in the animal brain (PANKSEPP 2002), and animal brain research might be recognized as a most powerful way for us to understand the fundamental neural nature of basic emotional feelings in humans as well as other animals. Although all human feeling are obviously accompanied by various cognitive activities, scientifically we may be well advised to accept that the classic distinction between emotions and cognitions is a realistic way to parse psychological space and thereby to incisively dissect key research questions into meaningful units. Although cognitivists are prone to see emotions as just another type of cognition, such lines of reasoning merely reflect a 'cognitive imperialism' which wishes to rule by semantic-fiat as opposed to providing robust strategies for understanding the essentially neurobiological nature of affect.

I personally do not understand how traditional 'dust-bowl'-cognitivism, focussing just on external information processing, can tell us much about the fundamental nature of thirst and hunger, even though it can, with the right kind of shift in research priorities, tell us a great deal about the thoughts and strategies we have when we are hungry and thirsty. Thus, increasing trends to cognitivize intellectual territories where they have made few fundamental contributions, while disregarding long-standing intellectual traditions that have tackled emotional issues head on, seems to be an imperiously unproductive modus operandi. For, the time being, the top-down 'Cognitive Neuroscience of Emotions' approaches (LANE/NADEL 2000), should be countered by bottom-up 'Affective Neuroscience of Cognition' ones. Of course, a synthesis of the two perspectives is desirable. Working together they may yield more robust and productive strategies than either alone (for a more complete critique of such issues, see PANKSEPP 2002). Of course, considering that the 'mind science' community is currently so vastly 'cognitivized', the possibility of a rapid paradigm-shift to focus on organic foundations of affect in animal models seems remote.

Keeping such issues in mind, I would argue that the cognitive–informational aspects of emotions are best studied in the context of an emerging 'affect-logic' (CIOMPI 1997; WIMMER/CIOMPI 1996) that has yet to be fully developed, at least at an empirical level. 'Logic' in this context means that different emotions and moods promote different patterns of thought. This has, of course, long been evident when strong bodily motivations are aroused: The various hungers and thirsts lead to obsessive strategies for acquiring more and more and hoarding those resources. Even sodium-hunger enlivens the cognitive apparatus to dwell on the taste and the search for salt. Many investigators (my favorite classic being Alexander Shand whose ideas are highlighted in the following article), have proceeded to codify some of the possible relations. But now this analysis has to be taken to a new level of sophistication, where the details of specific cognitive–affective interactions need to be worked out empirically. For instance, even though it is obvious that feelings of anger can arouse thoughts of retribution—to reach out and strike someone who has thwarted one's desires or offended one's sense of propriety, do we really know how the brain–mind begins to dole out blame, to obsessively seek out the apparently external causes of offense? Do we yet know the time courses of such events and the best ways to counter such primitive mental energies? An enormous amount of basic work remains to be done at the cognitive level.

3. Affects Are Organic rather than Computational Processes

During the past few decades, the influence of computationalism has been growing more influential in all areas of psychology, while the import of deep neuro-organic views seems to be diminishing (albeit modern brain imaging is a hybrid exception). Partly, this is simply due to the fact that it is much easier to play with the software of computers than to acquire the 'wet-ware' skills necessary for animal brain research. Of course, opportunities for research on the inner workings of animal brains are quite scarce, especially now that one has to run the gauntlet of increasingly difficult approval committees at many research institutions. However, there is no way that the computational views can ever replace the organic approaches, even though they may certainly help refine our hypotheses. The grand assumption of computationalism that general-purpose, substrate-neutral 'algorithms' for most psychological processes, capable of being instantiated via stepwise iterations on computers, remains attractive to many, even among those who hold deeply evolutionary perspectives (see PINKER 1997). Those viewpoints are shortsighted when it comes to the basic nature of emotions and motivations.

Of course, all aspects of the conscious mind must be deemed 'complex systems' and scholars have struggled mightily to bring some kind of mathematically ordered structure to the understanding of complex systems (e.g., HOLLAND 1995). However, perhaps we have been fooled that organic nature can be computed, and if it is not computational (as HOLLAND and so many other complexity scholars seem to believe it is), then we really need to discuss emergence in rather different, and perhaps more humble, ways. Verbal/imagery metaphors (see LAKOFF/JOHNSON 1995) may be more important than equations in the mind/brain sciences (indeed, we should begin to recognize equations as just one precise and explicit exemplar of metaphor). I suspect there is still a massive 'physics envy' in cognitive science, where it is assumed that mathematical metaphors can be as successfully deployed for unraveling the nature of mind as they have been for deciphering inanimate nature. Unfortunately, there are no spectacular success stories (as there were in physics) that point to the unique fertility of that assumption. Accordingly, observation of nature rather than computation of its shadows should remain the prime enterprise of scientists and other natural philosophers. Attempts to empirically establish supervenience linkages between levels of organization may be essential for any type of grander complex systems theory to emerge.

Thus, I am of the mind that on the practical side, computationalism has been massively, and all too successfully, oversold in the mind/brain sciences. It has accordingly been sucking intellectual energies away from 'the bench' at an increasingly distressing level during the past three decades. I say this sympathetically, with the perspective that when I was a young scientist studying the processes of body energy balance regulation, I also went through a phase of computational enthusiasm (PANKSEPP/RITTER 1975; PANKSEPP 1973, 1978). However, the impact of all that for a lasting understanding of the key mechanisms of energy balance in the brain was miniscule. In retrospect it was largely a waste of time and effort that generated no powerful *new* ideas about organic nature, despite being an excellent way to summarize data and highlight the rigor of one's ideas. Meanwhile, molecular biological discoveries have been decisive in helping us formulate a lasting knowledge that has many therapeutic implications for appetite control disorders (CHIESI/HUPPERTZ/HOFBAURER 2001).

In my estimation, the likelihood that we will ever truly understand affective processes via computationalism seems remote. For instance, how shall we ever compute the experience of an orgasm? No matter how many facially expressive robotics and useful emotional detection tools are derived through those approaches, the assumption that one has injected feelings into robotic 'minds' has to remain an implausible assumption/illusion unless one can make new and powerful predictions about the nature of human feelings and hopefully their biological underpinnings. Clearly, new 'Turing Tests' will be needed to evaluate the affective nature of artificial minds to properly evaluate any future computational 'solutions.'

Where might computationalism really add something to understanding emotions? Since affective states are created in the brain by specific neurochemical systems that help establish global field dynamics with certain virtual 'shapes' (reflecting, perhaps, whole-body instinctual action tendencies), computationalism could have an important role in the topographical-mathematization of the fluctuating dynamics of those virtual neurodynamic 'shapes'. Of course, the empirical analysis of those brain processes may need to be pursued (which remains quite primitive) before effective computational modelling can succeed. I also tend to believe that computationalism may have an important role in decoding sensory, perceptual and cognitive processes (where the repetitive, monotonous micro-structurings of neocortical columns do have some superficial resemblances to RAM-type computer chips—certainly much more so than is evident in subcortical emotional systems). But even in the analysis of such cognitive processes, computationalism remains, after decades of effort, more of a promissory note than a major achievement in revealing the true nature of mind.

In this vein, it is noteworthy that one of the fathers of computationalism, Jerry FODOR has now shared a dire testimonial about the cognitive-computational revolution he fostered (see his 2000, 'The Mind Does Not Work That Way', which, of course, is a cheeky parody of Steven PINKER's anti-organic, pro-computationalist 'How the Mind Works'): As FODOR put it: "you might think that cognitive scientists would be worried a lot about the limitations of the Classical computational theory of mind. Speaking for myself, I'm worried half to death. In fact, it seems to me, much of the field is in deep denial" (p39), and "so far, what our cognitive science has found out about the mind is mostly that we don't know how it works" (p100). I suspect this is because the full force of bio-emotional and motivational processes, and the resulting developmental interactions with maturing cognitive systems, has really never been incorporated into the prevailing cognitive paradigms. Thus, it may be wise to ponder whether the ever increasing investments in computationalism within psychology and neuroscience will only give us fancier toys as opposed to knowledge of fundamental importance for understanding mind.

In Sum

The following article (PANKSEPP 2000a) summarizes my attempt to deal with the mystery of affective processes. By doing this, I do not in any way wish to minimize the importance of cognitive issues, but to open up intellectual space for dealing more vigorously with the deeper affective mysteries of the mind that are being tackled by only a few in human mind sciences (e.g., BUCK 1999). Obviously, our most unique human mental abilities reside in our remarkable capacity, though the aid of massive reiterating memory and symbolic capacities, to project behavioral strategies forward and backward in time.

However, I suspect that when the temporal distributions of mind activities are fully described across many individuals who are not required to do anything else, it will be evident that most folks spend exorbitant amounts of time dwelling on emotional and motivational concerns—on affective issues that were laid down as 'evolutionary urges' in the genetically dictated circuits of very ancient regions of their animal brains. Although we can plan our activities and

Author's address

Jaak Panksepp, Chicago Institute for Neurosurgery and Neuroresearch, Evanston, Illinois, and J. P. Scott Center for Neuroscience, Mind and Behavior, Department of Psychology, Bowling Green State University, Bowling Green, Ohio, USA.
E-mail: jpankse@bgnet.bgsu.edu

hence our futures with more sophistication than other animals, our feelings still remain like centers of gravity around which our cognitive apparatus tends to revolve—unless, of course, all our needs are satisfied.

Accordingly, in order to understand many cognitive processes, we will have to come to terms with the fundamental neurobiologically affective nature of our minds. Such issues should be of ultimate concern for a lasting science of human nature. Regrettably, there presently seems to be little consensus on the urgency and importance of such issues in our institutions of higher learning. Perhaps a little common sense, along with the global emotional traumas that are facilitated by the fantasy-revering cultures we have created (like the one we experienced since September 11th of this year), will help change that.

References

Adolphs, R./Tranel, D./Damasio, A. R. (2002) Dissociable neural systems for recognizing emotions. Brain and Cognition, in press.

Aharon, I./Etcoff, N./Ariely, D./Chabris, C. F./O'Connor, E./ Breiter, H. C. (2001) Beautiful faces have variable reward value: fMRI and behavioral evidence. Neuron 32:537–551.

Bejjani, B.-P./Damier, P./Arnulf, I./Thivard, L./Bonnet, A.-M./ Dormont, D./Cornu, P./Pidoux, B./Samson, Y./Agid, Y. (1999) Transient acute depression induced by high-frequency deep-brain stimulation. The New England Journal of Medicine 340:1476–1480.

Blood, A. J./Zatorre, R. J. (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Sciences 98:11818–11823.

Buck, R. (1999) The biological affects, a typology. Psychological Review 106:301–336.

Chiesi, M./Huppertz, C./Hofbauer, K. (2001) Pharmacotherapy of obesity: Targets and perspectives. Trends in Pharmacological Sciences 22:247–254.

Ciompi, L. (1997) The concept of affect logic. An integrative psycho-socio-biological approach to understanding and treatment of schizophrenia. Psychiatry 60:158–170.

Damasio, A. R. (1994) Descartes' error: Emotion, reason, and the human brain. Avon Books: New York.

Damasio, A. R./Grabowski, T. J./Bechara, A./Damasio, H./ Ponto, L. L. B./Parvizi, J./Hichwa, R. D. (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. Nature Neuroscience 3:1049–1056.

Fodor, J. (2000) The mind doesn't work that way: The scope and limits of computational psychology. MIT Press: Cambridge UK.

Heath, R. G. (1996) Exploring the mind-brain relationship. Moran Printing Inc: Baton Rouge LA,

Holland, J. H. (1995) The hidden order, How adaptation builds complexity. Addison-Wesley: Reading MA.

Kern, M. K./Jaradeh, S./Arndorfer, R. C.,/Jesmanowicz, A./ Hyde, J./Shaker, R. (2001) Gender differences in cortical representation of rectal distension in healthy humans. American Journal of Physiology—Gastrointestinal and Liver Physiology 281: G1512-G-1523.

Knutson, B./Adams, C./Fong, G./Hommer, D. (2001) Anticipation of monetary reward selectively recruits nucleus accumbens. Journal of Neuroscience 21:RC159, 1–5.

Lakoff, G./ Johnson, M. (1999) Philosophy in the flesh: The embodied mind and its challenges to western thought. Basic Books: New York.

Lane, R. D./Nadel. L. (Eds.) (2000) Cognitive neuroscience of emotion. Oxford University Press: New York.

LeDoux, J. E. (1996) The emotional brain. Simon & Schuster: New York.

Liotti, M./Brannan, S./Egan, G./Shade, R./Madden, L./Abplanalp, B./Robillard, R./Lancaster, J./Zamarripa, F. E./ Fox, P. T./Denton, D. (2001) Brain responses associated with consciousness of breathlessness (air hunger). Proceedings of the National Academy of Sciences 98:2035–2040.

MacLean, P. D. (1990) The triune brain in evolution: Role in paleocerebral functions. Plenum Press: New York.

Panksepp, J. (1973) A reanalysis of feeding patterns in the rat. Journal of Comparative and Physiological Psychology 82:78–94.

Panksepp, J. (1978) Analysis of feeding patterns: Data reduction and theoretical implications. In: Booth, D. A. (ed) Hunger models: Computable theory of feeding control. Academic Press: London, pp. 143–166.

Panksepp, J. (1985) Mood changes. In: Vinken, P. J./Bruyn, G. W./Klawans, H. L. (eds) Handbook of Clinical Neurology, Volume 45: Clinical Neuropsychology. Elsevier Science Publishers: Amsterdam, pp. 271–285.

Panksepp, J. (1998) Affective neuroscience: The foundations of human and animal emotion. Oxford University Press: New York.

Panksepp, J. (1999) Emotions as viewed by psychoanalysis and neuroscience, An exercise in consilience Neuro-Psychoanalysis, 1:15–38.

Panksepp, J. (2000a) The neuro-evolutionary cusp between emotions and cognitions, implications for understand consciousness and the emergence of a unified mind science. Consciousness & Emotion 1:17–56.

Panksepp, J. (2000b) Affective consciousness and the instinctual motor system: The neural sources of sadness and joy. In: Ellis, R./Newton, N. (eds) The caldron of consciousness, motivation, affect and self-organization, Volume 16: Advances in Consciousness Research. John Benjamins: Amsterdam, pp. 27–54.

Panksepp, J. (2002) At the interface of the affective, behavioral and cognitive neurosciences: Decoding the emotional feelings of the brain. Brain and Cognition, in press.

Panksepp, J./Ritter, M. (1975) Mathematical analysis of energy regulatory patterns of normal diabetic rats. Journal of Comparative and Physiological Psychology 89:1015–1028.

Panksepp, J./Panksepp, J. B. (2000) The seven sins of evolutionary psychology. Evolution and Cognition 6:108–131.

Panksepp, J./Panksepp, J. B. (2001) A continuing critique of evolutionary psychology: Seven sins for seven sinners, plus or minus two. Evolution and Cognition 7:56–80.

Pinker, S. (1997) How the mind works. Norton: New York.

Redouté, J./Stoléru, S./Grégoire, M.-C./Costes, N./Cinotti, L./Lavenne, F./Le Bars, D./Forest, M. G./Pujol, J.-F. (2000) Brain processing of visual sexual stimuli in human males. Human Brain Imaging 11:162–177.

Rolls, E. T. (1999) The brain and emotion. Oxford University Press: Oxford UK.

Small, D. M./ Zatorre, R. J./Dagher, A./Evans, A. C./Jones-Gotman, M. (2001) Changes in brain activity related to eating chocolate: From pleasure to aversion. Brain 124:1720–1733.

Wimmer, M./Ciompi, L. (1996) Evolutionary aspects of affective-cognitive interactions in the light of Ciompi's concept of 'affect-logic'. Evolution and Cognition 2:37–58.

Wimmer, M./Ciompi, L. (2002) Emotion, Kognition und Evolution. Interdisziplinäre biologische, psychologisch–psychopathologische und soziale Aspekte. Filander Verlag: Fürth, in press.

Jaak Panksepp

The Neuro-Evolutionary Cusp Between Emotions and Cognitions

Implications for Understanding Consciousness and the Emergence of a Unified Mind Science[1]

Concepts without factual content are empty; sense-data without concepts are blind. The senses cannot think, The understanding cannot see. By their union only can knowledge be produced.—Immanuel Kant, *The Critique of Pure Reason*

Abstract

The neurobiological systems that mediate the basic emotions are beginning to be understood. They appear to be constituted of genetically coded, but experientially refined executive circuits situated in subcortical areas of the brain which can coordinate the behavioral, physiological and psychological processes that need to be recruited to cope with a variety of survival need (i.e., they signal evolutionary fitness issues). These birthrights allow newborn organisms to begin navigating the complexities of the world and to learn about the values and contingencies of the environment. some of these systems have been identified and characterized using modern neuroscientific and psychobiological tools. The most fundamental emotional systems can now be defined by the functional psychobiological characteristics of the underlying circuitries—characteristics which help the organism coordinate behavioral, physiological and psychological aspects of emotionality, including the valenced affective feeling states that provide fundamental values for the guidance of behavior. The various emotional circuits are coordinated by different neuropeptides, and the arousal of each system may generate distinct affective/neurodynamic states and imbalances may lead to various psychiatric disorders. The aim of this essay is to discuss the underlying conceptual issues that must be addressed for additional progress in understanding the nature of primary process affective consciousness.

Key words

Brain, emotions, consciousness, affect, cognition, emotional disorders, neuropeptides, psychiatric implications.

Susanne Langer (1951) in contemplating the status of emotional processes in her own time, worried that everything that falls outside of the domain of analytic, propositional, and formal thought is merely classed "*as emotive, irrational, and animalian… All other things our minds do are dismissed as irrelevant to intellectual progress: they are residuals, emotional disturbances, or throwbacks to the animal estate*" (p246, my italics in this and all other quotes). Several generation later, Joe LeDoux (1996), "the leading expert on the emotional brain" (Gazzaniga et al. 1998, p516), promoted a modern variant of that intellectual tradition by suggesting that: *"The brain states and bodily responses are the fundamental facts of an emotion, and the conscious feelings are the frills that have added icing to the emotional cake"* (p302). In essence, LeDoux, as well as many other neuroscientists believe that affective experiences—those intangible and presumably unmeasurable subjective events —are of little importance for a scientific understanding of emotions.

My personal view on these issues is quite different: I accept the importance of all of the above levels of analysis, but would also suggest that an understanding of affective processes in both humans and other mammals should be deemed an essential ingredient for the field to consider. In accepting the likelihood that the basic emotional feelings are fundamental representations of complex, causally efficacious, organic processes within mammalian brains—emergent properties that are realized in the dynamic organizations of neuronal networks—I have called for my colleagues to "*consider one simple bit of logic: If affective feelings do exist in the minds of other organisms and have causal consequences on their behaviors, we will never adequately un-*

derstand their brains, or our own for that matter, unless we incorporate various new functional concepts into our thinking." (PANKSEPP 1999, p164).

In short, I believe that a neural understanding of emotional feelings—those apparent underlying regulators of many behavioral choices—remains one of the most important topics for our science and our society to pursue. In part, this position is based on the recognition that for any lasting understanding, complex brain phenomena need to be viewed from several mutually complementary perspectives. Neurophysiological terminology is not sufficient to conceptualize many global brain processes. The neurobiological nature of feelings can be scientifically approached through the conjoint cross-species implementation of *behavioral, psychological* and *neural* perspectives—namely, through a triangulation that is essential for the pursuit of a substantive *affective neuroscience* (PANKSEPP 1998a).

At the same time, I, along with most investigators in the field, accept as given that substantial amounts of emotional processing within the brain (e.g., unconditional responses) are achieved by neural networks that in themselves probably elaborate no conscious emotional feelings. Indeed, in line with FREUD's original suggestion, most investigators now agree that much of what goes on in the brain is dynamically on automatic pilot and unconscious. At the same time it seems likely that a great deal of time and effort in brain evolution was devoted to the establishment of intrinsic values—the various feelings of goodness and badness that are internally experienced indicators of survival utility, elaborated within ancient regions of the brain shared by all mammals in remarkably homologous ways. This is not to say that these emergent feelings are not thoroughly biological in their essential underlying form, but to accept that psychology has a major role to play in unraveling the nature of such processes within the brain. The simple fact that other animals avidly consume and get "hooked" on the same drugs as humans—becoming dependent on molecules that promote essentially similar neuropsychological processes in all mammals—provides one robust line of support for such a thesis. The essential substrates for such desires are subcortical (IKEMOTO/PANKSEPP 1999; MCBRIDE/MURPHY/IKEMOTO 1999; WISE 1996).

There are many other equally compelling lines of evidence that investigators who do not wish to confront the central issue of human and animal feelings, and other aspects of their psychological lives, commonly choose to ignore. At the same time, the pursuit of substantive knowledge in this area is remarkably difficult, and perhaps the prevailing agnostic view is a preferred tactic for various socio–political reasons (i.e., sustaining consensus and grant support). But it is ultimately not a wise path: All too commonly it fails to consider the whole corpus of evidence on such topics. It can undermine our ability and willingness to confront the reality of the emergent neurodynamics that may constitute primitive forms of consciousness. It also promotes a neodualism that is harmful for scientific understanding—sustaining a division of mind and matter introduced by DESCARTES for outdated religious and political reasons. The unified nature of basic emotional processes across all mammalian species—with essential subjective and objective aspects that must be studied conjointly—is a monistic perspective that I would encourage all to accept. My aim here will be to share an overview of how affective neuroscientific strategies for understanding mind could help us correct our neglect of those affective processes that Susan LANGER and other thoughtful observers of the human/animal condition encouraged us to consider during the past century.

Toward a confrontation with affective consciousness: background issues

Let me first affirm once more that it has long been obvious to critically minded observers that many of the emotional acts that humans and animals exhibit reflect no conscious intent. Many impulsive emotional acts are projectile, reflexive responses to environmental stimuli, and the rules for those responses are ingrained within seemingly straightforward types of neuronal circuits. For instance, animals do not learn to exhibit a startle response to a sudden sound or a withdrawal response to sudden pain or rapidly approaching threats. These responses are much closer to reflexes than intentional actions, even though a process is set in motion in the brain that can govern future actions. However, there is an intermediate class of emotional action tendencies that do contain a germ of intentionality—for instance, the tendency of animals spontaneously to seek resources. Although no higher level conscious intent is required for animals to become appetitively energized when confronted by an abundance of various environmental rewards, their behavioral engagements have more complex neural underpinnings than those that govern stimulus-bound reflexes. The underlying SEEKING system provokes animals to exhibit a flexible appetitive presence in the world. Many emotional responses reflect "intentions in action" to use SEARLE's (1983) discriminating terminology, even though they may not constitute "intentions to act" (which may require higher cognitive processes). Emotional feelings, I believe, are realized more in the neural substrates of the former than the later.

The exploratory and investigatory behaviors of animals seeking resources have an outward character suggesting that they emerge spontaneously from certain ingrained types of neural organization—see IKEMOTO/PANKSEPP (1999) for a most recent review. Such "instinctual" behaviors flow as naturally as a river cascading down a waterfall. Such spontaneous emotive behaviors have a flexibly characteristic *presence* suggesting they do represent the fundamental urges of an organism. There are many emotive behavior patterns such as this in the intrinsic behavioral repertoires of all species. The fact that such basic emotional action tendencies arise spontaneously from an intrinsic form of neural organization does not automatically mean that they do not have immediate repercussions on a primary-process form of consciousness. There are good reasons to believe such behaviors arise from neural systems whose substrates constitute the very foundation of all subsequent forms of consciousness: If these systems are damaged, the adaptive competence of animals is severely compromised. To the best of our knowledge, these neural systems create the experiential immediacy of an internally felt presence in the world, a proposition that can be tested in humans, the question being—when the underlying systems are artificially activated does internal experience have a quality of belongingness or one of artificial imposition?

Because such "intentions in action" are not created by "intentions to act," many behavioral neuroscientists, including those prominently interested in emotions, have chosen to remain skeptical of the possibility that other animals have consciously experienced emotional feelings. Indeed, many assert that if such mental faculties do exist, they may have little to do with the way brains control behavior. As already noted, LEDOUX (1996) has been a prominent advocate of such an epiphenomenalist perspective in cognitive neuroscience. In my estimation, *unsubstantiated disbelief*—skepticism about reasonable possibilities that have substantial empirical support—can be as much of a barrier to scientific progress as *unsubstantiated belief*. I would submit that the concept of emotional feelings is not in the later category, but that agnosticism on the issue *is* rapidly becoming an exemplar of the first. In areas such as emotion research, we should not remain eternally silent on such matters, as some behavioral neuroscientists would prefer. That is rapidly becoming a potentially immoral stance, as we recognize that the probable existence of emotional states in other animals is very high. In any event, to understand the brain, we must be willing to entertain the reality of various psychological processes, as created through poorly understood neurodynamics.

Certain investigators, as committed to a rigorous scientific understanding of the fundamental nature of emotions as the skeptics and agnostics, feel that there is no rational alternative but to seriously consider the existence of a primordial form of affective consciousness in other organisms and to analyze the role of such processes in behavioral choices (BUCK 1999; DAMASIO 1999; PANKSEPP 1998a, 1998b, 2000a). The weight of evidence is simply too large for us to ignore the possibility that affective feelings are fundamentally created by brain systems that generate "intentions in action" rather than simply being created by the higher associative and language abilities of the human brain, as claimed by some prominent investigators (ROLLS 1999). Here I will advance the view that primary-process "affective consciousness" is constructed fundamentally from the intrinsic capacities of certain neural circuits—the basic emotional systems of the brain—which operate in goal-directed and valenced ways, working in harmony with basic, thoroughly biological, self-representational systems of the brain.

These distinct points of view—one asserting that the topic of emotional subjectivity, especially in other animals, is simply not workable from any credible scientific perspective, and the other asserting that it is—are presently in deep conflict. In my estimation, the argument against animal feelings comes ultimately from an unforgiving, anthropocentric form of solipsism combined with a pernicious form of neo-dualism. It is remarkable that it is supported by so many neuroscientists, for there is no well-argued data base supporting that view… only the philosophical residue of CARTESIAN dualism. The evidence for animal subjectivity comes from i) an enormous number of approach and avoidance tests, ii) various consummatory choice and stimulus preference studies that have been conducted on many other mammals, and iii) the remarkable homologies in the neuroanatomies and neurochemistries for such affective tendencies in all mammals. The weight of existing evidence (although there is no "knockout" *final proof*—an impossible task in science) is that other animals do have internally experienced feelings that have consequences for their behaviors.

The ultimate resolution of this issue should be of considerable interest to investigators of the human mind, especially since an understanding of the "higher" forms of awareness may be critically dependent on our ability to understand some of the "lower" substrates of felt existence. Indeed, there is now an increasing enthusiasm to deal with those subtle brain processes we know as moods and feelings, which appear to be part of the genetic birthright we share with many other creatures, for they have powerful influences on the way our cognitive activities operate and hence in all aspects of the way we live our lives (DAMASIO 1994, 1999). These are the systems that create a foundation of meaning for higher life decisions. Some of us believe that a true understanding of the organization of mind and ground of being must be premised on a neuroscientific probing of those ancestral value-processes that evolution provided to help complex creatures like mammals navigate successfully through the world.

From an evolutionary perspective, honed by the remarkable recent advances in molecular biology, it is now certain that many of our fundamental abilities are remarkably similar to those of our brethren animals. The underlying "mechanisms/processes" can only be understood if we are willing to simultaneously take several perspectives to the organized nature of complexity—with one critical but often neglected one being a data based cross-species, experiential point of view. We can probably understand the nature of human hunger by studying the subcortical energy regulatory systems of rats. We can do the same for thirst, anger, fear, and the many other vexations and pleasures of the shared, primitive regions of mammalian brains.

A psychobiological confrontation with these ancient emotional systems, and the intrinsic values they create (as monitored indirectly via the various approach and avoidance behaviors animals exhibit), shall be of foremost important in decoding how consciousness first emerged on the face of the earth (PANKSEPP 1982, 1998b). Through the ability of emotional systems to conditionally encode every-day activities with values, many of our cognitive activities remain tethered to affective principles. As many have suspected, we tend to approach things *because* they have made us feel good (in the various ways that is possible), and we avoid things because they make us feel bad. Other animals presumably operate essentially in the same manner, even though the cognitive strategies we use to fulfill our desires and to avoid our travails are surely more sophisticated and long-sighted than in most other animals. Of course, the different species often employ very different sensory, motor and cognitive tools to achieve emotional and motivational homeostasis.

This naturalistic view of human and animal existence lost credibility abruptly with the success of the behaviorist revolution early in the 20th century and it was sustained by the subsequent advent of digital computational models of mind and the emotion-free cognitivism of the second half of the century. As a result of those schools of thought, which marginalized the importance of emotional and motivational feelings in the governance of human and animal lives, several generations of thinking along lines advocated here were lost. Only recently are some returning to reconsider such evolutionary roots of mental existence.

The issues I shall focus on here were aired by previous generations of thinkers (e.g., COGAN 1802; DARWIN 1998; FREUD 1981b; SHAND 1920 just to name a few), and they are re-emerging once again to the forefront of evolutionary and neuro-epistemological thought (DAMASIO 1999; MACLEAN 1990; PANKSEPP 1998a, 1998b). My basic premise here will be that the evolution of higher brain mechanisms was critically guided by the preexisting neurobiological exigencies of organisms (i.e., subcortical emotional and motivational abilities), which are generally more similar among living mammalian species than their higher cortico-cognitive functions which have diverged more considerably (see BUDIANSKY 1998 and HAUSER 2000, for a recent overviews of cognitive differences). However, even there we will find a great deal of convergent evolution because of the basic needs all animals share. In sum, my guiding premise is that a knowledge of the "lower" affective functions will constitute essential substrates for understanding the operation of higher brain–mind functions: Unless we come to terms with the deeply organic nature of our basic drives—the various emotions and motivations—we may never understand the multi-faceted nature of consciousness(es).

Evidence concerning many of the basic emotional systems we share with the other mammals has recently been summarized (PANKSEPP 1998a). Here I will largely focus on the types of higher brain dynamics (or cognitive "modules" in the debatable parlance of modern evolutionary psychology) that may exist in human cortico-cognitive areas that establish many of our more sophisticated behavioral priorities. In general, it will be important to have formal proposals, as well as empirical tests, of specific types of "affect-logic" that emerge from higher regions of the mind that have evolved to deal with basic emotional and motivational issues (CIOMPI 1997; WIMMER 1995; WIMMER/CIOMPI 1996).

There are presently three major strands of experimentally-based neurobiological thought in this area: i) One strand is emerging from modern evolutionary psychology which is postulating mind–brain "modules" that arise all too often from an "arm-chair" Pleistocene-oriented logical analysis of human mind and behavior (TOOBY/COSMIDES 2000). ii) Another is emerging from modern cognitive neuroscience, which generally takes a massive corticocentric focus, that often seems to be deny, or at least ignore, the existence of powerful emotional forces in the deep recesses of the brain–mind (e.g., GAZZANIGA et al. 1998; GAZZANIGA 2000). iii) Finally, the third and most solidly evolutionary approach is highly focussed on subcortical issues and has arisen from the recognition that the basic emotions may reflect organizational principles at the very foundations of the mammalian mind (DAMASIO 1999; MACLEAN 1990; PANKSEPP 1998a). The three could work well together toward a comprehensive mind science, but that will require better understanding, appreciation and integration of each others' premises and data bases than presently exists. The philosophical community is also becoming remarkably interested in such issues (e.g., GRIFFITHS 1997), and investigators should become immersed in *all* the available scientific evidence rather than constraining themselves to the most prominent human psychological tradition that was heavily based on a facial-analysis of emotions that emerged in the 70s (e.g., EKMAN 1998). The animal neurological traditions should not be ignored by philosophers, for that is the only way we can resolve the foundational processes.

My aim here will be to provide a historical overview of some of the above issues and to push forward the idea that one way we can come to understand the natural order of the human mind is to clarify a variety of interrelated themes that arise from the ancestral nature of the brain–mind: i) to decode the basic nature of the biological value-generating systems that are built into mammalian brains as ancestral birthrights, ii) to discuss how these and related systems actually generate "valence-tagging" of previously neutral perceptual events; iii) to understand how the aforementioned interactions govern more complex layers of thinking and perceiving; iv) to suggest how experimental work on the basic affective processes of the brain may interweave with those subtle brain process(es) generically known as consciousness; v) to discuss how the above lines of inquiry may have important implications for understanding the essential nature of volitional activities and free will, as well as vi) the understanding psychiatric disorders. This will be followed by vii) some examples of how the intrinsic plasticity in the underlying system may establish temperamental states and habitual ways of being within organisms. After providing overviews of each of these topics, I will conclude with my personal views on the possibility that affective processes will ever be simulated computationally.

I. Biological Value Encoding Processes of the Brain

I have recently summarized the nature of brain emotional systems both in synoptic (PANKSEPP 1982, 1991, 2000a, 2000b) and comprehensive archival ways (PANKSEPP 1998a), including recent chapters on separation distress (PANKSEPP et al. 1988), play (PANKSEPP 1993b), fear (PANKSEPP 1990) and seeking systems (IKEMOTO/PANKSEPP 1999), which are the specific emotional processes that have been the focus of my research during the past three decades. I will not aspire to any detailed coverage of facts here, but will simply highlight the main conceptual themes that the current evidence supports, including a revitalized form of psychoanalytic thinking.

A synopsis of affective neuroscience

In general, the executive emotional systems are conceived to generate a variety of internally experienced affective states and related "evolutionary operants" or instinctual behavioral tendencies that emerge from widespread brain systems that have at least 6 attributes: As I suggested in 1982, they are able to 1) directly evaluate the meaning of certain sensory inputs (e.g., the smell of predators in prey species); 2) they modulate attentional and sensory-motor sensitivities relevant for the evoked behavioral tendencies (e.g., hunger sensitizes olfactory acuity); 3) they control diverse physiological and hormonal conditions of the body which bring many organ systems in line with the concurrent behavioral demands (e.g., adrenaline secretion is adaptive for all behaviors that require motor arousal); 4) they sustain animals in specific feeling (mood) states for relatively prolonged periods of time (e.g., separation protest vocalizations and feelings of distress are typically be sustained until social reunion occurs or despair sets in). All of the above are also 5) modulated by various cognitive activities (i.e., appraisals can provoke emotions) and which 6) also modulate cognitive activities (i.e., emotions channel thoughts and code memories). The manner in which the seventh major attribute, affective experience, is generated remains most mysterious of all, but there is a great deal of relevant data that can guide our thinking. My best estimate is that centromedial mesencephalic systems, such as those of the periaqueductal gray (PAG) are absolutely essential (BANDLER/KEAY 1996; PANKSEPP 1998b), even though these system are in strong interaction with higher brain areas such as the cingulate, frontal and insular cortices, which surely elaborate felt emotional experiences. For a summary of details concerning this system, see WATT (1999b). In any event, the notion that affect is an irrelevant issue for understanding how the brain controls behavior will hopefully soon become a minority view, even among those rigorous animal behaviorists who have been trained to avoid any tinge of anthropomorphism. Now that we know how much we share genetically, behaviorally and probably psychologically, such issues need to be evaluated on a case by case basis rather than dismissed by fiat. Obviously, this strategy is bound to succeed more in the analysis of primitive brain systems that all mammals share homologously rather than higher systems where there has been much more evolutionary divergence.

I doubt if many investigators of the relevant system would claim that there are no intrinsic emotional systems in the brain as many did just a few years ago. It is now certain that the brain contains a variety of genetically ingrained emotional systems for generating specific classes of emotional behaviors (PANKSEPP 1998a). To all appearances, affective experience is a rather direct manifestation of the arousal of these systems. When these systems are electrically stimulated, humans report urges to act and describe emotional experiences that have a feeling of belongingness, as opposed to being alien to the self. Presumably, the arousal of the various emotional command systems could be distinguished subjectively from each other by humans as being fundamentally distinct feelings, but such issues were never addressed during the era when such brain stimulation studies were most actively pursued (see PANKSEPP 1985 for review). Of course, the essential role of these primitive systems should not be taken to mean that the higher cortical projection areas have no role in experienced feelings. The clinical evidence indicates that they most certainly do, but largely in a modulatory/regulatory capacity. The core mechanisms for affect appear to be subcortically situated.

The brain emotional "command" systems that have been provisionally identified in experimental animals, along with their major anatomies and neurochemistries are summarized in Table 1. I will not attempt to provide any more detail in the limited space available here, especially since they have been thoroughly summarized recently (PANKSEPP 1998a). However, I would emphasize that we are only on the near shore of substantive human work in this area and even critical animal work on such issues remains quite meager because there is currently little institutional support for work which is attempting to fathom how affect is elaborated within the mammalian brain. Most still believe that such issue reside in the realm of intuitive hunches rather than the predictive landscape of mainstream science. Considering the existence of cross-species affective neuroscience type research strategies, that bias is blatantly incorrect. In any event, there would have to be a major shift in both our research priorities and strategies for this type of work to proceed at a reasonable pace. For now, I would simply emphasize that the type of detailed knowledge of the underlying neural substrates that needs to be obtained simply can not be achieved without behavioral brain research in other animals, along with careful evaluation of permissible manipulations in humans (e.g., KNUTSON et al. 1998). I personally believe this kind of knowledge would be invaluable for a new and hopefully highly humanistic phase of biological psychiatry where psychoanalytic approaches become, perhaps for the first time, widely used tools for new types of inquiries into the psychodynamics of the human mind.

In any event, now that we know a great deal about these intrinsic emotional systems in the brains of other mammals, we could (at least in theory) arouse a variety of distinct emotional tendencies in humans by artificial means (i.e., various types of direct brain stimulation), and ask how those systems contribute both to behavioral choices and mental states. Of course most such work is ethically problematic, unless retired neuroscientists interested in such issues were more willing to be guinea pigs for future inquiries. However, to the extent that we can selectively arouse such systems using peripheral pharmacological maneuvers, we could validate that the behavioral indices we utilize in animals are not leading us astray. The development of molecules that can activate specific neuropeptidergically orchestrated emotional systems may eventually allow us to evaluate very discrete possibilities rather directly. They may also lead to a to a new revolution in biological psychiatry where very specific feelings can be modulated by

Basic Emotional Systems	Key Brain Areas	Key Neuromodulators
General + Motivation SEEKING/Expectancy	Nucleus Accumbens—VTA Mesolimbic and mesocortical outputs Lateral hypothalamus—PAG	DA (+), glutamate (+), many neuropeptides, opioids (+) neurotensin (+)
RAGE/Anger	Medial amygdala to Bed Nucleus of Stria Terminalis (BNST). medial and perifornical hypothalamic to dorsal PAG	Substance P (+), ACh (+) , glutamate (+)
FEAR/Anxiety	Central & lateral amygdala to medial hypothalamus and dorsal PAG	Glutamate (+), many, neuropeptides, DBI, CRF, CCK, alpha–MSH, NPY
LUST/Sexuality	Cortico-medial amygdala, Bed nucleus of stria terminalis (BNST) Preoptic and ventromedial hypothalamus Lateral and ventral PAG	Steroids (+), vasopressin, & oxytocin, LH–RH., CCK.
CARE/Nurturance	Anterior cingulate, BNST Preoptic Area, VTA, PAG	oxytocin (+), prolactin (+) dopamine (+), opioids (+/–)
PANIC/Separation	Anterior Cingulate, BNST & Preoptic Area Dorsomedial Thalamus Dorsal PAG	opiods(–), oxytocin (–) prolactin (–) CRF (+) glutamate (+)
PLAY/Joy	Dorso-medial diencephalon Parafascicular Area Dorsal PAG, Tectum	opioids (+/–), glutamate (+) ACh (+), Any agent that promotes negative emotions reduces play

Table 1. General summary of the key neuroanatomical and neurochemical factors that contribute to the construction of basic emotions within the mammalian brain. The monoamines serotonin and norepinephrine are not indicated since the participate in non-specific ways in all emotions. The higher cortical zones devoted to emotionality, mostly in frontal, cingulate, insular, and temporal areas, are not indicated. Key: CCK = choleocystokinin, CRF = corticotrophin releasing factor, DBI = diazepam binding inhibitor, ACh = acetylcholine, MSH = Melanocyte Stimulating Hormone, NPY = Neuropeptide Y.
– inhibits prototype, + activates prototype. (Adapted from PANKSEPP 1998a and WATT 1999a)

pharmacological influences on specific affect systems of the brain. These lines of thought raise the possibility of some very interesting *psychoethological* work that could be pursued in humans, perhaps implemented with psychoanalytic free-associative, active listening procedures as outcome measures.

I single out the psychoanalytic tradition because it was a line of 20th century thought that continued to accept the importance of affective processes in our attempts to understand the mind. Even though psychoanalytic approaches were not robust enough to reveal the internal nature of emotions, FREUD did develop a provisional conceptual scheme—of id, ego and superego functions—where the drive-like neural "forces" within the id were the primal powers around which the rest of the mental apparatus revolved. While FREUD realized that his system of thought depended critically on our ability to understand the id, he and his colleagues had no reasonable way to probe the internal neural structures of the various "drives" that constituted that conceptual brain–mind. However, FREUD presciently *suggested that the id, cut off from the external world, has a world of perceptions of its own*. I translate this to mean that affective processes, being very ancient in brain evolution, do not need cognitive structures in order to generate the psychological impact of raw feelings. However, without a substantive neuroscientific understanding of the id, the rest of FREUD's theory of psychology could not be adequately evaluated. Our recent understanding of basic emotional systems permits a potentially fruitful rapprochement between psychoanalytic and neuroscientific approaches to mind (KANDEL 1998; PANKSEPP 1999).

FREUD, no doubt, would be very favorably disposed to such views. As FREUD highlighted in his "Project" (1981a) when he discussed the pleasure and unpleasure of sexual release: "…a suspicion forces itself on us that… the endogenous stimuli consist of *chemical products,* of which there may be a considerable number." (p321). Although he neglected such issues for most of his career, toward the end of his life FREUD still asserted that "The future may teach us to exercise a direct influence, by means of particular chemical substances, on the amounts of energy and their distribution in the mental apparatus." (FREUD 1981c, p182). It would be an understatement to say that many such agents have now emerged, and there are many other possibilities in the wings. Indeed, at the heart of many of the brain's affect programs of the mammalian brain, "there are a variety of chemical codes (largely neuropeptidergic) that may eventually permit precise new modes of psychiatric intervention, and new ways to evaluate how feelings are constructed in the human brain" (PANKSEPP 1999, p44).

From this vantage, it would be important for us to be able to directly study the human mental processes on line—not just with direct brain measures such as EEG and fMRI but also psychoanalytically (i.e., free associative narratives in "normal" individuals induced to experience distinct moods). Although brain imaging techniques have given us remarkable snapshots of emotions—from psychostimulant craving (CHILDRESS et al. 1999) to sadness (GEORGE et al. 1996)—they have not yet added much to our understanding of how the dynamics of mind change during these states. Could the technique of free-association be utilized experimentally to reveal the affective dynamics of the human mind (DAHL 1998)? I suspect that a new field of inquiry, such as *psychoethology,* which would seek to characterize the normal affective-cognitive topography of the human mind under the sway of different emotions could help us scientifically characterize many basic psychological processes. But more on that later.

II. The Generation of Associative "Valence-Tagging"

I doubt if many investigators today would claim that human emotions are totally socially constructed. At the same time few would claim that social-construction of emotionality is either modest or negligible. Obviously, the genetically provided emotional systems of the brain, many of which become fully operative soon after birth in all mammalian species (Table 1), are continuously molded by organismic responses to new environmental objects and events. Brains can imbue initially neutral environmental events with values. This is commonly called valence-tagging or secondary/conditioned reinforcement—the ability of previously neutral events to assume the intrinsic values evoked by emotionally salient events (i.e., unconditional stimuli) through associative learning. Even though this may transpire by a variety of distinct mechanisms operating at several hierarchical levels within the brain/mind, leading to a complex developmental landscape, the simplest and most widely studied form is that arising from classical conditioning. This reflects one of the simplest forms of cognitive-emotional interaction that exists within the brain. To put it in everyday words—even though individuals at an instinctive level may know what they like and what they do not like (through unconditional pleasure-unpleasure responses), at a cognitive level they do not. However, through classical conditioning, cognitive systems learn quickly.

Let me also put this in more traditional psychological terms: The systematic pairing of neutral cues or conditional stimuli (CSs) with biologically important events or unconditional stimuli (UCSs) which spontaneously evoke instinctual behavioral and physiological changes, also known as unconditional responses (UCRs) can rapidly produce learning, or conditioned responses (CRs). The ability of the temporal pairing of CSs with UCSs to produce CRs, has been a staple of behavioral psychology since Pavlov systematized such knowledge. Now, there is widespread recognition that most emotional responses can be conditioned in this way. That has led to a cottage industry of behavioral researchers working to reveal the details of how fear responses condition in the amygdala (e.g., how tones and light paired with shock evoke conditioned withdrawal or or autonomic arousal responses). Generally, such investigators believe that the understanding of emotions is best achieved through the implementation of simple learning approaches that focus on associative learning issues as opposed to the intrinsic evolutionarily dictated nature of brain emotional systems. The techniques are very effective in both animals (LEDOUX 1996) and humans (ÖHMAN 1993), and they are well within the tradition of 20th century behavioral science.

Unfortunately, such work only tells us a great deal about how emotional responses can be molded by learning, but comparatively little about the intrinsic nature of the evolved emotional processes of the brain. The widespread use of such classical conditioning techniques has led to the recognition that much of emotional learning occurs at unconscious levels. That conclusion is generally accepted since many classically conditioned fear responses occur so rapidly that no subjectively experienced cognitive or affective processes is likely to have intervened between the presentation of a CS and the emission of the CR that is measured. However, it is rarely acknowledged that the long-term affective responses generated by such conditioning may also be influential in regulating the behavioral tendencies of animals. Indeed, the comparative neglect of long-term emotional responses within such research programs may now be promoting a misleading picture of the organization of emotions in the brain/mind, and the study of projectile classically conditioned responses needs to be supplemented by the recognition that affective feeling states are concurrently conditioned to the CSs, and that those states may have causal efficacy in the *sustained* regulation of subsequent behaviors.

For instance CSs that have been paired with painful stimuli, can subsequently intensify other fear responses commonly evaluated in the potentiated startle paradigm, and also promote longer term behavioral changes such as freezing. A study of this background fear is getting somewhat closer to the underlying affective process. It is now certain, as has been suspected for 30 years, that a hot-bed for such associative learning is in the local circuits of the lateral and central amygdala (LeDoux 1996), but it is gradually being realized that anxiety conditioning can occur in many other brain areas (Davis/Shi 1999; Maren 1999). The key synaptic chemistry which seems to promote both the conditioning and deconditioning of such associative responses are changes in glutamatergic transmission (Falls/Miserendino/Davis 1992). Indeed, it is likely that conditioning, or at least long-term sensitization (Adamec 1997), can occur along the whole length of emotional command systems (see previous section), so an enormous amount of work remains to be done before we truly understand how pervasive is the plasticity of this system along the neuroaxis.

Only modest headway has been made in implementing such strategies for the study of most other, especially positive, emotional processes. An exception is the recent work on the role of brain "reward" and dopamine systems in the conditioning of appetitive eagerness (for a recent reviews, see Ikemoto/Panksepp 1999; Schultz 1998). To highlight how effectively such processes can be used to study other subtle positive responses such as social "joy" or animal "laughter," Figure 1 summarizes a classical conditioning experiment with tickle-induced 50-KHz chirping in young laboratory rats (Panksepp/Burgdorf 1999). Of the four groups depicted (see legend), only the group with contiguous CS-UCS pairings exhibited a systematically incrementing pattern of learned behavior. In other words, the young rats chirped in anticipation of being tickled. We presently have no empirical knowledge of where such conditioning occurs, but the cingulate and orbitofrontal cortices, as well as nucleus accumbens, septal nuclei and bed nuclei of the stria terminalis (BNST) are likely places to look.

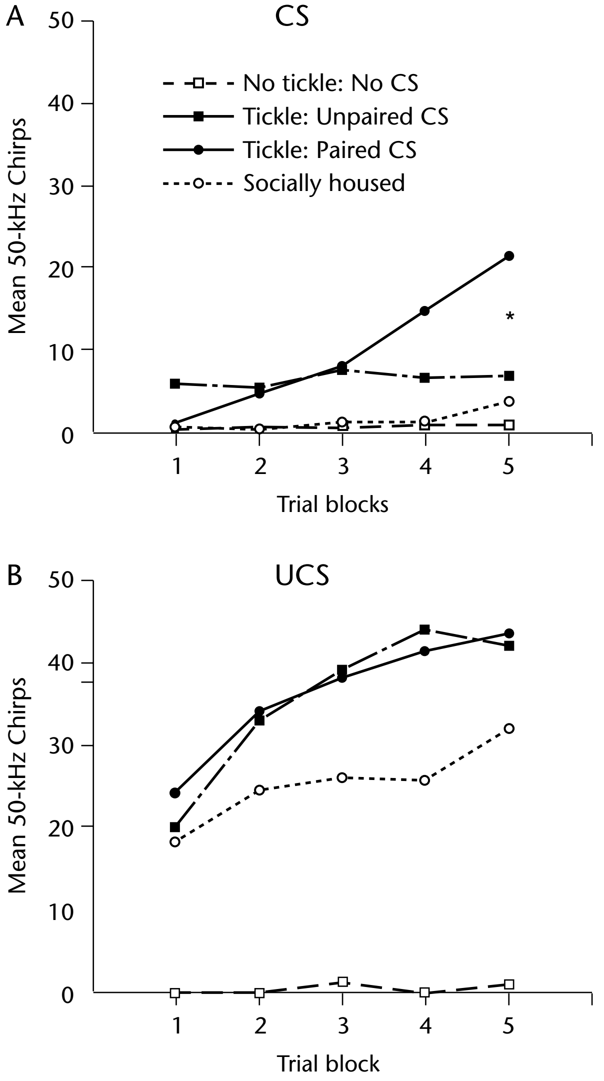

Figure 1. Mean (±SEM) levels of 50 KHz "laughter" type chirping during the first five trials of conditioning: "Tickle: Paired CS" animals were exposed to the conditioned stimulus—passive exposure to the the tickle hand—right before a 15 second period of tickling (data not shown, but rates of chirping were about 38 per 15 sec). The group of animals that received the CS followed immediately by tickling, exhibited significantly higher chirping rates than the other two control groups (data according to Panksepp/Burgdorf 1999).

Although the classical conditioning procedures tell us a great deal about how emotional values can be linked to neutral stimuli, we must also recognize that such approaches tell us practically nothing about the intrinsic nature of the emotional values that are mediated by the emotional command systems or UCR pathways (Table 1). To my way of thinking, that can only be achieved by the types of approaches highlighted in the previous section, as well as through various place-preference and place-avoidance conditioning procedures (for summary, see SCHECHTER/CALCAGNETTI 1993), as well as relevant operant learning tasks, with stringent schedules of reinforcement, to evaluate motivational strength. Animals seek out places where they have had positive affective experiences, and they avoid places where they have had negative ones.

Although many would like to believe that affective experiences occur within fairly high areas of the brain, such as neocortical zones that mediate working memory (LEDOUX 1996), the evidence so far seems to be that the affective content of experience can be elaborated quite low in the neuroaxis (e.g., OLMSTEAD/FRANKLIN 1997; PANKSEPP 1998a). Some may be tempted to suggest that such "affects" reflect unconscious processes, perhaps the "dynamic unconscious" postulated by FREUD, but I would rather view them as the essential foundations of consciousness. Consciousness must not only be conceptualized phylogenetically (CABANAC 1999; PANKSEPP 1990a), but also as ontogenetic processes of neuronal/psychological development. From this vantage, it is noteworthy that PET studies of infant brains have found much higher levels of metabolic activity in those primitive emotional areas of midbrain and diencephalon than in most cortical areas. However, with development, intense patterns of cortical activation gradually emerge (CHUGANI 1996). Are infants then unconscious, or are they simply operating primarily with primitive forms of affective consciousness? Pain studies of human infants tend to bear out the second alternative (ANAND 1997). The gradual development of working memory, with the ability to treat subcortical processes as tokens of information, presumably provides high-order regulation over emotional processes rather than constructing affect out of those inputs.

III. Emotions and Higher Order Psychological Processes

It is a straightforward tenet of folk-psychology that our emotions have robust effects on the way we think and what we think about. Because of such interactions, there is a regrettable tendency to conflate emotional and cognitive processes by people who do not work on the deep structures of the brain. Since so much of current work in experimental psychology is concentrating on higher cognition-emotion interactions in humans, and since animal work has comparatively little to say about those issues (i.e., thought processes are even harder than emotional ones to observe in animals—see HAUSER 2000), I will briefly cover one historical antecedent—the ideas of Alexander SHAND—that may help highlight reasonable ways to proceed at the human level. Then I will focus on one major conceptual issue—the nature of emotional *projections,* which may help us better understand how affective feelings interact with cognitive processes in very global ways. These lines of thought may help us establish some lawful relationship between emotions and cognitions.

First, let me indicate that three general laws of emotions that could be linked to a biological analysis were put forward by Charles DARWIN (1998). He suggested that each basic emotional system of the brain (i.e., his principle of *action, due to the constitution of the nervous system*), interacts with other systems (his principle of *antithesis*) and is also accompanied by the vast baggage of accumulated learning (his principle of *serviceable associated habits*). Contrary to modern investigators of emotions in animals, DARWIN was not hesitant to acknowledge that most probably a key feature of their emotional responses is a feeling tone.

Since then, there have been several attempts to codify the laws of emotions, as they operate at the psychological level (FRIDJA 1986) as well as how they operate at a deeper affective-logic level that has been related to psychiatric disorders (CIOMPI 1997). Unfortunately the empirical work has lagged far behind the general conceptualizations. To generate some simple straightforward experiments, it may be instructive for us to once more consider the systematization attempts of our predecessors, and I have been impressed by some of the ideas advanced by SHAND.

The "laws" of Shand

In 1920, Alexander SHAND published his monumental "*The foundations of character: Being a study of the tendencies of the emotions and sentiments*" which, in the midst of an impressive narrative, put forward 150+ laws that he believed accurately characterized human emotional feelings, their attending cognitions, and their various interactions. These "laws" were derived from everyday observations and personal insights rather than any systematic empirical analysis, but we should not dismiss them because of that. They still provide a source of many intriguing, empirically testable, hypotheses. I will only provide a sampling of his thought. His first few fundamental laws were those that he believed were the foundation for all other laws of character:

1) *"Mental activity tends, at first unconsciously, afterwards consciously, to produce and to sustain system and organization."* (p21)

2) *"Every primary impulse, whether it is independent or belongs to a primary emotion, is innately connected with the systems of fear, anger, joy, and sorrow, in such a way that, when opposed, it tends to arouse anger; when satisfied, joy; when frustrated, sorrow; and when it anticipates frustration, fear; these systems being similarly connected together."* (p38)

6) *"All intellectual and voluntary processes are elicited by the system of some impulse, emotion, or sentiment, and subordinated to its end."* (p67).

Although many of SHAND's laws seem straightforward, even self-evident, and hence perhaps not worthy of empirical analysis, there are good reasons they should be deployed for experimental studies. An empirically verified fact is worth a thousand reasonable assumption. Indeed, I am tempted to suggest that a new discipline, such as experimental philosophy, might be quite useful in this arena since experimental psychologists often seem not be temperamentally ready to tackle such issues. In any event, it they could be empirically substantiated, each of the following assertions could broaden and deepen our scientific understanding of basic emotional matters.

15) *"The joyful temper, in proportion as it is stronger than the ordinary disposition to joy, weakens sensibility to the opposite emotions of repugnance and sorrow, and by strengthening hope and confidence in the future, weakens the opposite emotions of despondency and despair."* (p153).

17) *"The joyful temper lowers the threshold of sensibility for joy, hope, and confidence, but raises it for sorrow, despondency, and despair."* (p.154).

20) *"The sorrowful temper lowers the threshold of sensibility for sorrow, despondency, and despair, but raises it for joy, hope and confidence."* (p154).

33) *"The universal end of Fear is merely to prevent the occurrence of some threatening event whether the danger be 'real' or 'imaginary.' "* (p215)

37) *"All varieties of anger tend to accomplish their ends by some kind of aggressive behavior."* (p250).

38) *"Fear and anger tend always to exclude one another, where both are referred to the same objects."* (p254).

69) *"Sorrow tends to be diminished by the knowledge that another sorrows with us."* (p341).

70) *"Sorrow tends to be increased by the knowledge that another rejoices at our suffering."* (p341).

Perhaps one reason such reasonable assertions never received much attention was because they utilized a host of affective concepts which were axiomatically accepted as emotional givens, with no replicable empirical evidence provided for their existence. Now that we are beginning recognize the neural circuits from which such feelings arise (PANKSEPP 1998a), we may finally be able to implement new research programs that try to highlight the types of intrinsic affective-cognitive regularities that are evident from an everyday folk-psychological perspective. One of the most workable general concepts is that of *projection*—the tendency of people to cast their feelings onto the world as if the world were the cause of their feelings.

Emotional projections into cognitive activities

It is now generally accepted that emotional and cognitive processes massively interact (GRAY 1990; PANKSEPP 1988, 1990b; PARROT/SCHULKIN 1993), and incisive empirical work on those topics is increasing (CHRISTIANSON 1992). Rather than detail those trends here, let me simply highlight one issue that may be empirically very workable—the ability of emotions to be projected onto objects and subjects in the world. Although the concept of "projection" was a mainstay of FREUDIAN theory that has been empirically neglected, its pervasiveness in human affairs should be receiving more prominent attention in the empirical analysis of how emotions and cognitions interact within the brain. It still has enormous psychiatric implications.

Although one could envision that "projection" would be largely a matter of how "valence tagging" between perceptions and cognitive impressions transpires (*vide supra*), it may be a more pervasive and dynamic a brain response than that. It may reflect direct actions of emotions on brain areas that mediate cognitive and perceptual processes. Since cognitive processes are designed to deal with moment by moment events in the external world, while affective feelings reflect evolutionarily provided value codes, it may be that the projection of feelings onto environmental events and objects was one of the simplest ways for evolution to persistently guide the perceptual priorities of the cognitive apparatus.

It is easy to imagine that this type of interaction operates through some type of global neurodynamic/neurochemical process in the brain, whereby basins and peaks of attractors mold the psychological landscape. It may reflect how emotions are "embodied" or broadcast widely in neural tissues rather than being informationally encapsulated linear programs such as those that operate in digital computers. Various widely acting neurochemical "spritzers" (e.g., norepinephrine and serotonin) as well as widely dispersed peptide system could be the substrates for these widespread effects in the brain.

If we correctly comprehend how the corresponding psychological processes are aroused neurobiologically, we should be able to develop major new ways to modify how people view the events of their lives and hence open up new avenues of psychotherapeutics. For instance, from the animal data, we could envision that sexual jealousy is aroused substantially by activation of brain vasopressin systems (WINSLOW et al. 1993). Many comparable psychobiological hypotheses concerning human mental dynamics could be generated from recent neuropeptide research in animals, (PANKSEPP 1993a, 1998a), which should eventually provide opportunities to modify specific emotions in fairly discrete ways, and to determine, mechanistically, how affects, and hence value priorities, are "projected" into the world. Some of the most dramatic forms of such projection will be found in disorders such phobias and post-traumatic stress disorders (VAN DER KOLK et al. 1996), and there are new ideas (e.g., "limbic permeability") how such processes emerge psychobiologically (ADAMEC 1997).

It is possible that many cognitive problems could be ameliorated simply by adjusting the underlying emotional feelings. Recent evidence for this comes from the widespread use of SSRIs (Serotonin Specific Reuptake Inhibitors) which dramatically reduces the tendency of people to experience negative emotions toward other people (KNUTSON et al. 1998). Indeed, marriages that have been on the verge of falling apart because of the negative feelings that spouses commonly project on each other have been saved by the ability of these agents spontaneously alleviate negative feelings, with no need for any further cognitive adjustments (KRAEMER 1993). In other words, affective states of consciousness may have such insistent effects on cognitive flow that direct interventions on the affective processes may, quite simply, be among the most robust and effective ways to rechannel cognitive resources. However, since cognitive attributions can re-evoke emotions once pharmaceuticals have worn off, the role of other therapeutic interventions in establishing long-term ways to solidify new levels of emotional homeostasis need to be implemented. Even strange new technologies such as vagal pacemakers and modulation of cortical activity through rTMS (GEORGE/BELMAKER 2000), not to mention traditional interventions such as exercise, dance and music, may help achieve such ends.

IV. Emotions and Affective States of Consciousness

Our scientific understanding of how emotions and cognitions interact will depend substantially on our ability to decode how consciousness is elaborated by neural tissues. We are finally in an intellectual era where the discussion of such issues is again encouraged, but we remain remote from any consensus on how such processes are instantiated within brains. I favor the view that several types of consciousness exist—with an essential distinction to be made between affective-feeling and cognitive-propositional forms of consciousness as well as the simple perceptual awareness of events in the world (PANKSEPP 1990, 2000a). The former may be integrally linked to global organic processes constructed partly from slowly firing neuropeptide networks of subcortical origin, while the latter may be more "digital" and based on rapid-fire, informationally restricted excitatory amino acid transmission.

Although the foundations of consciousness are, no doubt, constructed from unconscious neural processes, I believe that cognitive forms of consciousness (thoughts about the world) were evolutionarily premised on the prior evolution of affective forms of consciousness, which inform organisms what it might be worth thinking about. That form of mental activity, as described by Marian DAWKINS (1998, p97), may be essentially "a matter of attending to internal images or representations of objects and events… that an animal has some sort of inner representation of the external situation confronting it or that it has memories or anticipations of future situations. Thinking may lead to comparisons between two or more representations and to choices and decision about what to do next based on some sort of assessment of likely outcomes." The same may be said for affective consciousness, except the so-called "representations" may be evolutionarily provided action states that arise intrinsically from emotional systems of the brain. In their role of regulating behavioral output, these states may be much more similar across mammalian species than the the specific thoughts and behaviors animals exhibit.

The emergence of higher forms of consciousness in brain evolution may have been premised upon the preexisting action-readiness and affective dynamics of emotional systems. As I suggested a few decades ago: "I assume that the most primitive function of consciousness is to facilitate adaptive response selection from alternative courses of action: It allows organisms to cope with complex environmental situations in which several behavioral alternatives are competing, with comparable urgency, for a common output channel in the brain. Such a crisis of choice (if one can imagine a crisis on an evolutionary time scale) may have become most urgent to species that possessed executive brain mechanisms that could concurrently promote several adaptive behavior patterns to a single type of environmental challenge. As I have discussed more fully elsewhere… emotive command circuits may have such a characteristic. This flexibility could promote adaptive response-molding, perhaps by a 'reinforcement' mechanisms linked to fluctuating activities in the underlying executive circuits." (PANKSEPP 1982, p451).

The number of proposals on what it means in neural terms to have had emotional feelings are rather scarce. Some believe that feelings are nothing more than some type of information in comparatively recently evolved neocortical working memory systems (e.g., LEDOUX 1996), while others have preferred a JAMES-LANGE type of bodily feedback approach (DAMASIO 1994). The only reasonably well-developed alternative to that view is the possibility that emotional command systems can establish various distinct types of resonances in the neurosymbolic representation of a primordial body (the "SELF"), situated largely, at least in early neonatal development, within deep and ancient mesencephalic areas such as the PAG and surrounding tectal and tegmental systems (PANKSEPP 1998b). Parenthetically, DAMASIO (1999) has more recently moved toward this point of view, with his idea of core-consciousness which is very similar to the concept of the SELF (Simple Ego-type Life Form). The SELF is capitalized to highlight that this is a postulate concerning some type of primordial organization of the brain—a coherent neurosymbolic-homuncular schema of the organism, a virtual body heavily weighted toward the representation of basic motor-orientational and visceral processes. Within consciousness studies, this most central zone of the midbrain has often been neglected in preference to the adjacent Extended Reticular Thalamic Activating System (ERTAS) which is especially important in gating somatic-sensory information to the thalamus (NEWMAN 1997; STERIADE 1996; STERIADE/JONES/MCCORMICK 1997). However, for the "consciousness community," WATT (1998, 1999a, 1999b) has been aspiring to make the necessary adjustments.

This view, contrary to cortico-centric views of consciousness, situates the emergence of global emotional integrative abilities rather more medially in the neuroaxis than the ERTAS, but still situated in a way that can modulate the arousability of the ERTAS structures (i.e., an amalgam of cholinergic, catecholaminergic, GABAergic and glutamatergic systems). With a massive concentration of the affective SELF in mesencephalic levels as well as slightly higher reticular tissues such as hypothalamic and other basal forebrain zones as well as among intralaminar and other reticular tissues of the thalamus, we have an image of affective consciousness which is experimentally testable. Further, the possibility of such brain functions are reiterated in yet higher interconnected tissues, especially frontal cingulate and insular cortical areas, the type of global, organically "embodied" influence that emotions can have on the brain is finally being more widely considered in the literature (see DAMASIO 1999; PANKSEPP 1998b; as well as the recent e-mail seminar organized by WATT, with a target paper by SCHIFF/PLUM 1999 providing a focus for discussion).

Obviously, affective consciousness, just like all other forms of consciousness, do not rely on single nuclear groups but rather the patterned interactions of may brain areas that are all refined by experience. Thus, the seat of the "SELF" is presumably reiterated during brain/psychic maturation, so that affective processes (and hence primal values) continue to inundate the rest of the emerging neural apparatus, especially of frontal cortical zones that elaborate long-term intentions and plans. This would help explain why modern brain imaging procedures tend to largely highlight correlates of telencephalic arousal during emotional states, while the more causal animals studies that tend to manipulate systems directly, are highlighting the importance of subcortical circuits that are rarely visualized with the PET and fMRI imaging procedures. Clearly, those techniques are generating many false negatives, for neurological studies indicate that the subcortical areas are of decisive importance in both the generation of affect and primary-process consciousness (SCHIFF/PLUM 1999). Recent work with new brain-imaging approaches is affirming such conclusions (DAMASIO et al. 2000).

In sum, according to the present view, affective feelings arise from various neurodynamics, which are concentrated but not restricted to specific centromedial areas of the brainstem. This general view could easily have been empirically refined during the middle and later parts of the last century. However, the emergence of digital computers, the cognitive revolution with its informationally encapsulated-modular views of information processing, as well as the rapid decline of psychodynamic perspectives, led several generations of scholars to neglect such integrative concepts for understanding the nature of consciousness. From the long-term vantage, this neglect may have some hidden benefits: Now it allows investigators, steeped in modern neuroscience methodologies, to construct more precise neural image of such foundational issues than could ever have been done before. Such a revolution in our thinking is already fostering new disciplines such as Neuro-Psychoanalysis (see SOLMS/NERSESSIAN 1999).

In any event, the above view of emotional feelings could help explain why so many emotions are so readily projected into the world of sensory-perceptual affairs, and how minimal stimuli in the environment (e.g., the way someone glances at you, or the intonation in one's voice) can captivate the brain/mind in emotional turmoil. Within the deepest mesencephalic areas (e.g., in PAG-ERTAS interactions), we can easily envision how basic emotional and motivational processes control the attentional and information-processing capacities of the somatic-exteroceptive (i.e., sensory thalamic-neocortical) nervous systems. The neurodynamics of emotions can easily inundate the neurodynamics of perceptual systems. Unfortunately, the empirical evaluation of such issues remains rudimentary.

V. Affective Consciousness and the Evolution of Free Will

One of the ultimate issues of consciousness studies is how "free will" could ever emerge from mechanistic activities of the brain. To have true voluntary choice seems incomprehensible within practically all materialistic physiological or computational notions of how cognitive processes of the brain–mind might operate. Unless one is willing to entertain that psychological processes operate at the quantum level (e.g., BECK/ECCLES 1992) there is simply no place for an immaterial mind to intervene in the material processes of the brain. Perhaps the present view of emotions can provide a conceptual handle to how that remarkable a feat may be achieved within a complex materialistic framework which does not aspire to deny that basic biological values do exist within the nervous system. It requires us first to appreciate that consciousness is not only *caused by* but also *realized in* specific types brain systems that mediate "intentions in action" (to again deploy SEARLE's discriminating perspective on this troublesome issue). There is no immaterial mind. A reasonably satisfactory understanding of the issues can be had if we appreciate how the higher forms of cognitive consciousness (e.g., "intention to act"), reflecting some capacity for freedom of choice, are based fundamentally upon the more ancient forms of affective consciousness where such choice was minimally possible. Still, affect programs in the absence of any sophisticated cognitive abilities, can presumably permit simple choices by being able to coherently reflect values that may be only partially represented in the environment. In any event, the basic emotional circuits, without the addition of cognitive potentials, are mechanistically rather closed systems—they are relatively blind to opportunities for adaptive behaviors that may exist in complex environments. Only additional layers of brain evolution opened up opportunities for the type of flexible response selection that we traditionally conceptualize as free will or volitional action (also see, LIBET/FREEMAN/SUTHERLAND 1999).

This view could also help bring some resolution to other major philosophical problems in consciousness studies: The dilemma of how the unified experience of consciousness get "bound" within the brain could be solved if we conceptualize that the very foundation of an affective mind, namely the virtual body or SELF, was first established in evolution upon stable motor coordinates, capable of being modulated by basic emotional systems that generate various forms of action readiness, upon which additional complex perceptual and cognitive processes could be built (PANKSEPP 1998b).

To reiterate, the type of "solution" to the free will problem that I and others (e.g., DAMASIO 1999) favor is based firmly on an evolutionary view of consciousness which makes a distinction between having basic emotional feelings (a brain function that all mammals share) and having the ability to have thoughts about those feelings (which is much more highly resolved in humans than most other mammals). The probability that consciousness emerged rapidly in brain evolution is, of course, remote. It probably went through many stages of emergence (PANKSEPP 1990a), and to understand it, we must first understand the foundational stages.

If affective consciousness emerged fairly early in brain evolution, and it is fundamentally reflective of instinctual emotional action systems interacting with a primitive neural representation of the SELF and ancient bodily (i.e., evolutionary) memories, then, with higher brain evolution, those pieces of information could serve as symbolic token within the deliberative capacities of more recently evolved neural substrates of more cognitive (i.e., exteroceptively tethered) forms of consciousness. Animals that only have affective consciousness presumably do not have the neural complexity to exhibit free (i.e., self-directed) choice. On the other hand, when a receptive neurosymbolic field evolved where affective forms of consciousness could be used as tokens of information in higher levels of deliberation, the doors to "free will" were opened in brain evolution. If the brain substrates for "the SELF" bifurcated, to be well represented in cognitive networks, as they already had within affective networks, then it is possible to envision that decisions could be made several steps removed from the immediacy of one's basic urges. Hence, rather than simply having "intentions in action," to again use SEARLE terminology (1983), the capacity to have "intention to act" gradually emerged.

The capacity of the higher self-structures to entertain several conflicting emotional and motivational alternatives concurrently is, to my way of thinking, the essential foundation for those brain processes that we presently subsume under the concept of "free will." As others, I would seek such higher abilities within neural systems closely affiliated with frontal cortical working memory systems as well as more posterior parietal systems that elaborate multi-modal representations of the world (SPENCE/FRITH 1999).

Although this provides only a glimmer of the complexity that needs to be empirically unraveled, I do not find it problematic to believe that "free will" is fundamentally, the ability of higher brains systems to deliberate more fully on the affective issues confronting an organism than they would be capable of if they did not have the higher symbolic capacities of the cortico-cognitive apparatus. Although all of our choices may seem quite limited, especially when the affective urges are intense, the cognitive symbolization of such affective processes under calmer states of mind, would provide the opportunity for flexible characterological development in well-reared children and hence the widening of meaningful life choices they can eventually make.

Those who are especially committed to becoming masters of that cognitive terrain, can even make choices incompatible with survival. Although most would have little desire to entertain such options, they are potentially there to be entertained by all. Within such a view of complexity, I see no problem for a highly evolved brain like ours, to "freely" pursue options completely of their own making. Of course, we should anticipate that the level of unconscious materialistic control within such control systems will always remain more substantial than any civilized human would wish to admit, and it might be deemed wiser, as a matter of personal philosophy, to not voluntarily enslave oneself to those organic tethers (e.g., as advocated by UUS 1999 in his existential "Libertarian Imperative" option).

In sum, we have to be willing to see our nature from several, often contradictory, perspectives, and it may be worth considering once more what William JAMES (1961, p305) had to say about the diversities of consciousness that can co-exist within our minds: "[O]ur normal waking consciousness, rational consciousness as we call it, is but one special type of consciousness, whilst about, parted from it by the flimsiest of screens, there lie potential forms of consciousness entirely different. We may go through life without suspecting their existence: but apply the requisite stimulus, and at a touch they are there in all their completeness, definite types of mentality which probably somewhere have their field of application and adaptation. No account of the universe in its totality can be final which leaves these other forms of consciousness quite disregarded."

VI. Implications for Psychotherapeutics

The implications of such a vision of the emotional brain should have profound consequences for how we eventually envision certain psychiatric disorders. The existing diagnostic systems, such as DSM IV and ICD-10, are excessively weighted to lists of symptoms, with a rather striking neglect of the underlying neural and psychodynamic issues (JENSEN/HOAGWOOD 1997). The present views may encourage investigators to bring brain emotional aspects into prominence once more (for one potential scheme, see Table 2). If we do this well, we should be able to create more effective and more humane therapeutic approaches, where well-informed patients are full collaborators in the therapeutic enterprise, including the selection of medications based on what they would desire for their lives.

Psychological and somatic therapies would also have increasingly prominent interactive roles in treatment strategies. In addition to harnessing the medium of language and cognitive insights, clients

Basic Emotional System (see Panksepp 1998a)	Emergent Emotions	Emotional Disorders
SEEKING (+ & –)	Interest Frustration Craving	Obsessive Compulsive Paranoid Schizophrenia Addictive Personalities
RAGE (– & +)	Anger Irritability Contempt Hatred	Aggression Psychopathic tendencies Personality Disorders
FEAR (–)	Simple anxiety Worry Psychic trauma	Generalized Anxiety Disorders Phobias PTSD variants
PANIC (–)	Separation distress Sadness Guilt/Shame Shyness Embarrassment	Panic Attacks Pathological Grief Depression Agoraphobia Social Phobias
PLAY (+)	Joy and glee Happy playfulness	Mania ADHD
LUST (+ & –)	Erotic feelings Jealousy	Fetishes Sexual Addictions
CARE (+)	Nurturance Love Attraction	Dependency Disorders Autistic aloofness Attachment Disorders
The SELF—a substrate for Core Consciousness (see Panksepp 1998b).	A mechanism for all Emotional Feelings	Multiple Personality Disorders?

Table 2. Postulated relationships between basic emotional systems, common emotional processes, and major psychiatric disorders. The last two columns only provide best estimates of the major relationships. Obviously, multiple emotional influences contribute to each of the emergent emotions (e.g., jealousy is also tinged by separation distress and anger), and all the emotional disorders have multiple determinants. Plus and minus signs after each indicate major types of affective valence that each system can presumably generate. Capitalization is used to designate the various emotional systems to highlight the fact that these are instantiated as distinct neural entities rather than simply psychological concepts. The essential neural components constitute command influences that coordinate the basic behavioral, physiological and psychological aspects of each emotional response.

would also be increasingly guided to supplement the more standard treatments with musical interventions, dance, exercise, meditation and the various untapped powers of various placebo effects. Psychopharmaceuticals and direct brain stimulation (as with rapid Transcranial Magnetic Stimulation—see Pasqual-Leone et al. 1998; George/Belmaker 2000) might be used more rationally, with continuous structured client feedback about the quality of their lives. In such multi-modal approaches, psychopharmaceuticals might be more commonly used in lower doses, perhaps less frequently (more on demand when clients desire certain kinds of emotional support), and the re-structuring and balancing of emotional energies would be pursued in a much richer therapeutic structures of understanding than they have been for too much of this past century.

Although such re-structuring of mental-health programs may sound utopian, we can anticipate that there will soon be a new age of psychopharmaceuticals, especially as we develop specific modulators of peptide based emotional systems (Panksepp 1993a), where psychological effects of drugs may be highly dependent on the quality of the social–emotional environments in which people live. We have already encountered some of this in our attempts to perfect naltrexone in the treatment of autism: The efficacy of that medication may be dependent on the social sensitivity of care-providers (Panksepp et al. 1991). The notion that medications may be uniquely efficacious in certain emotional environments is an idea whose time will come.

I do hope a day will eventually emerge when the analysis of the human psyche, perhaps again on the proverbial Freudian couch can be implemented in scientifically meaningful ways—perhaps through some type of approach such as "psychoethology.". In part, new variants of psychoanalysis should be guided by our understanding of the basic emotional processes that we share with the other animals. Indeed, for maximal progress, it would be advantageous if institutes are developed where convergent human and animal studies can be conducted under the same physical (and intellectual) roof. At present that remains a rare scientific model. The blending of approaches will require a level of consilience that was never achieved during the past century. It should be one of our foremost goals for the next. Such a synthesis would require us to respect not only the enormous fixed gifts of heredity that we carry within us but the remarkable plasticity of the brain–mind as it interacts with different environments.

VII. The Plasticity of Emotional Systems and Temperament

The plasticity of the nervous system—its ability to be molded in diverse ways by environmental inputs—is increasingly being recognized and acknowledged. Unfortunately, in certain segments of the intellectual community, this is still all too often done with a neglect of the equally important proposition that the plasticity operates within certain genetically dictated limits. The developmental implications of such processes are especially noteworthy (PANKSEPP 2001). Since the available literature in the field is so vast, let me simply summarize three of my favorite recent discoveries from animal brain research that have implications for understanding how stress may affect long-term psychological adjustments.

It is now known that the long-term stress responsiveness of an organism is strongly related to maternal bonding/separation issues (SCHMIDT/SCHULKIN 1999), but the details are rather surprising. For instance, although rodents exhibit a very modest pituitary-adrenal (P-A) stress response during the early neonatal period (3–4 days of life), animals that had been stressed at those early ages exhibit an exaggerated stress response when they grow older. On the other hand, older neonates (11–12 days of age), who already show a vigorous P-A stress response, exhibit comparatively less stress at an older age (VAN OERS/DE KLOET/LEVINE 1998). Thus, the long-term developmental consequences of neonatal stress can be diametrically different depending on exactly when the stress occurred (HEIM et al. 1997). It is to be expected that such long-term changes in stress-reponsivity may have effects on how emotions and cognitions interact later in life (i.e., early trauma that is not remembered may have long-lasting effects on adult personality), but little evidence is presently available on such issues.

A second impressive recent study related to the long term consequences of stress, as induced by a major social event—a single instance of social defeat—found remarkably long lasting effect in rats (RUIS et al. 1999). Socially housed male rats were given one robust experience with defeat by being forced to intrude into the territory of another male. The behavioral and physiological consequences were followed for three weeks. An informative aspect of this study was that following that horrendous defeat, half the animal were returned to live with their normal social groups while half were forced to live individually. The animals that had friendly social companionship following the stressor fared much better. They lost less body weight, were behaviorally less fearful, and exhibited smaller P-A stress responses to new stressful situations. At the end of the three week experiment, the socially-housed animals had larger sex glands and smaller adrenal glands (indicating they had experienced less chronic stress). This effect is rather similar to what we might expect from SHAND's 69th law, and it would suggest that simple social comfort is enough to ameliorate the effects of devastating emotional episodes. Active cognitive support may not be needed. In sum, friendly social companionship protected even "lowly" rats from the deleterious effects of social stress. One can only imagine how long such stressors might last in humans that had little social support—years no doubt.

Finally, in the same vein, BRUIJNZEEL et al. (1999) recently evaluated the cerebral consequences of one prior stress as evaluated by the *number of neurons* that respond to a different stress. The experimental animals in this study were exposed to one experience with mild foot shock, while the controls received no shock. Two weeks later, a shock-probe (very different from the previous shock device) was placed into each animal's cage. In investigating this novel object, animals usually received one or two shocks, and after half an hour their brains were removed and processed for cFos immunoreactivity, highlighting how many cells had been aroused by this stressful experience. The animals that had had the single prior experience with shock, exhibited twice the overall brain reactivity as animals that had not been previously stressed. Thus, the neural consequences of one stress experience could be clearly detected two week later within a remarkably large number of brain areas. Using such techniques, we are coming to better appreciate the widespread consequences of emotional arousal within the brain, even throughout most of the cortex (BECK/FIBIGER 1995; KOLLACK-WALKER/WATSON/AKIL 1997).

In short, we are finally in a position to empirically determine how emotional experiences can modify the temperamental tendencies of animals. More basic studies along these lines are bound to appear, and they will allow us to think clearly about the way similar emotional processes operate within human brains. How emotional habits may emerge in one of our most important, but least recognized emotional systems—the appetitive motivational SEEKING system—is extensively discussed in IKEMOTO/PANKSEPP (1999). Similar scenarios can be constructed for all the other basic emotional systems of the mammalian brain.

Conclusions and evolutionary/computational perspectives

Andy CLARK (1997) has provided a compelling argument for viewing cognitive processes as much more than digital information flow, and has insisted on including the analog processes of the body and environment as essential components of mind. Such perspectives are equally germane for emotions. We need thoroughly organic concepts of emotions in order to come to terms with what the brain really does. The cognitive revolution, modeled upon the type of information that flows most easily in digital computers is insufficient to really get at the roots of those organic dynamics that create affective consciousness. Although "affective computing" is beginning to flourish (PICARD 1997), there are many reasons to suspect that the neural foundation of consciousness is so fundamentally analog and organic that present computational–functional approaches will fail in giving us a realistic image of the foundations of mind. However, computational approaches do provide excellent ways to envision how the "scaffolding" for many higher informational-knowledge processes are created in higher regions of the brain.

At this point in the development (and failure) of cognitive science, it is becoming ever more evident that we need alternatives to traditional digital algorithms of consciousness. As Walter FREEMAN (1995, 1999) has advised us, we may need to fathom the "shape" of mind through images of multiple chaotic attractors derived from real-life analyses of spatially resolved neurodynamics. But still, such computations may only provide surface images of brain functions that constitute psychological processes in action. Then again, maybe that is precisely what mind is—a global, spatio-temporally resolved neurodynamic envelope arising from cascades of local perturbations within multitudes of neuronal assemblies. Perhaps the external form is a fine representation of the inner content. In any event, such dynamic metaphors provide images we desperately need to envision the true complexity of brain–mind. However, to fathom the ancient emotional and motivational systems upon which our higher mental abilities are build, we may need to understand the synaptic tides that course through the fabric of our lower brain through the auspices of many interacting neurochemical systems. Understanding the computational chatterings of digital on–off switches in computer simulations will never provide the powerful knowledge afforded by a study of the underlying organic processes.

In any event, to make substantive progress on such issues we need to have more pluralistic points of view. In addition to progressing further and further upward in computational–representational space, we also need to develop downward views whereby mind is rooted in, and perhaps fundamentally situated within embodied brain processes. Even though we may be able to eventually monitor the fluctuating shapes of primordial aspects of mind in non-linear dynamics (LEWIS/GRANIC 2000), to really understand what is going on, I suspect we shall also have to conceptualize psychological processes in organic terms: Not only does the brain resonate with the abilities of the physical body and the dynamics of the world, as outlined by CLARK (1997), the mind is instantiated in neural nets which do not simply transmit information, but which create dynamic fields of action that proved useful in the evolutionary history of each species. I suspect that such fields of action are the key dimensions of mind which are currently missing in modern mind science, and which can be most readily implanted into our ways of thinking through a variety of primary-process emotional concepts which recognize that affectively tinged action readiness is a fundamental substrate of mind. Basic psychological processes reflect the ways in which the evolutionarily provided brain tools of an organism to reach out into the environment. I especially like William POWERS's (1973, 1998) *Perceptual Control Theory of Action*, which is quite compatible with such views.

The fundamental cleavage lines of the primordial mind will have to be fathomed through some type of evolutionary psychology that we presently do not have. The "massive modularity" of the TOOBY/COSMIDES (2000) tradition has yet to handle the foundational issues well. It has not yet provided an intellectual structure that offers clear predictions about the brain. All too commonly, modern evolutionary psychology is expending its intellectual resources on potential fantasies (see SAMUELS 1998)—evolutionary stories which may interface nicely with the digital-cognitive revolution, but which do not jibe well with what we already know about the ancient regions of the brain where emotional urges are truly elaborated.

Too much of current evolutionary psychology is convincing too many young scholars who have not been steeped in substantive neuroscience and behavior genetics traditions a bill of goods that may lead to another "century of misunderstanding" comparable to that foisted upon us by the behaviorist and cognitive revolutions—branches of which rapidly transformed into "dustbowl" varieties where facts were collected with inadequate guiding concepts (which Kant warned us against—see epigraph). No comprehensive brain–mind science has yet emerged that has given us the type of realistic general image of mind that psychoanalytic thought sought to advance during the 20th century. Of course, the psychoanalytic tradition, following FREUD's (1981a) abandonment of his neuro-theoretical *Project for a Scientific Psychology* also did not have the heart to immerse itself in substantive brain matters, a bias that is only gradually being coaxed to change (see the new journal *Neuro-Psycholanalysis*—SOLMS/NERSESSIAN 1999). In any event, recent efforts like DAMASIO's (1999) are welcome harbingers of a new age of reason.

My personal advocacy of the subcortical view is premised on the conviction that at that level of neural evolution, we will find the genetically ingrained "powers" that have guided all subsequent layers of brain–mind emergence. Accordingly, I remain suspicious of an evolutionary psychology that would aspire to find special-purpose "modules" in higher regions of the human brain (e.g., TOOBY/COSMIDES 2000) when the absolutely essential lower modules we share with the other animals are being ignored. There are many reasons to believe that higher heteromodal regions of the human cortex are more akin, at least at birth, to general-purpose computational devices rather than special-purpose cognitive tools. Perhaps we could here extend Andy CLARK's (1997) remarkable image of language evolution as something that was adapted to existing brain function, rather than the brain being adapted to language. Let us take that line of reasoning a step further back: Perhaps the neocortex, a general purpose information processor resembling a massive ensemble of digital computers is adapted to the exigencies of core subcortical functions—the basic genetically-ingrained survival issues—which took so much longer to construct during brain evolution than did the cortex. We should remember that the human neocortex expanded remarkably rapidly during the past 3 million years, resembling the swift pace at which the speed and memory size of our man-made digital computers has increased during the past half century. A general purpose knowledge machine, with evolutionarily refined perceptual and motor abilities, is much more useful for guiding adaptive behaviors than special purpose cognitive modules. In this context, we should remember that most of the basic emotional and motivational survivial modules had been "perfected" by evolution long before proto-humans strode the face of this earth.

The human neocortex may be better conceptualized not as the fundamental source of consciousness but as a remarkable general-purpose skill-box that is adapted to the types of subcortical functions that had existed for a much longer time. Obviously, it is also designed to perceive the world in specific ways, but I doubt if it could perceive anything if separated from the subcortical functions to which it is largely subservient. In the final accounting, it may only be a tool of a more primary form of consciousness, and quite incapable of sustaining any type of consciousness on its own. And even though it is a most magnificent tool (with vast perceptual and cognitive abilities), it remains, I suspect, a handmaiden for more primary forms of consciousness (DAMASIO 1999; PANKSEPP 1998a, b). Core-consciousness—a global brain dynamic built upon biological survival values—was constructed out of organic materials in brain evolution, and it may still be tethered to those analog processes in some very fundamental ways.

I know of no evidence that is inconsistent with such a view, and if it were more generally considered, we might be encouraged to start looking for the big answers to our big questions subcortically where the cognitive light is dim but the affective light is bright. Obviously the subcortex is incredibly SELF-centric and "myopic" and the remarkable brightening and focussing of perceptual images achieved by the recently emerged cortical abilities is just short of miraculous. However, the real miracle of mind—the seat of the SELF—resides within medially situated subcortical areas as neurological evidence has long affirmed (SCHIFF/PLUM 1999). "Only" the glittering raiments of the mind—the autobiographical self and its many attendants—are stitched together by cortico-cognitive computations (DAMASIO 1999).

In sum, there are presently compelling reason to believe that the foundation of mind is realized in organic processes that can only be superficially modeled (i.e., like a toy-airplane) using computational approaches. Such pursuits may run into a massive wall—the true complexity of life—which may never be simulated except in the most pedestrian of ways. And our most effective simulations may require some type of attempt to mimic the evolutionary and epigenetic processes that help create the real developmental landscapes within living organisms. Molecular biology is now confronting such a walls in its desire to genetically engineer away disease. The interactive complexity of genetic controls remains mind-boggling and will continue to baffle us long after the human genome has been sequenced.

However, our technological endeavors are bound to yield many remarkable products, and computationalists are bound to eventually claim success on the consciousness issue. Accordingly, we now need another Turing Test to provide a compelling screen to evaluate the presence of "real" emotional and motivational processes—mental life in its various affective forms. May I suggest a few? Might not a combination of the following suffice: i) responses on a visually based RORSCHACH-type test designed to simulate our own free-associative tendencies, ii) an auditory-aesthetic prosody/musical appreciation test, and iii) a somatosensory and sexual test to evaluate the pleasure of touch and the friction of skin between consenting beings. We might add gustatory tests to distinguish wonderful culinary delights from more mundane edibles and potables, and tests to evaluate the presence of true hungers, passions, desires as well as cognitive responses to such basic conditions of the flesh.

In any event, it is a pity that many of these issues in humans remain to be empirically characterized. To achieve that, we will need a generation of mind scientists willing to pursue such mysteries of the human mind—evaluating the parametric psychological and brain responses of humans to a large range of affective stimuli, through some type of integrative psychobiology that does not yet exist. Because of the current stranglehold of fine-grained neuroscience and computational cognitivism on available resources, such integrative approaches to the brain–mind have barely begun. In short, it will be as important to understand why colo-rectal distention arouses so much affective turmoil in our brain–minds (TRAUB et al. 1996) as why we aspire to have lofty thoughts.

We will never have a satisfactory understanding of the human mind until we have a reasonable grasp of the "emotional brain" that all mammals share. In other words, there may be something to the embodied nature of living existence at the subcortical level that will require young scholars, devoted to the pursuit of artificial mentality, to get immersed in brain research. Perhaps we should encourage all students interested in mind to return once more, with refreshed evolutionary perspectives, to the animal brain research laboratory as part of their obligatory apprenticeship. Brain emotion theory can guide insightful new observations concerning animal behavior and predictions concerning the feelings that exist in human minds. Without such perspectives, the present tsunami of affect-free, cognitive research, that reveals little about our deeply human/animalian condition, can only increase, and we will continue to have an inadequate appreciation of our deeply embedded place of mind in the living order.

Author's address

Jaak Panksepp, Memorial Foundation for Lost Children , J.P. Scott Center for Neuroscience, Mind & Behavior, Department of Psychology, Bowling Green State University, Bowling Green, OH 43403, USA.
Email: jpankse@bgnet.bgsu.edu

Note

[1] Previously published in 2000 as "The neuro-evolutionary cusp between emotions and cognitions: implications for understanding consciousness and the emergence of a unified mind science", Consiousness and Emotion 1:15–54. Copyright John Benjamins Publishing Company. Reprinted with their permission. Since we did not have the opportunity to work from the final copy-edited manuscript published in Emotions & Consciousness, the present version is based on the penultimate version of the author's manuscript. Except for precise wording and a few paragraphs that were not included in the previously published paper, this is essentially the same paper.

References

Adamec, R. (1997) Transmitter systems involved in neural palsticity underlying increased anxiety and defense—implications for understanding anxiety following traumatic stress. Neuroscience & Biobehavioral Reviews 21:755–765.

Anand, K. J. S. (1997) Long-term effects of pain in neonates and infants. In: Jensen, T. S./Turner, J. A./Wiesenfeld-Hallin, Z. (eds) Proceeding of the 8th World Congress on Pain. Progress in Pain Research and Management, Vol. 8. IASP Press: Seattle WA, pp. 881–892.

Bandler, R./Keay, K. A. (1996) Columnar organization in the midbrain periaqueductal gray and the integration of emotional expression. In Holstege, G./Bandler, R./Saper, C. (eds) The Emotional Motor System, Progress in Brain Research, Vol. 107. Elsevier: Amsterdam, pp. 287–300.

Beck, F./Eccles, J. C. (1992) Quantum aspects of brain activity and the role of consciousness. Proceeding of the National Academy of Sciences 89:11357–11361.

Beck, C. H. M./Fibiger, H. (1995) Conditioned fear-induced changes in behavior and the expression of the immediate early gene c-fos: With and without diazepam pretreatment. Journal of Neuroscience 15:709–720.

Bruijnzeel, A. W./Stam, R./Compaan, J. C./Croiset, G./Akkermans, L. M. A./Olivier, B./Wiegant, V. M. (1999) Long-term sensitization of Fos-responsivity in the rat central nervous system after a single stressful experience. Brain Research 819:15–22.

Buck, R. (1999) The biological affects: A typology. Psychological Review 106:301–336.

Budiansky, S. (1998) If a lion could talk: Animal intelligence and the evolution of consciousness. The Free Press: New York.

Cabanac, M. (1999) Emotion and phyologeny. Journal of Consciousness Studiess 6:176–190.

Childress, A. R./Mozley, P. D./McElgin, W./Fitzgerald, J./Reivich, M./O'Brien, C. P. (1999) Limbic activation during cue-induced cocaine craving. American Journal of Psychiatry 156:11–18.

Christianson, S.-A. (1992) The handbook of emotiona and memory: Research and theory. Lawrence Erlbaum: Hillsdale, NJ.

Chugani, H. T. (1996) Neuroimaging of developmental nonlinearity and developmental pathologies. In: Thatcher, R. W./Reid Lyon, G./Rumsey, J./Krasnegor, N. (eds) Developmental neuroimaging: Mapping the development of brain and behavior. Academic Press: San Diego, pp. 187–195.

Clark, A. (1997) Being there. Putting brain, body, and world together again. MIT Press: Cambridge MA.

Ciompi, L. (1997) The concept of affect logic. An integrative psycho-socio-biological approach to understanding and treatment of schizophrenia. Psychiatry 60:158–170.

Cogan, T. (1802) A philosophical treatise on the passions. S. Hazard: Bath.

Dahl, H. (1998) The voyage of el Rubaiyat to the discovery of FRAMES. In: Bornstein, R. F./Masling, J. M. (eds) Empirical studies of the therapeutic hour. Washington DC: American Psychological Association, pp. 179–227.

Damasio, A. R. (1994) Descartes' error. New York: G. P. Putnam's Sons.

Damasio, A. R. (1999b) The feeling of what happens: Body and emotion in the making of consciousness. Harcourt Brace: New York.

Damasio, A. R./Grabowski, T. J./Bechara, A./Damasio, H./Ponto, L. L. B./Parvizi, J./Hichwa, R. D. (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. Nature Neuroscience 3:1049–1056.

Darwin, C. (1998) The expression of the emotions in man and animals. Third Edition. Oxford University Press: New York. Originally published in 1872.

Davis, M./Shi, C. (1999) The extended amygdala: Are the central nucleus of the amygdala and the bed nucleus of the stria terminalis differentially involved in fear versus anxiety? Annals of the New York Academy of Sciences. 877:281–291.

Dawkins, M. S. (1998) Through our eyes only? The Search for animal consciousness. Oxford University Press: New York.

Ekman, P. (1998) Universality of emotional expression? A personal history of the dispute. In: Ekman, P. (ed) Third Edition of Charles Darwin's the Expression of the Emotions in Man and Animals, with introduction, afterwords, and commentaries. Oxford University Press: New York, pp. 363–393.

Falls, W. A./Miserendino, M. J. D./Davis, M. (1992) Extinction of fear-potentiated startle: Blockade by infusion of an excitatory amino acid antagonist into the amygdala. Journal of Neuroscience 12:854–863.

Freeman, W. J. (1995) Societies of Brain: A study in the neuroscience of love and hate. Lawrence Erlbaum: Hillsdale NJ.

Freeman, W. J. (1999) How the brain makes up its mind. Blackwell: London.

Freud, S. (1981a) Project for a scientific psychology. In: Strachey, J. (ed) The Standard Edition of the Complete Psychological works of Sigmund Freud. The Hogarth Press: London. Originally published in 1895.

Freud, S. (1981b) Introductory lectures on psychoanalysis. In: Strachey, J. (ed) The Standard Edition of the Complete Psychological works of Sigmund Freud. The Hogarth Press: London. Originally published in 1916–1917.

Freud, S. (1981c) An outline of psychoanalysis. In: Strachey, J. (ed) The Standard Edition of the Complete Psychological works of Sigmund Freud. The Hogarth Press: London. Originally published in 1940.

Fridja, N. H. (1986) The emotions. Cambridge Univ. Press: Cambridge UK.

Gazzaniga, M. S. (ed) (2000) The new cognitive neurosciences, second edition. MIT Press: Cambridge MA.

Gazzaniga, M./Ivry, R. B./Mangun, G. R.(eds.) (1998) Cognitive neuroscience: The biology of the mind. W. W. Norton: New York.

George, M. S./Kettner, T. A./Kimbrell, T. A./Steedman, J. M./Post, R. M. (1996) What functional imaging has revealed about the brain basis of mood and emotion. In: Panksepp, J. (ed) Advances in Biological Psychiatry, Vol. 2. JAI Press: Greenwich CT, pp. 63–114.

George, M. S./Belmaker, R. H. (eds.) (2000) Transcranial magnetic stimulation in neuropsychiatry. American Psychiatric Publishing Group: Washington DC.

Gray, J. A. (1990) Brain systems that mediate both emotion and cognition. Cognition and Emotion 4:269–288.

Griffiths, P. E. (1997) What emotions really are: The problem of psychological categories. University of Chicago Press: Chicago.

Hauser, M. D. (2000) Wild minds: What animals really think. Henry Holt & Co: New York.

Heim, C./Owens, M. J./Plotsky, P. M./Nemeroff, C. B. (1997) Persistent changes in corticotropin-releasing factor systems due to early life stress: relationship to the pathophysiology of major depression and psot-traumatic stress disorder. Psychopharamcology Bulletin 33:185–192.

Ikemoto, S./Panksepp, J. (1999) The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. Brain Research Reviews: 31:6–41.

Jensen, P. S./Hoagwood, K. (1997) The book of names: DSM-IV in context. Developmental Psychopathology 9:231–249.

James (1961) The varieties of religious experience. MacMillan: New York. Originally published in 1902.

Kandel, E. R. (1998) A new intellectual framework for psychiatry. American Journal of Psychiatry 155:457–469.

Knutson, B./Wolkowitz, O. M./Cole, S. W./Chan, T./Moore, E. A./Johnson, R. C./Terpstra, J./Turner, R. A./Reus, V. I. (1998) Selective alteration of personality and social behavior by serotonergic intervention. American Journal of Psychiatry 155:373–379.

Kollack-Walker, S./Watson, S. J./Akil, H. (1997) Social stress in hamsters: Defeat activates specific neurocircuits within the brain. Journal of Neuroscience 15:8842–8855.

Kraemer, P. D. (1993) Listening to Prozac. Viking: New York.

Langer, S. K. (1951) Philosophy in a new key: A study in the symbolism of reason, rite and art. Mentor Books: New York.

LeDoux, J. (1996) The Emotional Brain. They Mysterious Underpinnings of Emotional Life. Simon and Schuster: New York.

Lewis, M. D./Granic, I. (eds.) (2000) Emotion, Self-Organization, and Development. Cambridge University Press: New York.

Libet, B./Freeman, A./Sutherland, K.(eds.) (1999) The volitional brain: Towards a neuroscience of free will. Journal of Consciousness Studiess 6:1–298.

MacLean, P. D. (1990) The triune brain in evolution: Role in paleocerebral functions. Plenum Press: New York.

Maren S. (1999) Neurotoxic basolateral amygdala lesions impair learning and memory but not the performance of conditional fear in rats. Journal of Neuroscience 19:8696–8703.

McBride, W. J./Murphy, J. M./Ikemoto, S. (1999) Localization of brain reinforcement mechanisms: intracranial self-administration and intracranial place-conditioning studies.Behavioural Brain Research 101:129–152.

Newman J. (1997) Putting the puzzle together: Towards a general theory of the neural correlates of consciousness. Journal of Consciousness Studies 4(1):47–66 and 4(2):101–121.

Öhman, A. (1993) Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. In: Lewis, M./Haviland, J. M. (eds) Handbook of Emotions. Guilford: New York, pp. 511–536.

Olmstead, M. C./Franklin, K. B. (1997) The development of a conditioned place preference to morphine: effects of microinjections into various CNS sites. Behavioral Neuroscience 111:1324–1334.

Panksepp, J. (1982) Toward a general psychobiological theory of emotions. The Behavioral and Brain Sciences 5:407–467.

Panksepp, J. (1985) Mood changes. In: Vinken, P. J./Bruyn,G. W./Klawans, H. L. (eds.) Handbook of Clinical Neurology. Vol. 1. (45): Clinical neuropsychology. Elsevier Science Publishers: Amsterdam, pp. 271–285.

Panksepp, J. (1988) Brain emotional circuits and psychopathologies. In: Clynes, M./Panksepp, J. (eds) Emotions andPsychopathology. Plenum Press: New York, pp. 37–76.

Panksepp, J. (1990a) The psychoneurology of fear: Evolutionary perspectives and the role of animal models in understanding anxiety. In: Burrows, G. D./Roth, M./Noyes Jr, R. (eds) Handbook of anxiety, Vol. 3: The Neurobiology of Anxiety. Elsevier: Amsterdam, pp. 3–58.

Panksepp, J. (1990b) Gray zones at the emotion-cognition interface, a commentary. Cognition and Emotion 4:289–302.

Panksepp, J. (1991) Affective Neuroscience: A conceptual framework for the neurobiological study of emotions. In: Strongman, K. (ed) International Reviews of Emotion Research. Wiley: Chichester UK, pp 59–99.

Panksepp, J. (1993a) Neurochemical control of moods and emotions: Amino acids to neuropeptides. In: Lewis, M./Haviland, J. (eds) The Handbook of Emotions. Guilford: New York, pp. 87–107.

Panksepp, J. (1993b) Rough-and-tumble play: A fundamental brain process. In: MacDonald, K. B. (ed) Parents and Children Playing. SUNY Press: Albany NY, pp. 147–184.

Panksepp, J. (1998a) Affective neuroscience: The foundations of human and animal emotions. Oxford University Press: New York.

Panksepp, J. (1998b) The periconscious substrates of consciousness: Affective states and the evolutionary origins of the SELF. Journal of Consciousness Studies 5:566–582.

Panksepp, J. (1999) Emotions as viewed by psychoanalysis and neuroscience: An exercise in consilience, and accompanying commetaries. NeuroPsychoanalysis 1:15–89.

Panksepp, J. (2000) Affective consciousness and the instinctual motor system: The neural sources of sadness and joy. In: Ellis, R./Newton, N. (eds) The Caldron of Consciousness: Motivation, Affect and Self-organization. Advances in Consciousness Research. John Benjamins Pub. Co: Amsterdam, pp. 27–54.

Panksepp, J. (2001) The long-term psychobiological consequences of infant emotions: Prescriptions for the 21st century. Infant Mental Health Journal 22:132–173.

Panksepp, J./Burgdorf, J. (1999) Laughing rats? Playful tickling arouses high frequency ultrasonic chirping in young rodents. In: Hameroff, S./Chalmers, D./Kazniak, A. (eds) Toward a science of consciousness III. MIT Press: Cambridge MA, pp. 231–244.

Panksepp, J./Lensing, P./Leboyer, M./Bouvard, M. P. (1991) Naltrexone and other potential new pharmacological treatments of autism. Brain Dysfunction 4:281–300.

Panksepp, J./Normansell, L. A./Herman, B. Bishop, P./Crepeau, L. (1988) Neural and neurochemical control of the separation distress call. In: Newman, J. D. (ed) The physiological control of mammalian vocalization. Plenum Press: New York, pp. 263–299.

Parrot, W. G./Schulkin, J. (1993) Neuropsychology and the cognitive nature of emotions. Cognition and Emotion 7:43–59.

Pasqual-Leone, A./Tarazona, F./Keenan, J./Tormos, J. M./Hamilton, R./Catala, M. D. (1998) Transcranial magnetic stimulation and neuroplasticity. Neuopsychologia 37:207–217.

Picard, R. W. (1997) Affective computing. MIT Press: Cambridge MA.

Powers, W. T. (1973) Behavior: The control of perception. Aldine de Gruyter: New York.

Powers, W. T. (1998) Making sense of behavior: The meaning of control. Benchmark Pub. Inc: New Canaan CT.

Rolls, E. T. (1999) The brain and emotion. Oxford University Press: Oxford UK.

Ruis, M. A. W./te Brake, J. H. A./Buwalda, B./De Boer, S. F./Meerlo, P./Korte, S. M./Blokhuis, H. J./Koolhaas, J. M. (1999) Housing familiar male wildtype rats together reduces the long-term adverse behavioral and physiological effects of social defeat. Psychoneuroendocrinology 24:285–300.

Samuels, R. (1998) Evolutionary psychology and the massive modularity hypothesis. The British Jounral for the Philosophy of Science 49:575–602.

Schechter, M. D./Calcagnetti, D. J. (1993) Trends in place preference conditioning with a cross-indexed bibliography; 1957–1991. Neuroscience & Biobehavioral Reviews 17:21–41.

Schiff, N. D./Plum, F. (1999) The neurology of impaired consciousness: Global disorders and implied models, ASSC E-Seminar Target article. Retrieved from the internet at http://www.phil.vt.edu/assc/niko.html.

Schmidt, L. A./Schulkin, J. (eds) (1999) Extreme fear, shyness and social phobia: Origins, biological mechanisms and clinical outcomes. Oxford University Press: New York.

Schultz, W. (1998) Predictive reward signal of dopamine neurons. Journal of Neurophysiology 80:1–27.

Searle, J. R. (1983) Intentionality: An essay in the philosophy of mind. Cambridge University Press: Cambridge UK.

Shand, A. F. (1920) The foundations of character: Being a study of the tendencies of the emotions and sentiments. MacMillan and Co: London.

Solms, M./Nersessian, E. (1999) Freud's theory of affect: Questions for neuroscience. Neuro-Psychoanalysis 1:5–14.

Spence, S. A./Frith, C. D. (1999) Towards a functional anatomy of volition. Journal of Consciousness Studiess 6:11–30.

Steriade, M. (1996) Arousal: revisiting the reticular activating system. Science 272:225–26..

Steriade, M./Jones, E./McCormick, D. (eds.) (1997) Thalamus. Volume 1: Organization and Function. Volume 2: Experimental and clinical aspects. Elsevier: Amsterdam.

Tooby, J./Cosmides, L. (2000) Toward mapping the evolved functional organization of mind and brain. In: Gazzaniga, M. S. (ed) The New Cognitive Neuroscience, second edition. MIT Press: Cambridge MA, pp. 1167–1178.

Traub, R. J./Silva, E./Gebhart, G. F./Solodkin, A. (1996) Noxious colorectal distention induced-c-Fos protein in limbic brain structures in the rat. Neuroscience Letters 215:165–168.

Uus, U. (1999) The libertarian imperative. Journal of Consciousness Studies 6: 48–64.

Van der Kolk, B. A./McFarlane, C. M./Weisaeth, L. (eds.) (1996) Traumatic stress: The effects of overwhelming experience on mind, body, and society. Guilford: New York.

Van Oers, H. J. J./de Kloet, E. R./Levine, S. (1998) Early vs. late maternal deprivation differentially alters the endocrine and hypothalamic responses to stress. Development Brain Research 111:245–252.

Watt, D. (1998) Emotion and Consciousness: Implications of affective neuroscience for extended reticular thalamic activating system theories of consciousness. ASSC E-Seminar Target article. Sept. 21, 1998–Oct. 9, 1998. Retrieved from the internet at http://server.phil.vt.edu/assc/watt/default.html.

Watt, D. F. (1999a) Consciousness and emotion: Review of Jaak Panksepp's 'Affective Neuroscience'. Journal of Consciousness Studies 6:191–200.

Watt, D. F. (1999b) At the intersection of emotion and consciousness: Affective neuroscience and extended reticular thalamic activating system (ERTAS) theories of consciousness. In: Hameroff, S./Chalmers, D./Kazniak, A. (eds) Toward a science of consciousness III. MIT Press: Cambridge MA, pp. 215–229.

Winslow, J. T./Hastings, N./Carter, C. S./Harbaugh, C. R./Insel, T. R. (1993) A role for central vasopressin in pair bonding in monagamous prairie voles. Nature 365:544–548.

Wimmer, M. (1995) Evolutionary roots of emotions. Evolution and Cognition 1:38–50.

Wimmer, M./Ciompi, L. (1996) Evolutionary aspects of affective-cognitive interactions in the light of Ciompi's concept of "Affect-Logic". Evolution and Cognition 2:37–58.

Wise, R. A. (1996) Neurobiology of addicition. Current Opinions in Neurobiology 6:243–251.

Matthias Scheutz

Ethology and Functionalism

Behavioral Descriptions as the Link between Physical and Functional Descriptions

Abstract

In this paper, I introduce the notion of 'behavioral state' as a means to bridge the gap between functional specifications of cognitive architectures and their physical implementations based on ethological methods of describing animal behavior. After briefly sketching some of the problems resulting from mere functional descriptions of cognitive architectures, I define the notion of behavioral state and discuss some of the properties of behavioral states that are relevant for describing and modeling minds. I show that behavioral states can serve as mediators between functional and physical descriptions of cognitive systems, arguing in particular that behavioral states can capture 'distance in time', an essential aspect of real world cognition missing in mere functional descriptions.

Key words

Functionalism, ethology, behavioral states, causation, implementation level of description.

1. Introduction

Ever since DESCARTES, philosophers interested in the mind have divided the world into a mental and a physical realm and consequently contemplated the relation between these two realms, a topic today widely known as the 'mind–body' problem. While this problem is far from being resolved, today's most commonly held position on the 'mind–body' relation in the philosophy of mind is *functionalism*, a view, which despite its appearance in many different forms is based on the central common claim that mental states *are* functional states. The general understanding is that mental states (i.e., states such as 'believing that *p*' or 'desiring *x*', or even psychological predicates such as 'pain' or 'pleasure') can be explained in terms of functional states and functional architectures.

Besides the fact that to my knowledge no one has ever attempted to specify concepts from folk psychology *in detail* using a functional architecture, the cognitive scientist who wants to understand and model cognitive systems will still face significant problems even if a complete functional specification of a given cognitive system could be provided: for one, the question of how functional states are related to physical states remains unanswered. Usually, philosophers assume that functional states 'supervene' on physical states without paying particular attention to the question as to *how* (and consequently also *why*) they supervene.[1] In other words, what plays a secondary role (if at all), is of crucial importance to the cognitive scientist: (some) implementation details of the functional architecture of these very abstract mental states. For example, it is not clear whether functional states can be realized as computational states (maybe only combined 'computational–physical' states will realize functional states or maybe only physical states alone). And more generally, the question arises what the constraints are that a functional architecture imposes on systems implementing it: are functional descriptions besides being general enough to include all possible mental architectures specific enough to constrain the class of possible realizing systems in such a way as to suggest possible ways of implementing them?

It seems that relating functional states *directly* to physical states is very unlikely to succeed in the light of multiple realization arguments for functional architectures (the more complex the architecture gets, the less we will be able to see what kinds of possibly very diverse physical systems will share the functional specification).[2] The level of functional specification of the psychology of minds will be too high and abstract a level of description to suggest *possible implementations* of the functional states (not to mention all the problems connected with the involved notion of 'implementation' or 'realization' that seem to be largely ignored by the philosophical community).[3]

It is my conviction that functional specifications of psychologies are not sufficient to suggest ways of understanding and modeling minds. To be of any *practical importance* in modeling a mind at all, a level of description of a cognitive architecture has to incorporate at least *some* of the relevant physical properties of its possible implementations (which will constrain both possible implementations as well as functional architectures). In this paper, I will suggest such an intermediary level, which I call *the level of behavioral states*. This level of description is largely inspired by ethological studies of animal behavior (and to some extent by research in behavior-based robotics) and will therefore bear the insignia of its intellectual sources very visibly on its sleeves.

First, I will briefly point to one of the problems resulting from a mere functional description of a cognitive system (the 'implementation problem'). Then I will introduce the notion of 'behavioral state' and locate its place as mediator between functional and physical states, sketching briefly the role behavioral states could play in understanding, designing, and implementing (simple) cognitive architectures. Finally, I argue that behavioral states are sufficient to capture relevant aspects of cognition and, thus, provide an intermediary level of architectural specification located between functional and physical descriptions.

2. Functionalism

2.1 The functionalist picture

A functional specification of a cognitive architecture consists of a set of input states, a set of output states, and a set of 'inner' or 'functional' states together with a specification of how they are *causally* related. That way it is possible to determine what state a cognitive system will be in next, given the current state and all the input conditions.[4] While input and output conditions have to be tied to physical inputs and outputs, the functional states do not require a direct correspondence to their physical realizers as expressed in the phrase that "functional states supervene on physical states" (e.g., see KIM 1997). This lack of a 'direct' correspondence between functional and physical states is what gives functionalism its explanatory power, while keeping it metaphysically palatable: it combines advantages of behaviorist approaches to mind (i.e., considering solely the input–output behavior of an organism) with advantages of identity theories (i.e., mental state/event tokens are physical state/event tokens) leaving out the pitfalls of both such as the lack of being able to account for 'inner states' in the former, and the requirement of type identities between mental and physical state/event types of the latter. Yet, this strength comes at a price: it is not clear what it means to *implement* or *realize* a functional architecture.

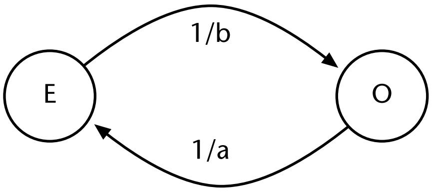

Figure 1. The even-odd transducer with two inner states.

2.2 Implementation of a functional architecture

So what are the implementation conditions for a functional architecture? To say that a system implements a functionalist description is to require that in addition to the input and output mapping, it has to get the mapping of the inner states right. Usually, these 'inner states' are assumed to be multiply realizable, i.e., many different, possibly very diverse physical systems will realize a given functional architecture. Therefore, the mapping between physical states and functional states has to be a many-to-one (very much in the spirit of CHALMERS 1997). Yet, inner states are viewed by functionalists as intrinsically relational states, being *mutually defined* by all states in the functional architecture (which is sometimes expressed by saying that they are defined by their 'causal role' in the functional architecture).

To illustrate this interdependence, consider, for example, the following automaton, which has two inner states 'E' and 'O' standing for 'even' and 'odd'. Depending on whether the number of '1's that the automaton has seen so far is even or odd, it outputs either 'a' or 'b', respectively.

A functionalist account (e.g., see BLOCK 1996) of what it means to be in state E would look like this:

Being in E $=_{\text{def}}$ Being an x such that $\exists P \, \exists Q$ [x is in $P \wedge$ (if x is in P and receives input '1', then it goes into Q and outputs 'b') $\wedge$ (if x is in Q and gets input '1', then it goes into P and outputs 'a')].[5]

Since it is only claimed that there has to be an arrangement of physical states that corresponds to the functional states in a way that preserves inputs and outputs as well as transitions between states, it is possible for one physical state to serve as the instantiation of more than one functional state (and vice versa). Therefore, the correspondence between physical and functional states is not necessarily that of a mapping between physical types and functional types (let alone a 1–1 mapping), but rather that of a relation that preserves state transitions. 'Implementation of a functional architecture', therefore, has to be viewed as some sort of 'bisimilarity' between functional and physical architecture rather than some sort of *isomorphic* relation from a functionalist point of view.[6] As a consequence, not every functional state might have a unique correspondence in the physical system, i.e., functional difference might not amount to physical difference, as it is possible that two different functional states are realized by the very same physical state (e.g., think of virtual memory systems in computers), a possibility that can complicate the search for a physical correlate of functional states (in section 4 I will address another essential difficulty of merely 'causal' descriptions, namely their failure to capture 'distance in time').

3. Behavioral States

3.1 An ethological perspective

To overcome the difficulties of tying functional specifications to physical implementations, I suggest to consider work done in animal behavior research as a venture point. According to animal behaviorists (e.g., McFarland 1981), animal behavior can be categorized in terms of

1. Reflexes (i.e., rapid, involuntary responses to environmental stimuli)
2. Taxes (i.e., responses orienting the animal towards or away from a stimulus)
3. Fixed-action patterns (i.e., time-extended sequences of simple responses)

While (1) and (2) are solely connected to external stimulation, (3) can have a contributing 'internal' component as well (fixed action patterns can be 'motivated'; take, for example, the 'egg-retrieving' behavior of the greyleg goose, see Lorenz 1981, or Lorenz/Leyhausen 1973). All three kinds of behaviors can be combined in complex ways to form hierarchies of behaviors (see figure 2).

In these behavioral structures, behaviors form 'competitive clusters', in which behaviors are mutually exclusive (e.g., in figure 2 the 'fighting behavior' is such a competitive cluster comprising the mutually exclusive behaviors 'chasing', 'biting', and 'display').To make these ideas of behavioral hierarchies more concrete, I will introduce the notion of *behavioral state*, which roughly corresponds to what is indicated by a 'circle' in figure 2. Putting it crudely, a behavioral state is a state an individual is in if it performs a particular behavior (e.g., such as 'food handling' or 'looking out for prey').[7] 'Behavior' is meant be understood in a *wide sense* to include behaviors that are not necessarily observable from the outside alone (such as 'memory recall' or 'thinking', in general). Hence behavioral states are not simply combined input–output states, but rather they are some sort of 'inner states' of an organism, states in which the organism is if it performs a particular kind of behavior. Note, however, that nothing is implied or claimed about a particular physical correlate of a behavioral state—it might or might not exist (I will return to this issue later).

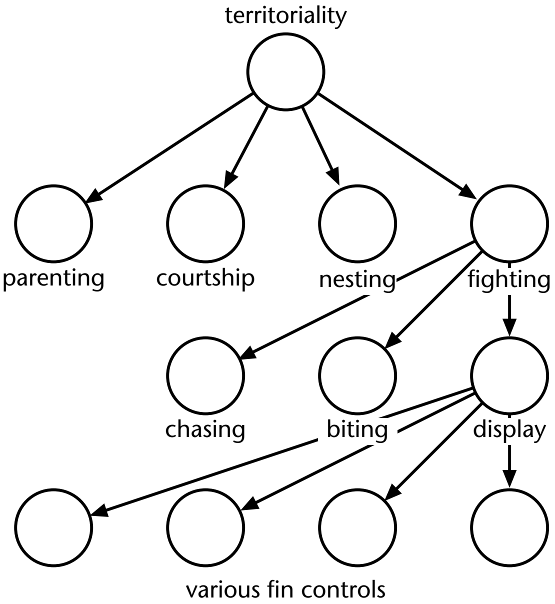

Figure 2. A part of a behavioral hierarchy for the male stickleback fish (see Lorenz 1981). The various fin controls can be divided further into rays of each fin, the muscle fibers for each ray, and the motor neurons for each fiber.

Behavioral states are not restricted to 'motor actions', but include sensory actions as well as more abstract proprioceptive and reflective actions (such as monitoring inner physiological states, generating images, producing plans, recalling poems, analyzing pictures, making logical derivations, etc.). The latter ones are more 'abstract behaviors', which are mostly (if not completely) internalized and often involve solely parts of the cognitive architecture; in fact, they might not result in any externally observable change at all (a mathematician contemplating abstract objects and manipulating their representations in her mind, for example, might not need any stimulation from the outside world in performing this task, nor might any motor action result from it—this 'brain in a vat'-idea with sustained cognitive activity whilst lacking external interaction seems to be at least conceivable in principle).

Memory and reflective processes, for example, are then viewed as special kinds of behavioral processes that lead to actions performed directly on the cognitive architecture, as opposed to the effectors of the individual which act on the environment.

In general, an individual will be in many behavioral states at the same time reflecting the fact that (1) some behaviors are contained in or shared among others (for example, searching for food as well as searching for a mate will both involve locomotion, despite the fact that the kind of search might be different), and (2) that many behaviors are performed in parallel (such as monitoring my hand as I move it to pick up an object).

3.2 Behavioral architectures

In a sense, the classical ethological picture outlined above is mainly concerned with the relation between various behaviors, it only depicts (some of the) causal relations between behaviors, and is, therefore, really a functional specification of the behavioral architecture. Yet, partly *implicit* in and partly *external* to this picture is information about the time constraints as well as the strength of interactions and influences among behaviors (as studied and gathered by animal behaviorists). In other words, the picture is *incomplete* in so far as it leaves out essential implementation details that cannot be retrieved from a picture like figure 2 alone. Without these implementation details, however, some behaviors would not be the kinds of behaviors they are, since what distinguishes them from other behaviors might just be constraints on timing and strength of response (take, for example, a retraction reflex caused by touching a hot plate with your finger as opposed to the same movement being performed very slowly). Furthermore, the strength and configuration of interactions between behaviors is an integral part of their defining characteristics, which cannot be captured by a causal structure alone: suppose behavior A *causes* behavior B. Then this can happen in many different behavioral arrangements, for example, by A enforcing B directly or A suppressing C, which in turn inhibits B, or by A enforcing D, which enforces C, etc. Implicit in A (as defined by an animal behaviorist, say) is already information, which of these possible arrangements are realized in the animal. Hence, the causal structure might get restricted by the behavioral structure if (some of) the information implicit in the definition of behaviors is made explicit. In the following, I will briefly sketch how behavioral states can be defined to explicitly incorporate some of the otherwise implicit aspects of behaviors.

3.3 The structure of behavioral states and networks

First and foremost, each behavioral state has an *activation level* and a *behavior* associated with it. This activation may depend on any of the following factors (and additional factors could be considered):

1. Its own activation level
2. The activation level of other states
3. Inputs from extereoceptive and proprioceptive sensors
4. Energy constraints (of the organism)
5. Decay over time

The behavior associated with a behavioral states can be simple (such as reflexes and taxes), or a more complex fixed behavior (such as fixed action patterns), or an even more complex adaptive behavior (which results from the interplay of fixed action patterns, reflexes, and taxes). The term 'adaptive' indicates that the latter kinds of behaviors can change over time, i.e., they can be learned, altered, etc. (utilizing the dynamic interplay of behavioral states).

Behavioral states are connected via inhibitory and excitatory links to other behavioral states and possibly to sensors (via 'information channels', i.e., filtering mechanisms that select parts of one or more sensory inputs and combine them in particular task-specific ways). Connections between behavioral states have a distance associated with them (expressed in terms of a time-lag), reflecting the 'distance in space' that a signal has to travel from one locus of action to interact with another, allowing temporal as well as spatial integration of incoming signals.

Groups of behavioral states that are connected via mutually inhibitory links form so-called 'competitive clusters'. They inhibit each other to various degrees, while usually entertaining excitatory connections to lower and upper level states (and possibly to some behavioral states of other clusters at the same level as well). In such a cluster the behavior associated with the highest activated state will become activate and all behaviors of the other states are suppressed.[8] This way hierarchical structures similar to the one in figure 2 can be defined which reflect the relationship between behaviors and in part also the complexity of each behavior associated with the various states (the lowest levels corresponding to simple reflex-like, reactive behaviors—this level has been explored in great detail in behavior-based robotics, e.g., see ARKIN 1992, or BROOKS 1986).

With respect to the spread of activation, networks of behavioral states are very similar to I(interactive) A(ctivation) and C(ompetition) networks (e.g., see RUMELHART/MCCLELLAND 1986). Therefore, results from connectionist research about effects such as 'blocking', 'settling', 'oscillation', 'hysteresis', and others (often) apply *mutatis mutandis* to behavioral networks as well. The essential difference between IAC networks and behavioral networks is that the behavior associated with a behavioral state could affect the activation level of the very state itself as well as the activations of other states *via environmental feedback*. For example, a behavioral node representing the 'search for black objects in visual field'-behavior might initiate ocular motor commands that lead to the detection of a small black object by another node, which in turn inhibits the search node, thus decreasing its activation, which in a mere IAC network (lacking environmental feedback) would have otherwise not decreased.[9]

As already mentioned, not all behaviors will involve physical effectors; in fact, only low level behaviors will directly exert influence on them (these are behaviors that would normally be localized in what roboticists refer to as 'reactive layer'). Higher level behavioral states will mostly operate on structures internal to the cognitive system (these states would be situated in the 'deliberative layer'). For example, a 'retrieve image of mother' node (assuming for a moment there is such a node), might initiate a search in long-term memory (possibly involving other behavioral states) for a particular image that is associated with the individual's mother. Or a 'project-hand-move-forward' node might initiate a 'simulated' hand movement in an emulator circuit, which is used to plan motions, resulting in a change in the circuit and as a consequence in other behavioral nodes (such as 'collision detectors' in the emulator circuit, etc.).[10] A behavioral network divided into a layered structure consisting of a reactive and a deliberative layer is schematically depicted in figure 3.

There are special cases of behavioral states that do not have any behavior directly associated with them. Instead of initiating an action directly, they contribute to behaviors indirectly by influencing other behavioral states, and can, therefore, assume the role of affective states. A state corresponding to 'hunger', for example, might receive inputs from proprioceptive sensors (i.e., a sensor monitoring the blood sugar or, more generally, the energy level) and exert positive influence on other states such as 'search-for-food' (e.g., see SCHEUTZ 2000). That way it is possible to entertain states that do not directly and immediately 'cause' the individual to act in a particular way, but might have indirect, long-term effects on the individual (e.g., depression, memory loss, etc.).[11]

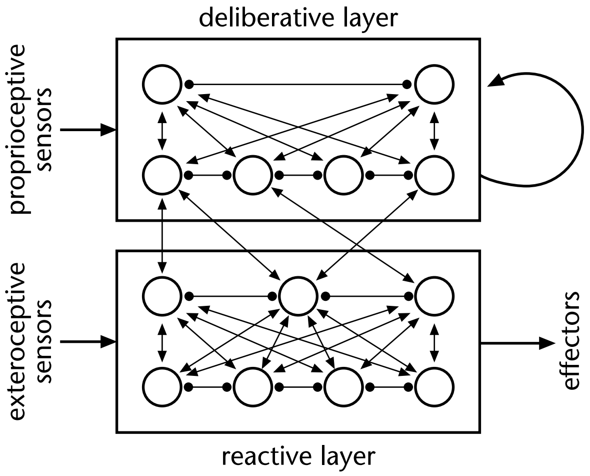

Figure 3. A hierarchy of behavioral states viewed as a two-layered architecture consisting of a deliberative and a reactive layer. Links with arrows indicate excitatory connections; links with circles inhibitory ones. While the behavioral units in the reactive layer operate on effectors (performing behaviors such as navigating through the environment, avoiding obstacles, etc.), behavioral nodes in the deliberative layer do not operate on effectors directly, but rather perform internal operations (such as memory lookups, symbolic combinations, etc.).

4. The Case for an Intermediate Level

4.1 The relations between physical, functional, and behavioral states

So far, I have not explicated how physical and functional states relate to behavioral states as defined above. From an implementation perspective, behavioral states can be realized in many ways in different physical substrates. In brains, for example, they could correspond to a single neuron or to a group of neurons. They could be realized solely neuronally or maybe by involving other systems (such as the hormonal system) as well. Another physical medium, in which behavioral states can be realized, is the silicone of computers: computers can implement behavioral states by virtue of computational processes.

Some behavioral states might be (directly) 'implemented' in the system in the sense that there exists a corresponding physical state or a set/sequence of physical states that are in *type correspondence* with the behavioral state. Other behavioral states might 'supervene' on physical states in that there does not exist such a type correspondence—note that programs running on modern operating systems with virtual memory architectures exhibit such supervenience relations: when a program does not entirely fit into physical memory, it is loaded in parts on an 'as-needed' basis, where different virtual memory locations get mapped onto the same physical memory location.

Another possibility for behavioral states to have no *fixed correlate* at all is to be only *partially* implemented (see SLOMAN 1998) or to depend on environmental conditions (e.g., in terms of other behavioral states and/or environmental states—an example might be my performing the multiplication algorithm using paper and pencil: I am in a behavioral state which is implemented by a number of other states such as states of the paper and pencil, several visual routines, rule-retrieving memory processes and rule-following routines, etc.).

Behavioral states implemented in (sequences of) physical states are tightly coupled to their physical realizers (still allowing for multiple realizations), while behavioral states supervening on physical states do not exhibit such a coupling at all. They are realized by some physical states, but they might not show any systematic correlation to their realizers. For example, consider two networks of behavioral states, which are *functionally* identical except for the fact that the first explicitly implements a higher level behavioral state called 'avoid-obstacle', which is active if the agent is engaged in obstacle avoidance behavior. The second one does not have such as state, but can still control the same obstacle-avoidance behavior. In this case, the behavioral state 'obstacle-avoidance' has a physical correlate in the former and no fixed physical correlate in the latter (what corresponds physically to the 'obstacle-avoidance' state in the latter is a complex sequence of patterns that might, under different circumstances, not correspond to this state at all, e.g., if the agent follows another agent, which is avoiding obstacles, and thus is a 'follow other agent' state, which by pure chance causes it to go through the same sequence of physical states… see also PFEIFER/SCHEIER 1999, ch. 12 for another example).[12]

This aspect of behavioral states seems very similar to the kinds of functional states about which philosophers tend to worry, and maybe most of the 'high-level' functional states such as 'belief states', etc. are not directly (i.e., physically) implemented in the system (often the term 'emergent' is used in this context). Even so, these kinds of rather abstract behavioral states still retain one aspect lost in the mere 'causation talk' of functional architectures, and that is *time*!

4.2 Causation and time

It has been pointed out by philosophers (e.g., see CHALMERS 1997) that there is an essential difference between functional descriptions of physical systems like clocks, combustion engines, CD players, etc. and the functionalist descriptions of minds: in the former case some aspects of the physical structure matter, they are essential to any system realizing the functional architecture. Thus, these physical aspects are (if not explicitly, so then implicitly) retained in the functional architecture, thereby constraining the set of possible realizers. In the latter case, however, it is the very functional structure itself—so it is claimed—that matters, that is, the patterns of causal organization regardless of the underlying physical structure. Therefore, only causal organization, or put differently, 'the flow of causation' is retained in functionalist abstractions from the physical as *the essential aspect* of minds. But is this really true?

Real minds are intrinsically tied to their environments and thus affected by the temporal structures imposed on them. Timing plays a crucial role in every aspect of a cognitive architecture pertaining to the proper functioning and survival of the organism. Many recent studies in cognitive science emphasize the importance of time as opposed to 'mere temporal order' (see, for example, PORT/VAN GELDER 1995).

What distinguishes *time* from mere (temporal) *order* (as implicitly provided by the notion of causality) is that in addition to order a *metric* is defined (on the set of time points), that is, a notion of *distance* in time. This notion of distance in time allows one to differentiate functions according to their temporal behavior that would otherwise be indistinguishable. Take, for example, two microprocessors that work at different clock speeds—functionally they are identical, yet there is an essential difference between them, which is usually also reflected by any price tag put on them: their speed (another example of a function, where time is the distinctive factor, would be vowel production and recognition).

Is it problematic that causation alone does not suffice to capture the temporal structure of cognitive architectures? I would claim: Yes. Imagine two different physical systems that share the same functional specification of a human mind, one a regular human, another the People's Republic of China 'implementing the human brain' at a much, much slower pace (to use BLOCK's example). A human body controlled by the People's Republic of China would fail terribly in the real world, because it could not react to its environment in due time.[13] Well, one might say, it would do just fine if everything surrounding it, that is, its environment had been 'slowed down' appropriately. This objection, however, strikes me as severely flawed, since it would entail *a completely new physics* (as in our physical universe certain processes have to happen at a certain speed otherwise they would not be the kinds of processes they are). Whether a 'slowed down version' of a human mind could control a 'slowed down version' in such a 'slowed down universe' (with possibly completely different physical properties) seems too speculative a question to be taken seriously. What seems to be a productive approach, however, is to ask whether it is possible to understand a certain architecture (that evolved or was designed to meet the temporal constraints of its environment) at a mere causal level? I suspect that the answer would be *no* for systems that are sufficiently complex (like brains of vertebrates or VLSI microchips, for that matter).[14]

If, on the other hand, causal structure were augmented by temporal constraints (i.e., information about distance in time between causally connected states), then this would in theory suffice to capture an essential aspect of possible physical implementations of the functional architecture. It would, for example, allow us to model the functional architecture computationally, i.e., to implement a virtual machine that abides to the temporal constraints (as many computational descriptions can handle temporal metrics, just take programming languages for real-time systems).

Behavioral states, therefore, seem to be an abstraction, which can be implemented computationally, and thus realized physically on computational systems. At the same time, behavioral states are abstract enough to capture aspects of minds that seem to be intrinsically connected to their causal structure and not to their physical realization ("organizational invariants", as CHALMERS 1997, puts it), thereby connecting them to functional descriptions of cognitive architectures.

5. Conclusion

The level of description of behavioral states is *intermediate* and *intermediary*, because it specifies states that could be realized in many different physical ways (in neural architectures, but possibly also in digital ones, and others), yet retains at least one crucial physical and causal aspect not retained in mere functional descriptions: (distance in) time! By explicitly incorporating time, behavioral states make it possible to model the temporally extended interactions among different parts of a cognitive system as well as interactions of the cognitive system with its environment. The level of description of behavioral states might, therefore, not only prove useful for constructing systems that exhibit complex causal interactions (such as minds), but also for explaining *how* functional states are related to physical states by viewing them as (not necessarily disjoint) collections of behavioral states.

Author's address

Matthias Scheutz, Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA, and Department of Philosophy of Science and Social Studies of Science, University of Vienna, Vienna, A-1090, Austria.
Email: mscheutz@cse.nd.edu

Notes

1 The questions of exactly how these states supervene on the physical and in what kinds of structures they are realized are rarely addressed in detail, let alone answered satisfactorily. This is most likely due to the fact that the notions of 'realization' and 'supervenience' are mostly used as unexplained 'primitive' terms in the philosophical literature (which is quite surprising given the theoretical importance and practical consequences that hinge upon them). Although some have attempted more or less precise definitions of 'realization'—e.g., KIM (1996), BLOCK (1996)—these definitions are not very helpful for those who, interested in building minds, are trying to understand the relation between architectures and their implementations.

2 Even with simple functionally specified objects this is problematic. Think of tables as functionally specified, for example, and consider all possible physical implementations of the specification 'table' and what they could possibly have in common at a physical level.

3 Note that this obviously does not hold for all functional specifications: a functional specification of an abstract finite state automaton, for example, can be easily related to physical states in a standard PC by 'implementing' the automaton in a programming language.

4 Of course, the behavior elicited by the organism realizing the cognitive system is specified as well.

5 Note that the existential quantifiers could be viewed as ranging over properties or as picking out particular physical states of the system.

6 The notion of 'bisimilarity' is defined as follows: let I and O be two finite sets (e.g., the sets of input and output states, respectively) and let $M_1 = \langle S_1, \rightarrow_1 \rangle$ and $M_2 = \langle S_2, \rightarrow_2 \rangle$ be two structures with domains S_1 and S_2, respectively, where relation $\rightarrow_1$ is defined over $S_1 \times I \times S_1 \times O$ and relation $\rightarrow_2$ is defined over $S_2 \times I \times S_2 \times O$. These structures are then said to be *bisimilar* if there exists a non-empty relation R between S_1 to S_2 such that for all $s_1 \in S_1$, $s_2 \in S_2$, $i \in I$, and $o \in O$ the following two conditions hold: (1) if $R(s_1,s_2)$ and $(s_1,i) \rightarrow_1 (t_1,o)$, then $(s_2,i) \rightarrow_2 (t_2,o)$ and $R(t_1,t_2)$, and (2) if $R(s_1,s_2)$ and $(s_2,i) \rightarrow_2 (t_2,o)$, then $(s_1,i) \rightarrow_1 (t_1,o)$ and $R(t_1,t_2)$. For a detailed elaboration of the role of bisimulation in a theory of implementation and functional realization, see SCHEUTZ (2001).

7 A note of terminology: while it is common usage to use 'mental states' and 'functional states' to refer to states of an individual's mind, the notion of state is not exclusively used to describe 'static' entities, but often times serves the role of a general term that subsumes states as well as events, i.e., processes. In a sense, the term 'behavioral state' should have been avoided in favor of 'behavioral processes', as the latter emphasizes the dynamic character of the activity taking place in the individual. Following established terminology, however, I will continue using the term 'behavioral state', even if (systematic) dynamic changes in the individual are being referred to.

8 There is evidence that similar mechanisms are at work in animals that inhibit all behaviors with lower activation values, e.g., see LORENZ (1981).

9 While environmental feedback can obviously be simulated with neural networks, the neural architectures that incorporate such feedback will be different from behavioral architectures that perform the same function because of their intrinsic embeddedness in the real world. Since it is one of the design principles of behavioral architectures that they can rely on environmental feedback resulting from the behaviors of activated behavioral states, this property has to be taken into account in modeling cognitive architectures.

10 I am currently investigating various possibilities of implementing simple emulator circuits using behavioral states.

11 Compare this to standard philosophical talk about "pain *causing* wincing and groaning, etc.', where it is never clear whether pain always causes all the behaviors, exactly when the effects surface, whether showing the effects is necessary and/or sufficient for the individual to have pain, etc.

12 Note that it should be possible to derive, beyond the causal propertied, the temporal properties of the 'obstacle-avoidance' state from the interaction of the (physically) implemented states.

13 Many parts of our cognitive system have especially developed to meet time constraints of the environment. There is evidence for neural as well as chemical internal clocks (that work at certain clock rates), oscillator circuits that adapt to external cycles, etc. None of this would work if the system ran at 1/10000th of its regular speed. The same is true for digital circuits that have been designed to work at certain clock rates.

14 It is easy to imagine that nature came up with all kinds of 'hacks' to solve timing problems which could and would have otherwise be implemented very differently. To give an example from computing, imagine a video conferencing system used to transmit video information across the internet. Because of current traffic on the net and to meet real time constraints it only sends partial information of each image, which has to be reconstructed as much as possible from previous images on the other side. It seems that it would be very difficult (if not impossible) to judge out what the system does from the program code alone.

References

Arkin, R. C. (1989) Motor schema-based mobile robot navigation. International Journal of Robotic Research 8(4):92–112.

Block, N. (1996) What is functionalism? In: Borchert, D. M. (ed) The Encyclopedia of Philosophy Supplement. Macmillan: New York, pp. 206–211.

Brooks, R. A (1986) Robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation, Vol. RA-2(1):14–23.

Chalmers, D. J. (1997) A computational foundation for the study of cognition. Retrieved on June 20, 2000, from the internet at http://www.u.arizona.edu/~chalmers/papers/computation.html

Kim, J. (1996) Philosophy of mind. Westview: Boulder CO.

Lorenz, K. (1981) The foundations of ethology. Springer Verlag: New York.

Lorenz, K./Leyhausen, P. (1973) Motivation and animal behavior: An ethological view. Van Nostrand Co.: New York.

McFarland, D. (1981) The oxford companion to animal behavior. Oxford University Press: Oxford, New York.

Port, R./van Gelder, T. (1995) Mind as Motion: Explorations in the Dynamics of Cognition. MIT Press, Cambridge.

McClelland, J. L./Rumelhart, D. E. (1986) Parallel distributed processing, Vol. 1 and 2. MIT Press: Cambridge.

Pfeifer, R./Scheier, C. (1999) Understanding Intelligence. MIT Press: Cambridge.

Scheutz, M. (2000) Surviving in a hostile multi-agent environment: How simple affective states can aid in the competition for resources. In: Hamilton, H. J. (ed) Proceedings of the Thirteenth Canadian Conference on Artificial Intelligence. Springer Verlag: Montreal, pp. 389–399.

Scheutz, M. (2001) Computational vs. causal complexity. Minds and Machines 11:543–566.

Sloman, A. (1998) Supervenience and implementation: Virtual and physical machines. Retrieved on June 20, 2000, from the internet at http://www.cs.bham.ac.uk/research/cogaff/Sloman.supervenience.and.implementation.pdf

David C. Geary and M. Catherine DeSoto

Sex Differences in Spatial Abilities Among Adults from the United States and China

Implications for Evolutionary Theory

Abstract

*Sex differences on tests of spatial abilities were examined for two samples of adults from the United States (U.S.) and China. In Study 1, an inconsistent pattern of sex differences emerged for tests that largely required subjects to mentally rotate representations of geometric figures in two dimensions. A male advantage on the Mental Rotation Test (MRT), a test that requires subjects to mentally rotate representations of geometric figures in three dimensions, was found for both the U.S. (*n *= 66) and Chinese (*n *= 40) samples. Study 2 included larger samples and replicated the sex difference on the MRT. It was also shown that in both the U.S. (*n *= 237) and China (*n *= 218), males were over-represented at the high end of MRT scores, and females were over-represented at the low end of MRT scores. The results support the position that the male superiority in 3-dimensional spatial cognition is not dependent upon culture.*

Key words

Sex differences, spatial abilities, evolution, culture.

A MALE ADVANTAGE IN the ability to generate and mentally manipulate spatial representations of geometric and other figures has been well established in studies conducted in North America and in a host of European nations (GEARY 1998; KIMURA 1999; LINN/PETERSON 1985; MASTERS/SANDERS 1993; VOYER/VOYER/BRYDEN 1995), including Austria (RESCHER/RAPPELSBERGER 1999), England (MCGOWAN/DUKA 2000), Germany (HAUSMANN et al. 2000), Hungary (KARADI et al. 1999), Norway (AMPONSAH/KREKLING 1997), Scotland (JAHODA 1979, 1980), and Sweden (HERLITZ/AIRAKSINEN/NORDSTROEM 1999). AMPONSAH, KREKLING and JAHODA found the same sex differences in Ghana, as did HALPERN/TAN (2001) in Turkey. Across these nations, the sex difference is particularly robust on tests of 3-dimensional spatial cognition (see also MOFFAT/HAMPSON/HATZIPANTELIS 1998), and is especially robust for the Mental Rotation Test (MRT; VANDENBERG/KUSE 1978). For less difficult and other forms of spatial task the pattern of sex differences is more mixed, although a male advantage is common (KIMURA 1999; EALS/SILVERMAN 1994; SILVERMAN/EALS 1992).

In any case, the finding of a robust cross-national male advantage on the MRT has been interpreted as consistent with evolutionary theory (GEARY 1995, 1996, 1998). Specifically, aspects of male–male coalitional competition (e.g., warfare; GEARY 1998) and other sex differentiated activities (e.g., hunting; SILVERMAN/EALS 1992; SILVERMAN et al. 2000) result in more navigation in unfamiliar territory for males than for females in preindustrial societies and presumably throughout human evolution. Sex differences in activities that involve a differential use of 3-dimensional physical space should, in theory, result in the evolution of sex differences in the cognitive and brain systems that enable navigation in and the mental representation of this space (GAULIN 1992; GAULIN/FITZGERALD 1989; GEARY 1995; SILVERMAN et al. 2000). A detailed analysis of the cognitive mechanisms supporting MRT performance suggests that items on the MRT may engage some of the same cognitive (and presumably brain) systems that support navigation in and the representation of 3-dimensional space (JUST/CARPENTER 1985, p165). Indeed, MOFFAT et al. found a large male advantage in the ability to navigate in a 3-dimensional virtual maze and that performance on the maze task was significantly correlated with MRT performance. SILVERMAN et al. found that males were better than females at wayfinding in an unfamiliar wooded area (i.e. knowing one's position in the area, relative to a start point) and that wayfinding performance was correlated with MRT performance, although DABBS and his colleagues reported that MRT performance was not correlated with performance on a paper-and-pencil navigation task (DABBS et al. 1998).

At this point, the relation between sex differences on the MRT and human sex differences, favoring males, in the ability to navigate in a novel 3-dimensional environment remains to be clarified (GEARY 1998; MOFFAT/HAMPSON/HATZIPANTELIS 1998). Nonetheless, it is clear that MRT performance is related to prenatal exposure to and circulating levels of sex hormones (e.g., GAULIN et al. 1997; HAMPSON 1990; HAUSMANN et al. 2000; MCCORMICK/TEILLON 2001; RESNICK et al. 1986; SILVERMAN/PHILIPS 1993; for a review see KIMURA 1999). The relation between sex hormones and MRT performance is, of course, in keeping with the view that the sex difference on this task reflects a more fundamental sex difference in the cognitive and brain systems that support 3-dimensional spatial cognition. The relation is also in keeping with the evolutionary model, as sex hormones are one of the primary mechanisms involved in the proximate expression of evolved sex differences (GEARY 1998).

Despite a clear theoretical rationale for expecting sex differences in complex spatial abilities and evidence that these differences are found across cultural context and covary with levels of sex hormones, CAPLAN/CRAWFORD (1997) recently asserted that sex differences in spatial ability, when they are found, are due to differences in the treatment of males and females and not to biological or evolutionary mechanisms. The above mentioned pattern of sex differences across many different national and cultural contexts argues against this position. Still, the proposition that the male advantage in 3-dimensional spatial abilities is a universal and evolved sex difference would be further bolstered with the demonstration that these differences are evident in East Asian, as well as North American, European, and African nations. However, it appears that sex differences on the MRT have only been assessed in one East Asian population, samples of high-school students from Japan (MANN et al. 1990; SILVERMAN/PHILLIPS/SILVERMAN 1996). In both studies, males outperformed females on the MRT in the Japanese and in U.S. samples. MANN et al. found no sex differences on a second spatial test—the Mazes subtest of the WECHSLER Intelligence Scale for Children–Revised (WECHSLER 1974)—in either nation, and SILVERMAN et al. found only a modest sex difference, favoring males, on a test of 2-dimensional spatial cognition. The results of MANN et al. and SILVERMAN et al. are consistent with a common finding that males outperform females on the MRT and that the pattern of sex differences is often more mixed for less complex spatial measures.

The current analyses were conducted as a response to CAPLAN/CRAWFORD'S (1997) denial of innate sex differences in spatial abilities, and in light of the paucity of information on spatial sex differences in East Asian populations. In the first study, we compare the pattern of sex differences on the MRT, and two other spatial tests, for samples of young adults from the U.S. and mainland China, and in the second assess sex differences on the MRT for larger Chinese and American samples. These comparisons provide an opportunity to replicate the sex difference on the MRT reported by MANN et al. (1990) and SILVERMAN/PHILLIPS/SILVERMAN (1996) for Japanese adolescents. In addition to providing another much needed assessment of sex differences in East Asia, a replication with Chinese samples is important in and of itself. This is because previous cross-national research suggests that individuals from Japan outperform individuals from China and the U.S. on tests of spatial abilities (MANN et al. 1990; SILVERMAN/PHILLIPS/SILVERMAN 1996; STEVENSON et al. 1985), indicating that it cannot be concluded that the sex difference on the MRT in Japan will generalize to other Asian nations. Moreover, unlike Japan, the cultural ethos in mainland China is that of equality of the sexes. Thus, if CAPLAN/CRAWFORD'S (1997) position is correct then a sex difference should not be found on the MRT in China.

Study 1

As noted, the goal was to examine the pattern of sex differences on a battery of spatial ability tests administered to college students in the U.S. and China, as part of a larger study of cross-national differences in arithmetical competencies (GEARY et al. 1996); sex differences were not analyzed as part of the larger study. Of particular theoretical interest is the question of whether the male advantage on the MRT will be found in China.

Methods

Subjects. The subjects were 40 (20 female, 20 male; age range 17 to 22 years) adults from China, and 66 (42 female, 24 male; age range 18 to 42 years) adults from the U.S. The American adults were recruited from undergraduate psychology courses at the University of Missouri, Columbia, and the Chinese adults were recruited from undergraduate courses at East China Normal University, Shanghai, China. The mean age was 19 ($SD = 1$) and 20 ($SD = 4$) years, respectively, for the Chinese and American samples ($p < .05$).

Ability Measures. All subjects were administered a battery of psychometric tests that spanned the Numerical Facility, Perceptual Speed, and Spatial Orientation ability factors (Ekstrom/French/Harman 1976), although only scores on the spatial tasks are considered here; cross-national differences are described in Geary et al. (1996). The spatial tests included the Card Rotations Test and the Cube Comparisons Test, both from the Educational Test Service battery of factor-referenced tests, as well as the MRT (Ekstrom/French/Harman 1976; Vandenberg/Kuse 1978). For each form of all three tests, subjects were allowed 3 min to match rotated test figures against a comparison figure. The Card Rotations Test requires the rotation of figures in 2-dimensional space, whereas the MRT requires the rotation of figures in 3-dimensional space. The Cube Comparisons Test requires the rotation of drawings of cubes, some of which need to be rotated in 2-dimensional space, others in 3-dimensional space. For each test, the score was the number of items correctly identified minus the number of items incorrectly identified (to correct or guessing).

Procedure. Translation. For all measures, the test stimuli were identical in the English and Chinese versions. To ensure comparability, an experienced translator first translated the English instructions into Chinese. Another experienced translator who was not familiar with the English instructions then back translated the Chinese version into English. Discrepancies between the original English instructions and the back-translated instructions were then discussed between the first author and the two translators, which resulted in a second Chinese version of the instructions. To ensure that these instructions were clear, the Chinese versions of all ability measures were then administered to two individuals who were not familiar with either the English or Chinese versions of the tests. Both individuals indicated that the instructions were readily understandable.

Administration. All ability measures were administered in small groups and under standard instructions. The entire testing session lasted about 50 min.

	United States				**China**			
	Male		**Female**		**Male**		**Female**	
Test	M	SD	M	SD	M	SD	M	SD
					Study 1			
CR	117	21	115	28	99	31	90	24
CC	21	10	14	9	12	8	15	9
MRT	21	9	12	8	16	11	12	9
					Study 2			
MRT	18	10	13	8	19	9	13	7

Table 1. Mean cross-national and sex differences for tests of spatial cognition.

Results

Mean test scores across nation and sex are shown in the top portion of Table 1. Scores on each test were submitted to a 2 (nation) by 2 (sex) analysis of variance. For the Card Rotations Test, neither the main effect for sex ($F(1,102) < 1$) nor the nation by sex interaction were significant ($F(1,102) < 1$), although the scores for the U.S. sample were higher than those for the Chinese sample ($F(1,102) = 15.83$, $p < .001$). For the Cube Comparisons Test, the main effect for sex was not significant ($F(1,102) = 1.21$, $p > .25$), but main effect for nation ($F(1,102) = 4.82$, $p < .05$) and the nation by sex interaction was ($F(1,102) = 6.62$, $p < .05$). Examination of Table 1 reveals that males had higher Cube Comparisons Test scores in the U.S. sample, but females had higher scores in the Chinese sample. Follow-up analyses revealed that the sex difference was significant for the U.S. sample ($F(1,102) = 8.22$, $p < .01$), but not for the Chinese sample ($F(1,102) < 1$). Finally, the main effect for sex, favoring males in both samples, was significant for the MRT ($F\ (1,102) = 14.55$, $p < .001$), but the main effect for nation ($F(1,102) = 2.15$, $p > .10$) and nation by sex interaction was not ($F(1,102) = 1.55$, $p > .20$).

Discussion

The findings for the Card Rotations Test and the Cube Comparisons Test suggest that for spatial tasks that largely require the mental rotation of 2-dimensional geometric figures, there may be no sex differences, or a different pattern of sex differences depending on culture or the particular sample assessed. These findings may not, however, generalize to more complex 2-dimensional spatial tasks, as COLLINS/KIMURA (1997) found a large male advantage for a 2-dimensional task that was more complex than the tests used in this study. In any case, the sex difference, favoring males, in the ability to mentally rotate 3-dimensional representations of geometric figures (i.e., MRT performance) was evident for young adults from both the U.S. and China. Moreover, the finding of a nonsignificant nation by sex interaction for the MRT indicated that the size of the male advantage did not differ across samples. Although the findings for the MRT are of theoretic interest, as noted earlier, they are in need of replication, given the small sample sizes.

Study 2

To assess further the cross-national sex difference for performance on the MRT, data from additional samples of young adults from the U.S. and China were re-analyzed (GEARY et al. 1999). As with the first study, the goal of the original research was to assess the pattern of cross-national differences in arithmetical competencies. The MRT was included as a contrast measure, that is a measure, unlike mathematical tests, for which there do not appear to be national differences comparing adults from the U.S. and China.

Methods

Subjects. The subjects were 237 (113 male, 123 female; one participant did not provide information on sex; 84% Caucasian) general psychology students from the University of Missouri, Columbia, and 218 (108 male, 110 female) undergraduate students from East China Normal University, Shanghai, China. The U.S. students received partial course credit for participating in the study, whereas the Chinese students received a small payment. The mean age of both samples was 19 years.

Ability Measures and Procedure. In addition to the MRT, all subjects were administered a battery of arithmetical computation and arithmetical reasoning tests, as well as an intelligence test; no other spatial tests were administered. The same translation and administration procedures described for Study 1 were followed (see GEARY et al. 1999).

Results

Mean sex differences. As shown in the bottom portion of Table 1, the pattern of cross-national performance and sex differences replicates that found for Study 1. More precisely, neither the main effect for nation (F (1, 450) = 1.05, $p > .25$) nor the nation by sex interaction (F (1, 450) < 1) were significant, but the main effect for sex was (F (1, 450) = 58.77, $p < .0001$).

Sex differences in high and low scores. The relatively large sample sizes allowed for an assessment of sex differences at the high and low ends of the distributions of MRT scores. For the U.S. sample, 75% (45 of 60) of the individuals in the top quartile were male, whereas 68% (40 of 59) of the individuals in the bottom quartile were female ($\chi^2(1) = 21.92$, $p < .001$). For the Chinese sample, 72% (41 of 57) of the individuals in the top quartile were male, and 72% (44 of 61) of the individuals in the bottom quartile were female ($\chi^2(1) = 22.89$, $p < .001$). The same trend was found for comparisons of the top and bottom deciles. For the US sample, 86% (24 of 28) of the individuals in the top decile were male, whereas 63% (15 of 24) of the individuals in the bottom decile were female ($\chi^2(1) = 12.96$, $p < .001$). For the Chinese sample, 81% (22 of 27) of the individuals in the top decile were male, and 64% (18 of 28) of the individuals in the bottom decile were female ($\chi^2(1) = 11.83$, $p < .001$).

Discussion

The cross-national sex difference, favoring males, on the MRT was replicated, and it was demonstrated that for both the U.S. and Chinese samples the majority of high scoring individuals were male and the majority of low scoring individuals were female. The differences were especially pronounced at the high end of the distributions of MRT scores. In the top ten percent of scores, the ratio of males to females was 6:1 in the U.S. sample and 4.4:1 in the Chinese sample. In the bottom 10 percent of scores, there were more than three females for every two males in the U.S. sample (female to male ratio of 1.67:1) and nearly two females for every male in the Chinese sample (female to male ratio of 1.8:1).

Summary and General Discussion

A male advantage on the MRT was found for young adults from the U.S. and China, and the magnitude of this sex difference did not differ across nations. The results are in keeping with the findings of MANN et al. (1990) and SILVERMAN/PHILLIPS/SILVERMAN (1996) of a male advantage on the MRT for samples of Japanese high-school students, and consistent with similar studies conducted in North America, Europe, and Africa (e.g., MCGOWAN/DUKA 2000; VOYER/VOYER/BRYDEN 1995). The current results thus confirmed the robustness of the male advantage on the MRT, and cast doubt upon the position that the sex difference in 3-dimensional spatial cognition is due to cultural factors or to differences in the treatment of males and females (CAPLAN/CRAWFORD 1997). The results for China are especially difficult to reconcile with CAPLAN and CRAWFORD's views, given the social ethos of equality of the sexes.

The sex difference, of course, does not mean that MRT performance, or spatial abilities in general, are not influenced by exploration of and navigation in physical space, as such exploration and other spatial-related experiences are clearly related to improved spatial abilities (MATTHEWS 1992). Rather, it is most likely that the male advantage on the MRT arises from both hormonal and experiential factors, and that the same hormonal mechanisms that facilitate spatial performance are likely to prompt males and females to differentially engage physical space. The sex difference, favoring males, in engagement in physical space, such as exploration of unfamiliar areas (see MATTHEWS 1992), is likely to interact with hormonal influences on the cognitive and brain systems that support spatial cognition to produce the observed differences (GEARY 1998).

Although the data do not directly address the position that male–male competition (e.g., warfare) and other sex-differentiated activities (e.g., hunting) has resulted in the evolution of a male advantage in 3-dimensional spatial competencies (GEARY 1998; SILVERMAN et al. 2000; SILVERMAN/EALS 1992), they are consistent with this position. Stated otherwise, the current studies provided an opportunity, especially in light of the ethos of sexual equality in mainland China, to reject the evolutionary hypothesis. That the hypothesis was not rejected provides further support for the view that the male advantage in complex spatial cognition has a biological basis, and makes cultural (e.g., stereotypes) models of these sex differences increasingly untenable (see also HALPERN/TAN 2001).

Authors' address

David C. Geary, Dept. of Psychological Sciences, 210 McAlester Hall, University of Missouri, Columbia, MO 65211-2500, USA. Email: gearyd@missouri.edu
M. Catherine DeSoto, University of Northern Iowa, Dept. of Psychology, 342 Baker Hall, University of Northern Iowa, Cedar Falls, IA 50614-0505, USA.
Email: cathy.desoto@uni.edu

Acknowledgments

We would like to thank Guo-Peng CHEN, Melissa MOON, and Ravindran SABAPATHY for their assistance with data collection and collation, and Fan LIU and Zhitang LIU for their assistance with the translations. The research was supported by grant 1R01-HD27931 from the National Institute of Child Health and Human Development and a grant from the Research Board of the University of Missouri, both awarded to the first author.

References

Amponsah, B./Krekling, S. (1997) Sex differences in visual-spatial performance among Ghanaian and Norwegian adults. Journal of Cross-Cultural Psychology 28:81–92.

Caplan, P. J./Crawford, J. B. (1997) Do sex-related cognitive differences exist, and why do people seek them out? In: Caplan, P. J./Crawford, M./Hyde, J. S./Richardson, J. T. E. (eds) Gender differences in human cognition: Counterpoints. Oxford University Press: New York, pp. 52–80.

Collins, D. W./Kimura, D. (1997) A large sex difference on a two-dimensional mental rotation task. Behavioral Neuroscience 111:845–849.

Dabbs, J. M. Jr./Chang, E.-L./Strong, R. A./Milun, R. (1998) Spatial ability, navigation strategy, and geographic knowledge among men and women. Evolution and Human Behavior 19:89–98.

Eals, M./Silverman, I. (1994) The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. Ethology and Sociobiology 15:95–105.

Ekstrom, R. B./French, J. W./Harman, H. H. (1976) Manual for kit of factor-referenced cognitive tests 1976. Educational Testing Service: Princeton NJ.

Gaulin, S. J. C. (1992) Evolution of sex differences in spatial ability. Yearbook of Physical Anthropology 35:125–151.

Gaulin, S. J. C./Fitzgerald, R. W. (1989) Sexual selection for spatial-learning ability. Animal Behaviour 37:322–331.

Gaulin, S. J. C./Silverman, I./Phillips, K./Reiber, C. (1997) Activational hormonal influences on abilities and attitudes. Evolution and Cognition 3:191–199.

Geary, D. C. (1995) Sexual selection and sex differences in spatial cognition. Learning and Individual Differences 7:289–301.

Geary, D. C. (1996) Sexual selection and sex differences in mathematical abilities. Behavioral and Brain Sciences 19:229–284.

Geary, D. C. (1998) Male, female: The evolution of human sex differences. American Psychological Association: Washington DC.

Geary, D. C./Liu, F./Chen, G.-P./Saults, S. J./Hoard, M. K. (1999) Contributions of computational fluency to cross-national differences in arithmetical reasoning abilities. Journal of Educational Psychology 91:716–719.

Geary, D. C./Salthouse, T. A./Chen, G.-P./Fan, L. (1996) Are American versus East Asian differences in arithmetical ability a recent phenomenon? Developmental Psychology 32:254–262.

Halpern, D. F. /Tan, U. (2001) Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. Brain and Cognition 45:392–414.

Hampson, E. (1990) Estrogen-related variations in human spatial and articulatory-motor skills. Psychoneuroendocrinology 15:97–111.

Hausmann, M./Slabbekoorn, D./Van Goozen, S. H. M./Cohen-Kettenis, P. T./Guentuerkuen, O. (2000) Sex hormones affect spatial abilities during the menstrual cycle. Behavioral Neuroscience 114:1245–1250.

Herlitz, A./Airaksinen, E./Nordstroem, E. (1999) Sex differences in episodic memory: The impact of verbal and visuospatial ability. Neuropsychology 13:590–597.

Jahoda, G. (1979) On the nature of difficulties in spatial-perceptual tasks: Ethnic and sex differences. British Journal of Psychology 70:351–363.

Jahoda, G. (1980) Sex and ethnic differences on a spatial-perceptual task: Some hypotheses tests. British Journal of Psychology 71:425–431.

Just, M. A./Carpenter, P. A. (1985) Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. Psychological Review 92:137–172.

Karadi, K., Szabo, I./Szepesi, T./Kallai, J./Kovacs, B. (1999) Sex differences on the hand mental rotation task for 9-yr.-old children and young adults. Perceptual and Motor Skills 89:969–972.

Kimura, D. (1999) Sex and cognition. MIT Press/Bradford Books: Cambridge MA.

Linn, M. C./Peterson, A. C. (1985) Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development 56:1479–1498.

Mann, V. A./Sasanuma, S./Sakuma, N./Masaki, S. (1990) Sex differences in cognitive abilities: A cross-cultural perspective. Neuropsychologia 28:1063–1077.

Masters, M. S./Sanders, B. (1993) Is the gender difference in mental rotation disappearing? Behavior Genetics 23:337–341.

Matthews, M. H. (1992) Making sense of place: Children's understanding of large-scale environments. Barnes & Noble Books: Savage MD.

McCormick, C. M./Teillon, S. M. (2001) Menstrual cycle variation in spatial ability: Relation to salivary cortisol levels. Hormones & Behavior 39:29–38.

McGowan, J. F./Duka, T. (2000) Hemispheric lateralisation in a manual-verbal task combination: The role of modality and gender. Neuropsychologia 38:1018–1027.

Moffat, S. D./Hampson, E./Hatzipantelis, M. (1998) Navigation in a "virtual" maze: Sex differences and correlation with psychometric measures of spatial ability in humans. Evolution and Human Behavior 19:73–87.

Rescher, B./Rappelsberger, P. (1999) Gender dependent EEG-changes during a mental rotation task. International Journal of Psychophysiology 33:209–222.

Resnick, S. M./Berenbaum, S. A./Gottesman, I. I./Bouchard, T. J. Jr. (1986) Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. Developmental Psychology 22:191–198.

Silverman, I./Choi, J./Mackewn, A./Fisher, M./Moro, J./Olshansky, E. (2000) Evolved mechanisms underlying wayfinding: Further studies on the hunter-gatherer theory of spatial sex differences. Evolution and Human Behavior 21:201–213.

Silverman, I./Eals, M. (1992) Sex differences in spatial abilities: Evolutionary theory and data. In: Barkow, J. H./ Cosmides, L./ Tooby, J. (eds) The adapted mind: Evolutionary psychology and the generation of culture. Oxford University Press: New York, pp. 533–553.

Silverman, I./Phillips, K. (1993) Effects of estrogen changes during the menstrual cycle on spatial performance. Ethology and Sociobiology 14:257–270.

Silverman, I./Phillips, K./Silverman, L. K. (1996) Homogeneity of effect sizes for sex across spatial tests and cultures: Implications for hormonal theories. Brain and Cognition 31:90- 94.

Stevenson, H. W./Stigler, J. W./Lee, S.-Y./Lucker, G. W./Kitamura, S./Hsu, C. C. (1985) Cognitive performance and academic achievement of Japanese, Chinese, and American children. Child Development 56:718–734.

Vandenberg, S. G./Kuse, A. R. (1978) Mental rotation, a group test of three-dimensional spatial visualization. Perceptual and Motor Skills 47:599–604.

Voyer, D./Voyer, S./Bryden, M. P. (1995) Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. Psychological Bulletin 117:250–270.

Wechsler, D. (1974) Manual for Wechsler Intelligence Scale for Children–Revised. Psychological Corporation: San Antonio TX.

Theresa Schilhab

Man And The Mirror

The Anthropocentric Perspective on Animal Self-Recognition in Mirrors

Abstract

This paper addresses how attribution of mind-reading abilities to human children and nonhuman primates differs in studies on mirror-based self-recognition employing the so-called mark test. In children, the employed criteria are anthropocentric, since the verbal response is accentuated. A close scrutiny of the test procedure also reveals methodological inadequacies thus questioning the widespread use of mentalistic interpretations. In nonhumans, the criteria are actually a set of anthropomorphic behaviours. This renders the application of criteria of self-recognition irrelevant to nonhumans, and leaves the results vulnerable to nonmentalistic interpretations. Finally, I briefly discuss what the behavioural resemblance of nonhuman organisms to humans signals about their minds and why our intuition gives guidelines concerning attribution of minds in nonhuman organisms.

Key words

Mirrors, self-recognition, mark test, children, nonhuman primates, anthropocentrism, anthropomorphism.

Introduction

In the past two decades, interest in theory of mind in nonhuman primates has increased considerably, as it could potentially reveal the origin of the human mind. The "theory of mind"–hypothesis claims that nonhuman primates reason about minds, implying that they think about themselves and others in terms of mental states (see HEYES 1998). Several empirical studies have addressed this hypothesis by focussing on abilities such as deception (e.g., WOODRUFF/PREMACK 1979; CHENEY/SEYFARTH 1990; MITCHELL 1991); perspective-taking (HARE et al. 2000; POVINELLI/NELSON/BOYSEN 1990; POVINELLI, NELSON/BOYSEN 1992; POVINELLI/PARKS/NOVAK 1991; PREMACK/WOODRUFF 1978; TOMASELLO/CALL/HARE 1998) or self-recognition with aid of mirrors (e.g., GALLUP 1970; LETHMATE 1973; LIN/BARD/ANDERSON 1992; POVINELLI/RULF/LANDAU/BIERSCHWALE 1993; POVINELLI 1994; WESTERGAARD/HYATT 1994; SHILLITO/GALLUP/BECK 1999). (For a review on self-recognition in nonhuman primates with aid of mirrors see DE VEER/VAN DEN BOS 1999, for a list of self-recognition studies in apes, see TOMASELLO/CALL 1997, pp. 332–333).

However, opinions on the interpretations of the empirical results diverge. While many endorse mentalistic interpretations (GALLUP 1977; GALLUP/POVINELLI 1993; EDDY/GALLUP/POVINELLI 1996; POVINELLI et al. 1997), HEYES (1998) claims that the results are equally compatible with a nonmentalistic interpretation. This point of view has been advanced particularly with results obtained on self-recognition (EPSTEIN/LANZA/SKINNER 1981; HEYES 1993, 1994, 1995). Despite prolonged discussions between proponents and opponents of mentalistic interpretations of how nonhuman primates reason about themselves and others, the debate has not ceased yet. In what follows I trace some reasons why unambiguous evidence of mind reading abilities in nonhuman primates as they appear in mirror-guided self-recognition is difficult to achieve. Since we are not telepathists, attribution of mind-reading abilities rests on behavioural criteria with both humans and nonhuman primates. In the words of primatologist Andrew WHITEN (1996, p277): "…the recognition of another's state of mind must somehow rest on observation of certain components within the complex of others' *behaviour patterns* together with their *environmental context*: that's all we can see —we can't see their minds in the direct way suggested by the idea of telepathy". I claim that while attribution of self-recognition in humans is naturally based on or at least incorporates verbal behaviour as criterion, it rests on non-verbal behaviour in nonhuman primates. However, being criterial of mind-reading abilities there is a qualitative difference between verbal and non-verbal behaviour. If a human makes a claim about another person's mental states, this statement is indicative of mind-reading abilities. Just as when a child at the age of four claims, that the belief of a peer is the cause of his behaviour[1]. To make such a claim *is* to reason about minds, and therefore these children have mind-reading abilities. There is no comparable situation when dealing with non-verbal primates. For instance, we can never be absolutely convinced that a lower-ranking animal's intense gazing or begging attitudes towards a dominant animal show mind-reading abilities in the beggar, since the behaviour could be elicited without reasoning about minds. Therefore, we do not have to assign mind-reading abilities to the beggar. Being based exclusively on non-verbal behaviours the results obtained with nonhuman primates will seldom be immune to nonmentalistic interpretations, unless we have reasons to believe that the behaviour is causally linked to certain mental states. Even if we are convinced, the assumption most probably stems from the familiarity of the attitude to human behaviour—that is, the familiarity of "certain components within the complex of others' behaviour patterns together with their environmental context". The inclination to attribute mental states to nonhumans from behavioural familiarity to human behaviour is normally named "anthropomorphism"—that is "attribution of human characteristics to nonhuman things and events" (GUTHRIE 1997). Among philosophers as well as ethologists it is generally believed that anthropomorphism is a fallacy and as such should be shunned, since it involves invalid methods and counteracts "objective science" (KENNEDY 1992). For a critical discussion of this position see FISHER (1991).

However, I hold that for self-recognition in nonhumans we apply anthropocentric criteria and contrary to common practice, I will provide a justification of this propensity. In the following, I will press these points by focussing on representative examples from empirical tests of self-recognition as developed by Gordon G. GALLUP Jr. (1970) done with nonhuman primates and children. This so-called "objective technique for determining the presence of a self-concept" (AMSTERDAM 1972, p297) is generally considered to be similarly applied to humans as well as nonhumans. However, comparison between attribution of self-recognition to children and to nonhuman primates reveals that the employed criteria differ qualitatively. My focus will be on how these criteria differ to show that criteria of attribution of self-recognition are largely anthropocentric. To combat the charge of the anthropomorphic fallacy I sketch a theory, which I have named "phylogenetic stance", that could provide the necessary theoretical basis for applying anthropocentric criteria of self-recognition in nonhuman primates (Schilhab submitted).

An "Objective Technique for Determining the Presence of a Self-Concept"

The "mirror and mark test" was introduced as an experimental analysis of self-recognition in chimpanzees by GALLUP (1970). In this initial study, four preadolescent chimpanzees (*Pan troglodytes*), that had no experience with mirrors, were tested for over a week at a total exposure time of 80 hours to the mirror. Normally chimpanzees confronted with a mirror respond with social behaviour such as threatening, vocalisation and bobbing. But gradually self-directed behaviour such as grooming parts of the body that is not visually accessible without the aid of the mirror, takes over. A likely explanation of this phenomenon is that the chimpanzee learns to recognise the mirror image as an image of self.

To substantiate the hypothesis that self-directed behaviour is associated with self-recognition, the chimpanzees were sedated and while being unconscious they were marked with an odourless dye on the uppermost part of one eyebrow and the opposite ear. After having fully recovered the chimpanzees were observed for 30 minutes without the mirror to account for any spontaneous touches of the marked area. Then the mirror was reintroduced and the chimpanzees were again observed for mark-directed behaviour.

The number of incidences of mark-directed behaviour rose from one during the post anaesthesia mirror-less period to four to ten in the period where the mirror was reintroduced. This was interpreted as a clear indication of chimpanzees being capable of self-recognition with the aid of the mirror.

Following the empirical method devised by GALLUP and its derivatives, other studies on chimpanzees have come to the same conclusion (e.g., EDDY/GALLUP/POVINELLI 1996; LETHMATE/DÜCKER 1973; LIN/BARD/ANDERSON 1992; POVINELLI/RULF/LANDAU/BIERSCHWALE 1993; SUAREZ/GALLUP 1981) including the closely related bonobo[2] (HYATT/HOPKINS 1994; WESTERGAARD/HYATT 1994) although not all chimpanzees can be shown to exhibit this capacity (SWARTZ/EVANS 1991, 1994). In addition orang-utans have been shown to elicit the reported type of behaviour in front of a mirror, suggesting that the behaviour is not idiosyncratic to chimpanzees among the greater apes (SUAREZ/GALLUP 1981). On the other hand, neither gibbons (LETHMATE/DÜCKER 1973) nor gorillas (SHILLITO/GALLUP/BECK 1999; SUAREZ/GALLUP 1981), also belonging to the hominoids, have consistently succeeded in the mirror and mark tests (for an exception on the gorilla results see PATTERSON/COHN 1994 and on gibbon results see UJHELYI et al. 2000).

Children

In an equivalent study on humans (AMSTERDAM 1972), children between three and 24 months were subject to a version of the mark test, developed prior to GALLUP's research, but resembling it in important aspects (AMSTERDAM 1972, p304, note 1). In this test a spot of rouge was placed on the child's nose serving as "….a point of reference for evaluating self-recognition in the mirror. The spot helped to focus attention on the face, the part of the body that it was desired that the child should examine and recognize" (AMSTERDAM 1972, p297). The claim is that from three to 24 months children go through three discrete though overlapping phases in their reaction to the mirror image. In phase one—from three to 12 months of age they treat it as a "sociable playmate". In the second phase from 13 to 20 months the children mostly withdraw from the mirror image in avoidance. In the third phase from 20 to 24 months children show recognition of their image (for a developmental sequence of self-awareness and its relation to mirror self recognition, see PARKER/MITCHELL 1994). "Self-recognition was scored if the child touched the actual spot or used the mirror to examine his nose" (AMSTERDAM 1972, p297)[3].

When working with children, the mother of the child plays a crucial role demonstrated by the instructions offered by the researcher. "The observer then said, " I'm going to show you what I'd like you to do with your child"…. "In just a minute I'd like you to come around here with (child's name). When you are next to the mirror, place him facing the mirror"… "then you point to his face in the mirror and say: "See, see, see". The observer pointed in the mirror each time she said "See"… "Then point to his face again and ask him, Who's that?"… Each observation was timed from the moment that the subject was placed in front of the mirror. After 2 min, the mother was asked to say "See" again three times, and ask, "who's that?" This was repeated a third time after another 2 min elapsed. Each subject was given three trials before the mirror. If the subject cried for more than about 30 sec, his mother was asked to pick him up and hold him until it was time for the next trial when he was again placed before the mirror" (AMSTERDAM 1972, p298).

Interpretation

Though the method adopted in the self-recognition test is considered similar for children and nonhumans (for researchers that compare the results directly, see LIN/BARD/ANDERSON 1992) the criteria differ radically[4]. This is evident from the study by AMSTERDAM, in which the entire performance is linguistically attended. However, my focus will be on what behavioural attitudes count as indicative of self-recognition in nonhuman primates and children respectively.

In studies of nonhuman primates, it is indicative of self-recognition if mark-directed behaviour increases considerably with the mirror present. However, GALLUP (e.g., GALLUP 1994; POVINELLI/GALLUP 1993) emphasises that passing the mark test was never meant as a decisive criterion of a sense of self (for a discussion of the criteria, see MITCHELL 1995). GALLUP/POVINELLI (1993, p327) claim: "…the use of unobtrusively applied facial marks was developed only as a means of validating impressions which arose out of seeing the animals use mirrors to respond to themselves in ways which suggested that they realized that their behavior was the source of the behavior being seen in the mirror". Thus, successful performance in the mark test is but one criterion of a set that should be met to be assigned self-recognition abilities.

The next question of interest is, what does self-recognition ability more specifically refer to? According to GALLUP (1970, p87): "…insofar as self-recognition of one's mirror image implies a concept of self, these data would seem to qualify as the first experimental demonstration of a self concept in a subhuman form". The rationale is that the mirror guidance of mark-directed behaviour is possible only if the primate has a concept of self (GALLUP 1977). Following ANDERSON (1984, p36):"…in order for an individual to correctly appreciate the source of the individual depicted in the mirror, there must be a mental representation of self onto which this perception of the reflection is mapped".

However, some sceptics claim that what the chimpanzees learn while exposed to the mirror, is nothing but an acknowledgement of a correlation between images in the mirror and real world co-ordinates—a position named the "associationist hypothesis"

(KENNEDY 1992, p105). In the words of HEYES (1998, p105): "…even if there were evidence that certain primates have this capability (mirror self-recognition) it would not imply the possession of a "self-concept". What makes certain primates perform successfully on the mirror self-recognition test is the possession of a "body-concept", which entails that "an animal must be able to distinguish, across a fairly broad range, sensory inputs resulting from the physical state and operations of its own body from sensory inputs originating elsewhere" which "does not relate to a mental category". Thus successful performance in the test occurs on purely nonmentalistic grounds (HEYES 1994, 1998). However, MITCHELL (1993, p315) who actually acknowledges the relation between self-awareness and mirror self-recognition claims that "…this self-awareness need not be as elaborate as GALLUP believes".

In comparison, in the self-recognition studies of children, self-directed behaviour (touching the mark on the nose) in front of a mirror *or* verbal claims to who is in the mirror are indices of self-recognition (AMSTERDAM 1972; PRIEL/DE SCHONEN 1986; ROBINSON et al. 1990). Especially verbal claims are considered valid indicators of a self-concept (BIGELOW 1981). This is also reflected in MANS/CICCHETTI/SROUFE (1978, p1247): "A child of two years commonly uses self-referents such as "mine" and "me" thereby demonstrating that a self-concept has been well formulated". In the study of AMSTERDAM (1972), self-directed responses were accepted *because* of the contingent naming of self. (For studies of preverbal children where increased mark-directed responses are accepted as criterion of self-recognition, see MANS/CICCHETTI/SROUFE (1978) and HILL/TOMLIN (1981)).

ROBINSON et a. (1990) supports the notion of attributing decisive weight to naming of self as criterion for self-recognition. This study showed that children develop a skill for using the mirror as a perceptual tool *before* reaching the age where they start using mirrors for self-recognition as demonstrated by naming of self. Children could locate objects without verbal self-identification of their mirror image. Thus, children can make sense of mirror images of objects without understanding the identity of the reflected self image.

This agrees with studies claiming that self-recognition develops successively in infancy (BERTENTHAL /FISCHER 1978), which considers mark-directed response to be the penultimate stage and verbal naming to be the ultimate stage of self-recognition. In accordance, division between making sense of mirror images and naming of self was demonstrated by an experiment on children aged 18–28 months (BIGELOW 1981). Presented with mirror images, ongoing images on videos, play back videos of self and of others in this order, children who recognised their own mirror images as measured by the mark test did not respond adequately verbally. They named all child images, self-images.

Crucial to the argument is that apart from JOHNSON (1983), researchers seem to think that while self-directed behaviour by itself is not indicative of self-recognition, naming of self is. This shows that verbal behaviour is decisive with respect to attribution of self-recognition.

The Decisive Feature of Verbal Reports

Since researchers on self-recognition in children ascribe decisive weight to verbal behaviour, they implicitly give priority to verbal behaviour as a warrant of mind-reading abilities. Is this warranty convincing? Is it not possible to think of a child that can name "self" in front of a mirror but do not recognise the image as self? What restricts one from arguing that when human children pass mirror tests, their performance could just as easily be explained in terms of simple behavioural patterns as expressed by nonmentalistic terms? What restricts one from the obvious conclusion that what the researcher sees as documentation of a concept of the self, is but documentation of the children having acquired the rules of application of a verbal concept of self? The children simply demonstrated that at the age of two they learned to answer the mothers request "who's that?" by applying a certain word (their name) in front of a mirror. The use of the personal pronoun and self-naming in front of the mirror by children could just as well be a learned socially reinforced response. Perhaps at the age of two this is all there is to self-recognition.

It *is* indeed thinkable that naming of self in front of mirrors does not imply self-recognition. The findings of BIGELOW (1981) and JOHNSON (1983) clearly demonstrate that naming of self is not always a warrant of self-recognition, since some children applied their own name or personal pronoun when viewing the image of another child.

The acceptance of verbal response as a sign of self-recognition has been questioned (e.g., ANDERSON 1984). However, when researchers take naming of self as showing recognition of self, they basically employ a theory of "reliability of claims about first-person experiences" which is characteristic of human interaction. They simply rely on verbal reports. Despite it is accepted that the warrant is questionable with young children, the scepticism disappears with older children. Thus, ANDERSON (1984 p. 40) claims; "…few would doubt that verbal self-reference is a sure indicator of self-recognition in older children and adults".

According to the philosopher Alwin I. GOLDMAN (2000), a sort of theory about the reliability of claims is systematically used in parts of the cognitive sciences (a process in which conversion of information from the first-person to third-person perspective is essential). When humans report on their experiences, the report is seldom disputed. This feature of communication is essential to the theory of "heterophenomenology" about conversion of first-person to third-person perspectives by the philosopher Daniel DENNETT (1991).

However, when analysing the consistency between the statements and the actual performance in 3- and 4-year-olds about where to find hidden candy, consistency is lacking in the 3-year-olds (POVINELLI/DEBLOIS 1992). While they can use the pointing of the experimenter to find the candy, they fail in responding correctly to control questions that test their knowledge justification. Thus, calls for reservations regarding the verbal responses of two year olds seem urgent.

Implications

To summarize: Verbal reports alone cannot warrant self-recognition. Unless other criteria such as internal consistency between mirror-related behaviour and reference to self has been demonstrated. This is exactly my point—attribution of mirror self-recognition is reasonable if a complex set of criteria are met. In research on humans, justification of self-recognition involves a number of behaviour patterns, verbal as well as nonverbal.

Though it is crucial to the discussion of how to assign mental states to others, the *complexity* of criteria that have to be met seems to escape the debate on the justification of the mark test. Indirectly GALLUP acknowledges this as demonstrated from his emphasis on the mark test as a *validation* of impressions (POVINELLI/GALLUP 1993), thus principally leaving the mark test as a mere corroboration of former impressions. Still, as demonstrated by the empirical research in nonhumans, the test is in fact treated as the single criterion being met.

However, if it is generally accepted, that verbal self-identification is decisive of recognition of self in humans, comparable investigations of non-verbal organisms will suffer from the incommensurability of verbal criteria. Thus, from the anthropocentric perspective results obtained with non-verbal organisms can never be immune to behaviouristic objections. The conclusion is unavoidable although children and most great apes display similar behaviours in front of mirrors. Conversely, interpretations of the performance of children though displaying similar behaviour are fundamentally mentalistic, because of the simultaneous appearance of verbal self-naming behaviour.

The propensity to find verbal responses conclusive evidence of certain mental states is expressed by the philosopher Peter K. SMITH (1996, p350): …the reason the performance of the human children is convincing (for me) is not so much the rapid learning… but the fact that they can verbalise their reasons ('your eyes aren't there', you can't see me')".

The Phylogenetic Stance

If the criteria are anthropocentric, does this mean that there is no way of demonstrating self-recognition in organisms different from humans? If we accept, that attribution of self-recognition abilities consists of a set of behaviours this needs not be so.

I believe that we already possess a measure by which we can infer self-recognition in nonhuman organisms. I also believe that this is what proponents of self-recognition in nonhuman primates implicitly employed all along but never spelled out. With the investigations on self-recognition in nonhumans and children in mind, most of the observed behaviours, which GALLUP refers to as "impressions" are obviously nearly identical. When GALLUP (1970, p86) observes: "picking extraneous material from the nose by inspecting the reflected image" or "making faces at the mirror" as examples of self-directed behaviours these are strikingly similar to "observes own images as he moves body or part of it" from the "mirror behaviour checklist" by AMSTERDAM (1972, p300). The behavioural resemblance is no coincidence but a direct manifestation of the common ancestry of chimpanzees and humans.

The reason we intuitively attribute self-recognition to nonhuman primates is that these organisms resemble humans in profound ways. I call this inclination to perceive certain nonhuman behaviours as human-like, the adoption of "the phylogenetic stance"[5] (SCHILHAB, submitted). The resemblance of the term to the vocabulary invented by DENNETT (e.g., 1996) is intentional. By this I wish to claim, that the phylogenetic stance is a point of view (biologically implanted and implicitly employed) that helps us categorise living organisms. By this reasoning, those nonhuman organisms being more behaviourally similar to humans are likely to show more mental-like abilities (self-recognition being a subcategory) than those sharing less.

Now, the above stipulation of this inclination goes under the name of anthropomorphism, a position deemed untenable in the days of behaviourism (ROLLIN 1998). Despite renunciation of anthropomorphism the above analysis of mirror self-recognition as well as empirical research suggest that people actually anthropomorphise when dealing with questions of mental states in nonhumans (EDDY/GALLUP/POVINELLI 1993; HERZOG/GALVIN 1997; MITCHELL 1997). Though, even if some scientists and non-scientists do adopt anthropocentric criteria, it is not obvious that they should do so or is it? The philosopher Elliott SOBER argues that in some cases anthropomorphism is preferable since it aligns with the scientific ideal of parsimony (SOBER 1998).

Thus, the phylogenetic stance comprises more than simply reflecting the human inclination to perceive other organisms from a human perspective. I claim that the familiarity of behaviour as reflected by the phylogenetic stance is widely based on real-life familiarity, i.e., phylogenetic closeness, because of the adaptive value of behaviour (for a thorough cladistic analysis of these points, see SOBER 2000). For instance, a common behaviour in great apes not shared with other mammals is the manner of locomotion, which depends on their much more mobile arms (BYRNE 1995). The ape shoulder blade can slide over the rib cage and allows apes to hang beneath branches with ease. Though, convergent evolution of behaviours—such as flight of bats and birds—that are similar to each other for reasons other than common ancestry—could pose a problem to the validity of the phylogenetic stance.

However, I have two answers to this challenge, one based on methodology and one based on theory. Convergent evolution as expressed by homoplasy is likely to result in false interpretations only in so far the traits at stake are taken one by one. Nevertheless, the intuition as expressed by the phylogenetic stance usually involves more than a single feature. Fundamental differences are therefore likely to reveal themselves.

A more theoretical consideration results from conceiving of particular behaviours as adaptations. In this, I follow the lines of thought as devised by the philosopher Lawrence SHAPIRO (2001). To conceive of the behaviour "tree swinging" (the particular ape locomotion mentioned above) as an adaptation is to claim, that "tree swinging" is to be identified with that physical structure that does the function we associate with "tree swinging". This implies that intuition about phylogenetic closeness in so far it is based on adaptations are based on physical structure *as well* as function (see also SOBER 2000, p374).

More to the point of this paper, when GALLUP's chimpanzees behaved similarly to humans in front of a mirror (as compared to the mirror behaviour checklist), the behaviour warranted similarity in experiences, despite the lack of verbal behaviour. To the contrary, the rhesus monkeys investigated in the same study (GALLUP 1970), showed no mark-directed responses despite 14 days of mirror image confrontation. Therefore, one can conclude, that rhesus monkeys share less mirror related experiences with human children than chimpanzees do.

The phylogenetic stance also provides a theoretical basis for the scepticism expressed by researchers on self-recognition in pigeons (EPSTEIN/LANZA/SKINNER 1981, p695): "Although similar behaviour in primates has been attributed to a self-concept or other cognitive processes, the present example suggests an account in terms of environmental events". By stating that self-recognition abilities embrace a complex set of behaviours in certain contexts the performance of great apes will be radically different from that of pigeons. Thus, ideally the phylogenetic stance results in a gradation of different levels of mind-reading abilities.

The degree of divergence in self-recognition behaviour is then simply a matter of phylogenetic distance. Therefore, humans understand each other better than humans understand chimpanzees and they understand chimpanzees better than they do rhesus monkeys, although both chimpanzees and rhesus monkeys belong to the order of primates.

Conclusion

This paper has addressed how attribution of mind-reading abilities to nonhuman primates and human children differ in studies on mirror-based self-recognition. From the comparison some conclusions can be drawn.

First, in human children the employed criteria are basically anthropocentric, since the verbal response is accentuated. This renders the application of criteria of self-recognition irrelevant to nonhumans, and leaves the results vulnerable to nonmentalistic interpretations. However, the analysis is neutral with respect to possible mentalistic interpretations of the studies of children. The verbal response does not by itself warrant mentalistic interpretation. The use of the personal pronoun and self-naming in front of the mirror by children could be a learned socially reinforced response and therefore would add no additional weight to the evidence that children truly understand that the mirror image is a reflection of themselves.

Second, in nonhumans the criteria are *not* exclusively successful performance on the mark test, but a set of anthropomorphic behaviours, which the mark test validates.

Third, I have sketched a theoretical basis named "the phylogenetic stance" that can accommodate our intuitions about what the behavioural resemblance of nonhuman organisms to humans signals about their minds. The theory attempts to provide an explanation of why our intuition gives fine guidelines concerning attribution of minds in nonhuman organisms.

Author's address

Theresa Schilhab, Centre for the Philosophy of Nature and Science Studies, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100, Denmark.
Email: schilhab@nbi.dk

Notes

1 In the so-called Sally-Anne test, children pass if they claim that another persons belief is the cause of their search strategy (e.g., Povinelli/deBlois 1992).

2 This study did not include the mark test. The results are obtained exclusively from ethograms of the animals in the presence and absence of mirrors.

3 In this research, self-recognition seems to be interchangeable with self-consciousness. It is shown indirectly by certain behavioural categories: "Self-consciousness is used here to denote those behaviours showing embarrassment or some form of self-admiration", (Amsterdam 1972, p303).

4 Gallup himself takes exception to comparison of his method and that of Amsterdam (Gallup 1994; Povinelli/Gallup 1993) however his objections differ from those I present.

5 Frans de Waal develops a somewhat similar position, which he names "evolutionary parsimony". This principle should come into play "especially when both humans and apes exhibit traits not seen in monkeys, and two explanations are proposed where one may do" (de Waal 1996, p65).

References

Amsterdam, B. (1972) Mirror self-image reactions before age two. Developmental Psychobiology 5(4):297–305.

Anderson, J. R. (1984) The Development of Self-Recognition: A Review. Developmental Psychobiology 17(1):35–49.

Bertenthal, B. I./Fischer, K. W. (1978) Development of Self-Recognition in the Infant. Developmental Psychology 14(1):44–50.

Bigelow, A. E. (1981) The correspondence between self-and image movement as a cue to self-recognition for young children. The Journal of Genetic Psychology 139:11–26.

Byrne, R. W. (1995) The thinking ape. Oxford University Press: Oxford.

Cheney, D./Seyfarth, R. (1990) Attending to behaviour versus attending to knowledge: Examining monkeys' attribution of mental states. Animal Behaviour 40:742–753.

de Veer, M. W./van den Bos, R. (1999) A critical review of methodology and interpretation of mirror self-recognition research in nonhuman primates. Animal Behaviour 58(459–468).

de Waal, F. (1996) Good natured. Harvard University Press: Cambridge.

Dennett, D. C. (1991) Consciousness Explained. Penguin Books: London.

Dennett, D. C. (1996) Kinds of Minds. Weidenfeld & Nicholson: London.

Eddy, T. J./Gallup Jr., G. G./Povinelli, D. J. (1993) Attribution of cognitive states to animals: Anthropomorphism in comparative perspective. Journal of Social issues 49:87–101.

Eddy, T. J./Gallup Jr., G. G./Povinelli, D. J. (1996) Age Differences in the Ability of Chimpanzees to Distinguish Mirror-Images of Self From Video Images of Others. Journal of Comparative Psychology 110(1):38–44.

Epstein, R./Lanza, P. L./Skinner, B. F. (1981) "Self-awareness" in the Pigeon. Science 212:695–696.

Fisher, J. A. (1991) Disambiguating Anthropomorphism: An Interdisciplinary Review. In: Bateson, P. P. G./Klopfer, P. H. (eds) Perspectives in Ethology., Vol. 9. Plenum Press: New York, pp 49–85.

Gallup Jr., G. G. (1970) Chimpanzees: Self-Recognition. Science 167:86–87.

Gallup Jr., G. G. (1977) Self-Recognition in Primates. American Psychologist May:329–338.

Gallup Jr., G. G. (1994) Self-recognition: Research strategies and experimental design. In: Parker, S. T./Mitchell, R. W./Boccia, M. L. (eds) Self-awareness in Animals and Humans. Cambridge University Press: Cambridge, pp 35–50.

Gallup Jr., G. G./Povinelli, D. J. (1993) Mirror, mirror on the wall which is the most heuristic theory of them all? A response to Mitchell. New Ideas in Psychology 11:327–335.

Goldman, A. I. (2000) Can Science Know When You're Conscious? Journal of Consciousness Studies 7(5):3–22.

Guthrie, S. E. (1997) Anthropomorphism: A Definition and a Theory. In: Mitchell, R. W./Thompson, N. S./Miles, H. L. (eds) Anthropomorphism, Anecdotes and Animals. State University of New York Press: Albany, pp 50–58.

Hare, B./Call, J./Agnetta, B./Tomasello, M. (2000) Chimpanzees know what conspecifics do and do not see. Animal Behaviour 59:771–785.

Herzog, H. A./Galvin, S. (1997) Common Sense and the Mental Lives of Animals: An Empirical Approach. In: Mitchell, R. W./Thompson, N. S./Miles, H. L. (eds) Anthropomorphism, Anecdotes and Animals. State University of New York Press: Albany, pp 237–253.

Heyes, C. M. (1993) Anecdotes, training, trapping and triangulating: Do animals attribute mental states? Animal Behaviour 46:177–188.

Heyes, C. M. (1994) Reflections on self-recognition in primates. Animal Behaviour 47:909–919.

Heyes, C. M. (1995) Self-recognition in primates: further reflections create a hall of mirrors. Animal Behaviour 50:1533–1542.

Heyes, C. M. (1998) Theory of mind in nonhuman primates. Behavioural and Brain Sciences 21:101–148.

Hill, S. D./Tomlin, C. (1981) Self-Recognition in Retarded Children. Child Development 52:145–150.

Hyatt, C. W./Hopkins, W. D. (1994) Self-awareness in bonobos and chimpanzees: A comparative perspective. In: Parker, S. T./Mitchell, R. W./Boccia, M. L. (eds) Self-awareness in Animals and Humans. Cambridge University Press: Cambridge, pp 248–253.

Johnson, D. B. (1983) Self-Recognition in Infants. Infant Behaviour and Development 6:211–222.

Kennedy, J. S. (1992) The new anthropomorhism. Cambridge University Press: Cambridge.

Lethmate, J./Dücker, G. (1973) Untersuchungen zum selbsterkennen im Spiegel bei Orang-utans und einigen anderen Affenarten. Zeitschrift für Tierpsychologie 33:248–269.

Lin, A.C./Bard, K. A./Anderson, J. R. (1992) Development of Self-Recognition in Chimpanzees (Pan troglodytes). Journal of Comparative Psychology 106:120–127.

Mans, L./Cicchetti, D./Sroufe, L. A. (1978) Mirror Reactions of Down's Syndrome Infants and Toddlers: Cognitive Underpinnings of Self-Recognition. Child Development 49:1247–1250.

Mitchell, R. W. (1991) Deception and Hiding in Captive Lowland Gorillas (Gorilla gorilla gorilla). Primates 32(4):523–527.

Mitchell, R. W. (1993) Mental models of mirror-self-recognition: Two theories. New Ideas in Psychology 11(3):295–325.

Mitchell, R. W. (1995) Self-recognition, methodology and explanation: A comment on Heyes (1994). Animal Behaviour 51:467–469.

Mitchell, R. W. (1997) Anthropomorphic Anecdotalism As Method. In: Mitchell, R. W./Thompson, N. S./Miles, H. L. (eds) Anthropomorphism, Anecdotes and Animals. State University of New York Press, Albany, pp 151–169.

Parker, S. T./Mitchell, R. W. (1994) Evolving self-awareness. In: Parker, S. T./Mitchell, R. W./Boccia, M. L. (eds) Self-awareness in Animals and Humans. Cambridge University Press: Cambridge, pp 413–428.

Patterson, F. G. P./Cohn, R. H. (1994) Self-recognition and self-awareness in lowland gorillas. In: Parker, S. T./Mitchell, R. W./Boccia, M. L. (eds) Self-awareness in Animals and Humans. Cambridge University Press: Cambridge, pp 273–290.

Povinelli, D. J. (1994) Comparative studies of animal mental state attribution. A reply to Heyes. Animal Behaviour 48:239–241.

Povinelli, D. J./deBlois, S. (1992) Young Children's (Homo sapiens) Understanding of Knowledge Formation in Themselves and Others. Journal of Comparative Psychology 106(3):228–238.

Povinelli, D. J./Gallup Jr., G. G. (1993) Mirror, mirror on the wall which is the most heuristic theory of them all? A response to Mitchell. New Ideas in Psychology 11(3): 327–335.

Povinelli, D. J./Gallup Jr., G. G./Eddy, T. J./Bierschwale, D. T./Engstrom, M. C./Perilloux, H. K./Toxopeus, I. B. (1997) Chimpanzees recognize themselves in mirrors. Animal Behaviour 53:1083–1088.

Povinelli, D. J./Nelson, K. E./Boysen, S. T. (1990) Inferences of guessing and knowing by chimpanzees (Pan troglodytes). Journal of Comparative Psychology 104:203–210.

Povinelli, D. J./Nelson, K. E./Boysen, S. T. (1992) Comprehension of role reversal in chimpanzees; Evidence of empathy? Animal Behaviour 43:633–640.

Povinelli, D. J./Parks, K. A./Novak, M. A. (1991) Do Rhesus Monkeys (Macaca mulatta) Attribute Knowledge And Ignorance to Others? Journal of Comparative Psychology 105(4):318–325.

Povinelli, D. J./Rulf, A. B./Landau, K. R./Bierschwale, D. T. (1993) Self-Recognition in Chimpanzees (Pan troglodytes): Distribution, Ontogeny, and Patterns of Emergence. Journal of Comparative Psychology 107(4):347–372.

Premack, D./Woodruff, G. (1978) Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences 4:515–526.

Priel, B./de Schonen, S. (1986) Self-Recognition: A Study of a Population Without Mirrors. Journal of Experimental Child Psychology 41:237–250.

Robinson, J. A./Conell, S./McKenzie, B.E./Day, R. H. (1990) Do infants use their own mirror-images to locate objects reflected in a mirror? Child Development 61:1558–1568.

Rollin, B. E. (1998) The unheeded cry. Iowa State University Press: Ames.

Shapiro, L. (2001) Mind the Adaptation. In: Walsh, D. M. (ed). Cambridge University Press: Cambridge.

Shillito, D. J./Gallup Jr., G. G.,Beck, B. B. (1999) Factors affecting mirror behaviour in western lowland gorillas Gorilla gorilla. Animal Behaviour 57:999–1004.

Smith, P. K. (1996) Language and the evolution of mind-reading. In: Carruthers, P./Smith, P. K. (eds) Theories of theories of mind. Cambridge University Press: Cambridge, pp 344–354.

Sober, E. (1998) Morgan's canon. In: Allen C Cummins, D. D. (eds) The Evolution of Mind. Owford University Press: New York, pp 224–242.

Sober, E. (2000) Evolution and the problem of other minds. The Journal of Philosophy 97(7):365–386.

Suarez, S. S./Gallup Jr., G. G. (1981) Self-recognition in Chimpanzees and Orangutans, but not Gorillas. Journal of Human Evolution 10:175–188.

Swartz, K. B./Evans, S. (1991) Not all chimpanzees show self-recognition. Primates 32:483–496.

Swartz, K. B./Evans, S. (1994) Social and cognitive factors in chimpanzee and gorilla mirror behavior and self-recognition. In: Parker, S. T./Mitchell, R. W./Boccia, M. L. (eds) Self-awareness in Animals and Humans. Cambridge University Press: Cambridge, pp 289–307.

Tomasello, M./Call, J. (1997) Primate Cognition. Oxford University Press: Oxford.

Tomasello, M./Call, J./Hare, B. (1998) Five primate species follow the visual gaze of conspecifics. Animal Behaviour 55:1063–1069.

Ujhelyi, M./Merker, B./Buk, P./Geissmann, T. (2000) Observations on the behavior of gibbons (Hylobates leucogenys, H. gabriellae, and H. lar) in the presence of mirrors. Journal of Comparative Psychology 114(3):253–262.

Westergaard, G. C./Hyatt, C. W. (1994) The responses of bonobos (Pan paniscus) to their mirror images: Evidence of self-recognition. Human Evolution 9(4):273–279.

Whiten, A. (1996) When does smart behaviour-reading become mind-reading? In: Carruthers, P./Smith, P. K. (eds) Theories of theories of mind. Cambridge University Press: Cambridge, pp 277–292.

Woodruff, G./Premack, D. (1979) Intentional communication in the chimpanzee: The development of deception. Cognition 7:333–362.

Scott Atran

A Cheater–Detection Module?

Dubious Interpretations of the Wason Selection Task and Logic

Abstract

People usually fail the W*ASON selection task, choosing P and Q cases, when attempting to validate descriptive rules having the form "If P, then Q". Yet they solve it, selecting P and not-Q cases, when validating deontic rules of the form "If P, then must Q". The field of evolutionary psychology has overwhelmingly interpreted deontic versions of the selection task in terms of a naturally-selected, domain-specific social-contract or cheating algorithm. This work has done much to promote evolutionary psychology as an alternative to a mindblind sociobiology that ignores the computational structure of cognitive mechanisms in producing people's behaviors. Nevertheless, evolution-minded researchers outside cognitive psychology know little of the ample literature challenging this interpretation and uncritically cite the 'cheater–detection module' as a key insight into human cognition. Although a priori arguments for a specially evolved cheater–detection module are plausible, the selection task provides no direct evidence for such a module.*

Key words

Adaptation, cheater–detection, logic, modularity, relevance.

Introduction: Selecting Out Cheaters

Humans and other animals practice multiple forms of cooperation, or 'reciprocal altruism', among genetically unrelated individuals of the same species (TRIVERS 1971). One answer to why nonkin cooperate is derived from the concepts of 'trade' and 'tradeoff' in economics and game theory (AXELROD 1984, FRANK 1988). If the benefit of being assisted outweighs the cost of giving assistance, then individuals who practice mutual aid can outproduce others who don't. For example, vampire bats live in large, stable social groups and recognize one another by voice. To survive, vampire bats foray each night in search of a blood meal. Chances of success are highly variable and a bat will die if unfed for sixty hours. To reduce this variance and prevent starvation, bats with blood-filled stomachs will regurgitate some of this valuable and hard-to-get resource to other hungry bats. The best predictor of whether or not a bat will share with a needy nonrelative is whether or not the nonrelative has previously shared food (WILKINSON 1984).

Vampire bats may be able to recognize cheaters during grooming, when they can best perceive whose stomachs are most distended with food and yet are not sharing. It is unclear, however, whether a bat that fails to regurgitate is recognized as a 'cheater' only by individuals the bat has denied, or acquires a 'reputation' as a defector when other 'cooperators' observe the bat's denial to those in need. It is also unclear whether cooperation is a *quid pro quo* or "from each according to its ability", whether cheaters recognize the consequences of their 'defection', and whether cheaters or would-be cheaters learn from the 'punishments' meted out.

A group of individuals that always cooperated would not likely survive an invasion of cheaters, unless the cooperators could identify and exclude the cheaters. Otherwise, the cooperators would be in effect subsidizing cheaters at significant cost to themselves and thus driving themselves to extinction. A group of individuals that always cheated would not likely survive an invasion of cooperators that could reliably discriminate cooperators from cheaters, because cheaters would always be denied the resources that cooperators obtained from one another. Detecting cheaters usually carries some cost in time or energy allotted to marking, monitoring and punishing or defending against them. As a population tended to full cooperation, the (selection) pressures to pay the cost of detecting cheaters would lessen, but opportunities for cheaters to invade undetected would thereby increase. As cheaters began succeeding, cooperators would again have to evolve cheater–detectors or die out. These antagonistic selection pressures make it unlikely than any population involving interactions between nonkin would consist wholly of cooperators or cheaters.

In any event, cooperation could not work without a cognitive system that directs an organism's attention to information that could enable it to decide whether or not it was being cheated (COSMIDES/TOOBY 1992). In humans, the opportunities for cooperation and cheating can range over the exchange of virtually any material or intellectual commodity, including money, ideas, time and baseball cards. Consequently, a human cognitive system for detecting cheaters cannot be restricted to perceptions of food intake and outake, but requires a more abstract computational algorithm for representing the 'benefits' and 'costs' that accrue to the exchange of any 'good'. One such candidate algorithm for human cooperation takes the form: "If Party A takes some benefit from Party B, then Party A must pay back the cost to Party B". The corresponding algorithm for cheater–detection would be: "A benefitted from, but failed to pay the cost to, B".

A Modular Interpretation of the Selection Task

In 1966, Peter WASON introduced a selection task to study logical reasoning about conditionals that has become the most widely used instrument for the experimental exploration of the psychology of human reasoning. The task presents subjects with a conditional rule in either a descriptive form ("If an item has property *P*, then it has property *Q*") or a deontic form ("If an item has property *P*, then it should have property *Q*"). A paradigm example of the descriptive form is: "if there is a vowel on one side of the card, then there is an odd number on the other". Four double-sided cards are placed on a table so that the subject can see only one side of each card, such as 'A', 'B', '1' and '2'. The experimenter instructs the subject to indicate which of the four cards must be turned over to find out whether the rule is true or false. If the conditional rule is interpreted in terms of formal logic, then only the *P* card ('A') and the *not-Q* card ('2') need to be turned over. Nevertheless, WASON and numerous subsequent researchers found that, for such "abstract" versions of the descriptive form of the selection task, most subjects turn over the *P* card ('A') and the Q card ('1'). This seems to suggest that ordinary human inference does not generally obey the laws of propositional logic.

In deontological contexts that convey social obligation, however, studies show that subjects often perform 'logically'. For example, subjects are asked to imagine a policeman checking bars to see if bargoers obey the deontic conditional: "If you drink alcohol, then you must be over twenty-one years of age". Suppose the policeman reads the cards "Beer", 'Coke', '22 years', '16 years'. In these situations, most subjects choose the *P* card ("Beer") and the *not-Q* card ('16 years') (GRIGGS/COX 1982). To obtain a genuine deontic interpretation, the deontic rule has to make sense. A 'must' formulation is only suggestive, but neither necessary or sufficient.

In another deontic task, subjects are asked to imagine a worker who signs on with a firm under assurance that: "If an employee works on the weekend, then that person gets a day off". The subject is then asked to verify if the contract is upheld by selecting from the following cards: "Worked Weekend" (*P*), "Worked Only During Week" (*not-P*), "Got Day off" (Q), "Did Not Get Day Off" (*not-Q*). Most subjects 'correctly' pick the P and *not-Q* cards. But when asked to take the perspective of the employer, rather than the worker, most subjects pick the *not-P* and Q cards (COSMIDES 1989, GIGERENZER/HUG 1992; MANTKELOW/OVER 1991; POLITZER/NGUYEN-XUAN 1992; FIDDICK/COSMIDES/TOOBY 2000). It is counterintuitive to conclude that the only logical answers are those given from the worker's perspective and not the employer's perspective. Intuitively, both perspectives seem equally rational and reasonable.

These results have led some to surmise that deontic versions of the selection task differ from abstract descriptive versions by giving practical content to 'pragmatic schema', and that humans are naturally more attuned to reasoning with pragmatic schema than with abstract logic (CHENG/HOLYOAK/NISBETT/OLIVER 1986). According to Leda COSMIDES, however, 'content-effects' and 'pragmatic schema' are nebulous concepts that can arguably be reduced to a simple evolutionary imperative, "find the cheater". On her account, the deontic tasks are often naturally interpreted as social contracts. Verifying the violation of a social contract requires picking the 'benefit taken' and 'cost not paid' cards, whatever the logical form of the contract. For GIGERENZER/HUG (1992), even the notion of a social-contract algorithm is too broad; rather, their studies in perspective-shift suggest an evolved, domain-specific adaptation for a 'module' dedicated to discovery of cheating.

An Evolution-Sensitive Alternative

Perhaps the most forceful challenge to the selection task as evidence of a cheater–detection module comes from studies by SPERBER/CARA/GIROTTO (1995). Their proposal is that performance on *all* versions of the selection task—descriptive as well as deontic—can be readily explained by universal, cross-domain forms of reasoning that operate in conjunction with context-specific aspects of the task. These cross-domain processes are the same as those routinely used by all humans (excepting serious pathology) in ordinary linguistic communication (SPERBER/WILSON 1986). The general idea is that individuals attempt to ascertain the relevance of any new information within a context of pre-existing knowledge and expectancies (EVANS 1989).

SPERBER et al. experimentally manipulated the logical form, propositional content and pragmatic context of the selection task to test effects on judgments of relevance and card choice. If the new information, together with the background context, leads to new beliefs, or to rejection of old beliefs, then the information is relevant to that context. Subjects are intuitively confident in their spontaneous judgments of relevance and choose cards in conformity with those judgments.

Relevance is a matter of degree and involves cognitive tradeoffs. The greater the cognitive effects result form processing the information (e.g., the more the number of new beliefs or the wider the range of potential inferences), the greater the information's relevance to the individual; but the greater the cognitive effort needed to produce the effects (e.g., the more time it takes to process the information), the lesser its relevance on that occasion. For example, a student might be informed that: (1) classes start in early September, (2) classes start the morning after Labor Day, (3) classes start no more than thirty-six hours following the start of the last national holiday in summer. In most contexts, (2) is more relevant than one, because (2) implies (1) but not vice versa and so (2) has more associated inferences than (1). Statement (2) is also more relevant than (3) because it takes less time to process and understand (2), although (2) and (3) are inferentially equivalent. The resultant cognitive balance usually allows the communicator to formulate, and the auditor to comprehend whatever information the communicator's formulation is intended to convey, in a relatively rapid, economical and efficient way.

SPERBER et al. argue that subjects in the selection task behave no differently than people in ordinary communicative settings. Subjects attend to the most relevant information that is being made available to them by the experimenter, and attempt to interpret it within a context of assumptions that will maximize this relevance. When the experimenter presents subjects with the abstract version of the selection task, and a conditional of the form *If P then Q*, the first thing subjects try to do is simply find out whether or not there are relevant instances of *P* and *Q*. If there weren't, then the fact that the experimenter mentioned them at all would be an apparent violation of the tacit assumption that underlies all human communication: convey information in a relevant way. From a purely logical standpoint, subjects appear to be initially interpreting the experiment not as a problem of falsifying the universally quantified statement, "For all x ($Px \rightarrow Qx$)", but as verifying the existentially quantified statement, "There exists x (Px & Qx)". Moreover, if the subjects interpret the rule as applying only to the four cards in front of them, rather than to the task or cards in general, then subjects interpreting the conditional as an existentially quantified statement can falsify the rule just by turning over the *P* and *Q* cards in order to discomfirm that there is a *P* & *Q* card.

When the consequent is negated, however, the majority of subjects perform 'logically'. Take the statement: "If there is a vowel on one side of the card, then there is not an even number on the other side". Subjects pick the *P* card ('Vowel') and the *not-Q* card ('Even Number'). In general, when subjects are presented abstract versions of the selection task in the form *If P then Q*, "most subjects choose the matching cards *P* and *Q* and thus apparently solve the [logical] problem" (EVANS 1989, p57). SPERBER et al. argue that in such cases, subjects reconstruct the assumption that the rule denies, namely, that there are cases of *P* & *Q*: "This interpretation of the rule as a denial causes them to make the correct selection" (SPERBER et al. 1995, p79).

What is different about deontic contexts is that the pragmatic context shifts the emphasis from rule verification to rule violation. This creates a situation much like the negative-consequent version of the descriptive task, except that in deontic and thematic tasks, the content and the context—rather than any explicit negation in the consequent—indicate what assumption is being denied by the rule. For example, the fact that a person is over twenty-one years of age and drinks alcohol is not very informative to most people in our society. But the context (e.g., that there is a police officer checking) raises the possibility that there might be underage drinkers. In this context, underage drinking would have cognitive effects; therefore, it would be more relevant to interpret the information as *forbidding* underage drinking: not-(*Drinks Alcohol* & *not-Over 21*). In general, the logic form of subjects' interpretation of deontic versions of the selection tasks is: not-[There exists x (*P*x & *not-Q*x)].

Notice that the same logical interpretation could arise in contexts that do not involve social contracts or cheating detection. For example, take the statement: "If a person wins a professional boxing match, then that person must be sober". The prediction is that subjects would pick the *P* card ('Wins Match') and the *not-*Q card ('Drunk') because information concerning a winning but drunk professional boxer more likely has cognitive effects than information concerning a successful sober boxer (cf. ALMOR/SLOMAN 1996).

FIDDICK/COSMIDES/TOOBY (2000) suggest that certain conditionals used in selection-task format might be interpreted in terms of a fitness-preserving hazardous-management module rather than a cheater–detection or social contract module. The general algorithm for a fitness-preserving conditional is: "If in a hazardous situation that is costly to fitness, then take the benefit of precaution". FIDDICK et al. predict (and experimentally confirm predictions) that the majority of subjects will pick the *P* ('Hazardous Situation') and *not-Q* ('No Precaution') cards.

FIDDICK et al. deny that relevance theory can reliably predict patterns of performance on deontic versions of the selection task, such as perspective shifts involving cheater–detection or fitness-preservation. For example, take the rule: "If you make poison darts, then you may use the rubber gloves". In the Privilege condition, subjects are primed to take the perspective of an anthropologist checking to see if tribespeople are abusing the privilege of wearing gloves. In this condition, subjects tend to make the apparently illogical *not-P* and Q selection. In the Risk condition, subjects take the perspective of an anthropologist checking if tribespeople are unduly risking their lives. In this condition, subjects make the logical *P* and *not-Q* selection.

Fiddick et al. argue that relevance theory must hold that either the rule is pragmatically awkward, or it is not. But if the rule is pragmatically awkward, then subjects should pick the illogical *not-P* and *Q* for both conditions. And if the rule is not pragmatically awkward, then subjects should pick the logical *P* and *not-*Q for both. In brief, relevance theory is seemingly faced with two contradictory outcomes: "because of its reliance upon logical formulae, relevance theory is placed in the unenviable position of having to invoke contradictory principles". As an alternative to relevance theory, FIDDICK et al. suggest that two different domain-specific schema are operating. In the Privilege condition, the anthropologist is using a *cheater–detection device* to see if tribespeople are abusing a privilege. In the Risk version, the anthropologist is using a *hazard-management device* to see if tribespeople are unduly risking their lives.

FIDDICK et al. fail to consider that, from the standpoint of relevance theory, both conditions may be taken as *implying* reciprocity:

A. Explicitly: For all x [*P(costly risk*) x → should take Q *(precautionary benefit)* x]

B. This implies: For all x [Q*(precautionary benefit)* x → should take *P(costly risk)* x]

In the Privilege condition, subjects look for a Violation of B:

Not-{there exists [Q*(Benefit)* x & *not-P(Cost)* x]}

In the Risk condition, subjects look for a Violation of A:

Not-{there exists [*P(Cost)* x & *not-Q(Benefit)* x]}

In another experiment, FIDDICK et al. present a bartering situation where people cannot *express* conditionals of the form "if P then Q". A farmer from one tribe nevertheless understands gestures from people in the other tribe indicating "We want potatoes" and he gestures back in ways that the people form the other tribe understand as "I want corn". The argument against relevance theory is that "I want P/ we want Q" does not work explicitly on logical form. Because relevance theory supposedly works only on logical form, relevance theory can't explain the results. But as FIDDICK et al. note themselves, surface form may differ from underlying logical form. Relevance theory makes no claim that surface form and logical form are isomorphic. Nor is such an isomorphism to be expected from the study of language. In syntactic theory, for instance, the command "Eat!", when understood as an imperative in English, entails two arguments that are absent in the statement's surface form but obligatorily present in its underlying logical form, namely, a subject and an object.

From the standpoint of relevance theory, the surface form of the barter, A, may be interpreted as implying the logical forms, B and C:

A. Explicit form: P (farmer give potatoes)/Q (tribe give corn)

B. Implicit form: For all x [*P(farmer's potatoes)* x → ought give Q *(tribe's corn)* x]

C. This further implies: For all x [*Q(tribe's corn)* x → ought give *P(farmer's potatoes)* x]

So, when asked to indicate when the farmer is cheating, subjects interpret this to mean that there is a Violation of C:

Not-{there exists [*Q(tribe gives)* x & *not-P(farmer gives)* x]}

When asked to indicate when other tribespeople are cheating there is a Violation of B:

Not-{there exists [*P(farmer gives)* x & *not-Q(tribe gives)* x]}

Leda COSMIDES and her colleagues are right to point out that interpretation of reasoning tasks involve various content-dependent algorithms that may have evolutionary import. As another example, consider the statement: "if a heavy object is projected up into the air up, then that object must come down". Arguably, subjects could invoke yet another domain-specific algorithm, a 'folkphysics module' (cf. SPELKE/PHILLIPS/WOODWARD 1995). It is unclear what module, if any, would cover selections for the following statement: "If a person wins a multimillion dollar lottery, then that person must be happy/sad". No cheating or hazard appears to be involved, although interpretation may involve evolution-linked affect schema.

A relevance-based interpretation of the WASON task does not require a strictly non-modular interpretation, such as interpretations that draw on mental-model theory or semantic networking. Neither does a relevance-based interpretation preclude the possibility of a cheater–detection module. Rather, the selection task evidence is simply better explained in terms of relevance than in terms of cheater detection. Relevance-guided comprehension may itself be evolutionarily specialized as a 'meta-representation module' (SPERBER 1994). But theoretical claims and empirical evidence for relevance-guided modularity are independent of claims about cheater–detection.

Logic: Another Dubious Cheater Catcher

In a recent essay, SPERBER (in press) turns the argument by COSMIDES and colleagues almost on its head. He proposes that some kind of argumentative logic is part of a naturally-selected adaptation that arose during an evolutionary arms race between communicators attempting to persuade (and deceive if need be) and audiences trying to evaluate messages as truthful or not. In the communication arms race, sophistry contributed as much as honesty to this adaptation for persuasion-and-coherence-checking. It is a nice a story, but so far *just* a story.

Historically, the formalization of deductive logic began with Aristotle, who sought an effective form of argumentation as a rhetorical device to parry the sophists and promote a better, scientific analysis of evidential relationships. The ensuing formalization was more than just a standardization of folk reasoning. It subtly but profoundly changed the use of everyday terms, such as 'if', 'and', 'or', and of words indicating inferential relationships, such as 'therefore', 'since', 'nonetheless'. In doing so it created counterintuitive truth tables for material implication. Thus, if the antecedent is false, the whole statement is true. Ordinarily, folk would probably conclude that the statement is indeterminate. True, if you put gold in aqua regia it either dissolves or doesn't (here folk intuition and formal logic agree). But if you don't put gold in aqua regia to begin with, then an ordinary intuition is that there's no argument possible. SPERBER points out (personal communication, 2001) that at least some cases of the antecedent being false and the whole statement being true seem to be amenable to ordinary intuition: for example, so-called Dutch conditionals ("If you're right, then I'm the Pope—or Dutch"). Nevertheless, logically, *all* such statements must be equally coherent—and they aren't.

Moreover, formal logic makes no distinction between ecologically valid and invalid reasoning. For example, the proposition "All ravens are black", has both an intuitive and formal relationship to evidence that bears on its truth, such as verifying that anything identified as a raven is indeed black. But the formally equivalent proposition, "All non-black things are non-ravens", has little pertinence for any real-world process of verification or evidentiary evaluation (if would be absurd and infinitely time consuming to actually go out and see if all things that are not black are indeed not ravens).

Finally, if logic is an adaptation for persuasion and cheater-detection, it's a pretty weak one. The advertising industry employs other types of reasoning and reflection on evidentiary relationships that side steps or easily overrides logic (e.g., you hair has protein, our shampoo has protein, therefore our shampoo is good for your hair). And, as Doug MEDIN notes (personal communication, 2001), perfume readily carries the day over argumentative displays of coherence, if circumstances are right, no matter how coherent the argument.

Conclusion: No Multiplication of Modules Beyond Evolutionary Necessity

If selection-task performance varies according to contexts that differentially draw upon existential verification, cheater detection, fitness preservation, folkphysics, affect programs or general encylopedic knowledge, then it is difficult to see how the selection task disambiguates or privileges any *particular* domain-specific mode of processing. The advantage of relevance theory over other accounts of the selection task is that relevance theory applies predictably to all versions of the selection task, without denying or privileging the effects of additional domain-specific competencies in any given context: "comprehension mechanisms and domain-specific mechanisms jointly contribute to subjects' performance, but… their effects are, as things stand, confounded" (Sperber et al. 1995, p88).

The selection task mobilizes various information processing devices at the interface between our relevance-guided system of comprehension and any number of other task-specific modules. As our ability to identify modular systems becomes more secure, evidence of activity at the interface will undoubtedly add detail and refinement to our understanding of how modular mechanisms function and interact. But as with the interaction between the language faculty and various other cognitive systems, the interface is not the place to begin to understand modularity (Chomsky 2000).

Finally, the counter proposal that logic itself emerged as a module for coping with an evolutionary arms race between persuasion and cheater-detection also lacks independent support (say, of the kind that could be garnered for an emotion-based alternative). In general, I think that mental modules should be invoked as explanations only when independently converging arguments for modularity can be sustained. This could include evidence for: cross-cultural universality, early and rapid ontogenetic acquisition, ease and rapidity of cultural transmission, hyperactivity and difficulty in inhibiting operation even in the face of contrary instruction, selective cerebral localization or impairment, evolutionary analogies, functional phylogenetic homologies, imperfect but ecologically performative design (Atran 1998, 2001). These are not necessary and sufficient causal criteria for establishing modularity, only an imperfect but indicative community of evidentiary symptoms. This somewhat parallels the cautious strategy that George Williams (1966) introduced with respect to the concept of adaptation in evolution, and which has led to substantial progress in understanding evolutionary process and structure.

Author's address

Scott Atran, CNRS-Institut Jean Nicod, Paris, 9 rampe de l'observatoire, 66660 Port Vendres, France, and ISR-The University of Michigan, Ann Arbor, USA.
Email: satran@umich.edu

Acknowledgments

Thanks to Doug Medin, Dan Sperber, Larry Fiddick, Dick Nisbett, David Hull and the anonymous reviewers for earlier comments and suggestions.

References

Almor, A./Sloman, S. (1996) Is deontic reasoning special? Psychological Review 103:374–380.

Atran, S. (1998) Folkbiology and the anthropology of science: Cognitive universals and cultural particulars. Behavioral and Brain Sciences 21:547–609.

Atran, S. (2001) The case for modularity: Sin or salvation? Evolution and Cognition 7:46–55.

Axelrod, R. (1984) The evolution of cooperation. Basic Books: New York.

Cheng, P./Holyoak, K./Nisbett, R./Oliver, L. (1986) Pragmatic versus syntactic approaches to training deductive reasoning. Cognitive Psychology 18:23–328.

Chomsky, N. (2000) Minimalist inquiries: The framework. In: Martin, R./Michaels, D./Uriagereka, J. (eds) Step by step. MIT Press: Cambridge MA, pp 89–155.

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./Tooby, J. (1992) Cognitive adaptations for social exchange. In: Barkow, J./Cosmides, L./Tooby, J. (eds) The adapted mind: Evolutionary psychology and the generation of culture. Oxford University Press: New York, pp. 163–228.

Evans, J. (1989) Bias in human reasoning. Erlbaum: Hillsdale NJ.

Fiddick, L./Cosmides, L./Tooby, J. (2000) No interpretation without representation: The role of domain-specific representations and inferences in the Wason Selection Task. Cognition 75:1–79.

Frank, R. (1988) Passions within reason. W.W. Norton: New York.

Gigerenzer, G./Hug, K. (1992) Reasoning about social contracts: Cheating and perspective change. Cognition 43:127–171.

Griggs, R./Cox, J. (1982) The elusive thematic-materials effect in the Wason Selection Task. British Journal of Psychology 73:407–420.

Mantkelow, K./Over, D. (1991) Social roles and utilities in reasoning with deontic conditionals. Cognition 39:85–105.

Politzer, G./Nguyen-Xuan, A. (1992) Reasoning about promises and warnings. Quarterly Journal of Experimental Psychology 44A:401–421.

Spelke, E./Phillips, A./Woodward, A. (1995) Infants' knowledge of object motion and human action. In: Sperber, D./ Premack, D./Premack, A. (eds) Causal cognition. Clarendon Press: Oxford, pp. 44–78.

Sperber, D. (1994) The modularity of thought and the epidemiology of representations. In: Hirschfeld, L./Gelman, S. (eds) Mapping the mind: Domain specificity in cognition and culture. Cambridge University Press: New York, pp. 39–67.

Sperber, D. (in press) An evolutionary perspective on testimony and argumentation. Philosophical Topics (special issue in honor of Alvin Goldman).

Sperber, D./Cara, F./Girotto, V. (1995) Relevance theory explains the selection task. Cognition 57:31–95.

Sperber, D./Wilson, D. (1986) Relevance. Blackwell: Oxford.

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

Wason, P. (1966) Reasoning. In: Foss, B. (ed) New horizons in psychology. Penguin: Harmondsworth, pp. 135–151.

Wilkinson, G. (1984) Reciprocal food sharing in the vampire bat. Nature 308:181–184.

Williams, G. (1966) Adaptation and natural selection: A critique of some current evolutionary theought. Princeton University Press: Princeton.

Todd K. Shackelford and Gregory J. LeBlanc

Sperm Competition In Insects, Birds, and Humans

Insights from a Comparative Evolutionary Perspective

Abstract

An understanding of the adaptive problems, evolved psychologies, and manifest behaviors of one species can offer insights into the evolved psychologies of other species. Sperm competition provides an arena within which to assess the heuristic value of such a comparative evolutionary perspective. Sperm competition occurs when the sperm of two or more males simultaneously occupy a female's reproductive tract. We describe mechanisms of sperm competition in insects and in birds. We suggest that the adaptive problems and evolved solutions in these species provide insight into evolved human sexual behaviors and psychology. We review recent theoretical and empirical arguments for the existence of human sperm competition. Using a comparative evolutionary psychological approach, we discuss features of male psychology and female psychology that may have evolved as solutions to the adaptive problems presented by sperm competition. We conclude with a discussion of future directions for work on human sperm competition, highlighting the heuristic value of a comparative evolutionary psychological approach in this field.

Key words

Sperm competition, evolutionary psychology, cognitive mechanisms, comparative psychology.

The application of evolutionary thinking to animal behavior and psychology offers a number of benefits. A comparative evolutionary psychological approach can provide insight into differences and similarities between species by encouraging rigorous examination of the adaptive problems and evolved solutions to these problems (Daly/Wilson 1998). Different species sometimes share a similar suite of evolved mechanisms because they recurrently confronted similar classes of adaptive problems. In this sense, the study of non-human animal behavior and psychology can shed light on human behavior and psychology (Wilson 1997). Trivers' (1972) theory of parental investment and sexual selection, for example, has been successful in explaining sex differences in courtship behaviors in a number of species, including humans. Analogously, Sperm Competition Theory has generated a wealth of information about the mating systems of a number of species (Birkhead/Møller 1998; Parker 1970a, 1970b). In this review, we present a comparative approach to evolutionary psychology, demonstrating that an understanding of the adaptive problems, evolved psychologies, and manifest behaviors of one species can offer insight into the evolved psychologies of other species. The focus of this article is on cross-species similarities, although a comparative evolutionary perspective also can be valuable in identifying cross-species differences (Daly/Wilson 1995).

Along with the benefits of a comparative evolutionary perspective come a number of difficulties (see Daly/Wilson 1998). For example, there are no rules for identifying whether a set of conditions that thwart reproduction across different species represent (a) a similar adaptive problem confronted by these different species, or (b) different adaptive problems that share qualitative features. In addition, there are no rules for judging the similarity or dissimilarity of the behaviors displayed by different species. As a final example, apparently similar adaptive problems confronted by different species are not always solved by the same evolved mechanisms. For example, many species faced the adaptive problem of thermal regulation. Solutions to this adaptive problem vary across species, however, ranging from sweating in humans, to the adjustment of wings and feathers in birds.

Our goal in this article is not to resolve these and other difficulties associated with adopting a comparative evolutionary perspective (see DALY/WILSON 1998). Instead, our goal in this article is to demonstrate that, these difficulties notwithstanding, a comparative evolutionary perspective can offer some insight into the adaptations and manifest behaviors of different species that have recurrently confronted similar classes of adaptive problems. Of key interest in this article is the value of a comparative evolutionary perspective for gaining a better understanding of human behavior and psychology. For example, because humans share with some insects the adaptive problem of mate retention, BUSS (1988) looked for and identified several behavioral similarities across these species that appear designed to prevent a mate from copulating with a rival. It is our intention to demonstrate in this article that sperm competition generated similar adaptive problems for some species of insects, many species of birds, and for humans. We suggest that cross-species behavioral similarities often may be the output of psychological mechanisms designed to solve qualitatively similar adaptive problems.

Brief Overview of Sperm Competition

Sperm competition is the result of the simultaneous occupation of sperm from two or more males in a single female's reproductive tract (PARKER 1970a, 1970b). Although important contributions to Sperm Competition Theory come from researchers studying many different species, PARKER's initial review was based on observations of three species of flies. PARKER's discovery of sperm competition in insects was facilitated by the fact that insects can store sperm for long periods of time in specialized organs, coupled with the ease with which females can be observed in multiple matings. PARKER's interest in game theory led him to consider the potential for an evolutionary arms race. On one hand, there would have been selection pressure for males to remove from a female the sperm deposited by previous suitors. On the other hand, selection would have favored those males that were able to prevent removal of their own sperm. PARKER noted that the last male to inseminate a female typically would fertilize 80% of her eggs. Investing more time and energy into copulating with a single female could produce a fertilization rate of 100%. Observations of male behaviors led PARKER to conclude that males maximized their offspring production not by monopolizing a single female, but by mating with multiple females. These observations and theoretical considerations formed the basis of Sperm Competition Theory (PARKER 1970a; for theoretical extensions and refinements, see PARKER 1982, 1984, 1990a, 1990b, 1993, 1998; PARKER et al. 1996; PARKER/BEGON 1993; PARKER/SIMMONS/KIRK 1990).

Sperm competition serves as a good 'test case' for the heuristic value of a comparative evolutionary psychological perspective. There is a wealth of information about sperm competition in many non-human animals, notably in insects and in birds (see, e.g., BIRKHEAD/MØLLER 1992, 1998). In addition, there is a growing body of work suggesting that sperm competition was an important selective force in shaping modern human psychology and behavior (see, e.g., BAKER/BELLIS 1995; SHACKELFORD et al. in press). One approach for furthering our understanding of human sexuality is to apply a comparative evolutionary psychological perspective to the study of sperm competition. When we examine human sexuality using a comparative evolutionary psychological approach to sperm competition, we can better understand our similarities to and differences from other sexually reproducing species.

Sperm Competition and Sexual Selection

In addition to natural selection, Charles DARWIN (1871) proposed sexual selection. Sexual selection describes (a) the competition between males for the opportunity to mate with females (intrasexual selection), and (b) the selection of male mates by females (intersexual selection). SMITH (1998) suggested that a more precise way of thinking about sexual selection is that it is not a competition between males to gain sexual access to females, but rather a competition between the ejaculates of different males to fertilize a female's eggs. In this sense, sperm competition can be considered an integral component of sexual selection (MØLLER 1998).

Sperm competition and sexual selection likely played an important role in the evolution and design of genitalia and reproductive physiology (EBERHARD 1985; SHORT 1979). For example, sexual selection could have acted upon ejaculate size. Larger ejaculates often are more successful in sperm competition (BAKER/BELLIS 1995). Comparatively, the ejaculates of more monogamous primate species experience less opportunity for sperm competition. Males in these species have smaller testicles relative to body weight (and consequently produce smaller ejaculates) compared to species with higher rates of sperm competition (GOMENDIO/HARCOURT/ROLDÁN 1998; HARCOURT et al. 1981).

Penis size and shape also may have evolved in response to the pressures of sperm competition and sexual selection. A longer penis that deposited sperm nearer to the site of fertilization may be the result of such selection (EBERHARD 1985; SHORT 1975). The shape of the human penis may be designed to remove sperm previously inseminated by a rival male (BAKER/BELLIS 1995). Empirical work on other species provides comparative, corroborative support for this hypothesis. The penis of the damselfly, for example, is designed such that it removes up to 99% of the stored sperm in the female before the male deposits his own ejaculate (WAAGE 1979).

Sperm Competition in Insects and in Birds

The first examination of sperm competition was in insects (PARKER 1970a). Insects provide a model of sperm competition that is applicable to a number of species. Male insects, like the males of other species, benefit by avoiding sperm competition. This avoidance can be accomplished in several ways. One method is to avoid competition with sperm that have been stored in a female by previous copulation with another male. Another method is to avoid competition with the sperm of males that might copulate with the female in the future (SIMMONS/SIVA-JOTHY 1998).

Many insects have anatomical features that facilitate avoidance of sperm competition. The penis of the male tree cricket, for example, is structurally designed to deliver an ejaculate into the anterior portion of the female's spermatheca (the sperm storage organ). Sperm that were stored there previously are collected on the male's penis and are removed prior to insemination of the male's own ejaculate (ONO/SIVA-JOTHY/KATO 1989; see also GAGE 1992).

Another method of avoiding sperm competition may exist. This is killing or incapacitating rival sperm. Some evolutionary biologists suggest that this mechanism operates in humans (BAKER/BELLIS 1995). Different sperm morphs are thought to 'seek and destroy' the sperm of rivals within a female's reproductive tract. No conclusive evidence has been collected to support this idea, however (for additional criticisms and a review of related empirical and theoretical work, see GOMENDIO/HARCOURT/ROLDÁN 1998).

In addition to the evolution of mechanisms designed to avoid competition with sperm already stored inside a female, some male insects are equipped with mechanisms that are designed to prevent the opportunity for future sperm competition (THORNHILL/ALCOCK 1983). These 'mate guarding' mechanisms can be classified as proximate or remote. Proximate mate guarding involves a male staying within close physical distance of the female with whom he has recently copulated (TSUBAKI/SIVA-JOTHY/ONO 1994). Remote mate guarding involves tactics designed to aid the male in avoiding competition with sperm from rival males when he is not physically present with his partner to do this guarding himself. In insects, for example, a copulatory plug is sometimes formed after copulation (DRUMMOND 1984). This plug is the result of chemical reactions between the seminal fluid and oxygen in the air. The plug blocks the female's genital orifice and thereby prevents future insemination by other males. Another form of remote mate guarding involves the post-copulatory release of a substance in the seminal fluid that reduces the receptivity of the female to rival males (THORNHILL 1976).

Birds have been studied extensively within the context of sperm competition. There are similarities between the mechanisms of sperm competition documented in birds and the mechanisms documented in insects. For example, the males of many species of birds and insects recurrently had to solve problems of avoiding sperm competition. Sperm competition mechanisms in birds also have important implications for the study of human sperm competition. Not only have male birds and male humans faced many of the same adaptive problems with respect to avoiding sperm competition, but also the majority of bird species engage in social monogamy, the mating system characteristic of humans (BAKER/BELLIS 1995; BIRKHEAD/MØLLER 1992). Social monogamy is a mating system in which a male and a female form a long-term pair bond. Within this mating system, males benefit by gaining exclusive sexual access to the reproductive resources of a female, whereas females benefit by gaining exclusive access to the investment of the male in her and her offspring (BIRKHEAD/MØLLER 1992; TRIVERS 1972). Human males and the males of many bird species invest substantially in offspring, which places these males at risk of investing in offspring to whom they are genetically unrelated. These and other similarities suggest the possibility of discovering similar mating behaviors, motivated by similar evolved psychologies, in birds and in humans.

Paternity guarding is a term used to describe behaviors designed to avoid sperm competition, or to avoid the key costs of sperm competition—failure to transmit genes into the next generation and paternal investment in genetically unrelated offspring. Researchers have identified two broad categories of paternity guarding in birds. These are paternity guarding through mate guarding and paternity guarding though frequent copulation (BIRKHEAD/MØLLER 1992). It was once thought that males guard their mates as a result of the pair bond that they form (BIRKHEAD/MØLLER 1992). Sperm Competition Theory offers a clearer understanding of the underlying mechanisms and processes. As males spend greater amounts of time away from their partners, the incidence of female extra-pair copulation increases (ALATALO/GOTTLANDER/LUNDBERG 1987). Accordingly, male birds adjust the intensity with which they guard their partners—particularly their proximate mate guarding behaviors—so that this mate guarding is most intense when their partners are at peak fertility (MØLLER 1987). But there are costs to mate guarding. Male birds engaged in mate guarding expend time and energy that could be used to locate food or to acquire additional mates. For example, significant weight loss has been documented in male ducks who spend more time mate guarding and, consequently, less time feeding (ASHCROFT 1976).

Although many species of birds engage in mate guarding, not all do. Ecological constraints prevent some birds of prey from mate guarding. The male osprey, for example, locates and acquires for his partner the majority of the food she eats (BIRKHEAD/LESSELS 1988). These males cannot simultaneously and successfully collect food for their partners and guard their partners. Furthermore, some male birds trade food for copulations, increasing not only their paternity certainty, but also the health of their putative offspring (TASKER/MILLS 1981). For the males in this species, frequent copulation can serve as a successful paternity guard. This is because (a) the number of copulations is positively related to the number of offspring sired by the male, and (b) the last male to copulate with a female fertilizes a relatively high percentage of her eggs (BIRKHEAD/MØLLER 1992).

Like mate guarding, the psychology motivating copulatory frequency appears to be sensitive to female fertility. Male ospreys copulate as many as 10 times per day with their partner when she is at peak fertility—just prior to egg laying. Copulatory frequency approaches zero after the first egg is laid, when her fertility is at a minimum (BIRKHEAD/LESSELS 1988). These behavioral patterns suggest that these birds are sensitive to the potential costs of sperm competition, and that they behave so as to reduce the likelihood of incurring these costs.

The preceding sections introduced key physiological and behavioral mechanisms of sperm competition. We presented information that provides a basis for comparing the adaptive problems and evolved solutions in insects and in birds to those in humans. The similarities in the mating system and parenting system of some birds and humans generated similar adaptive problems for these species in the domain of sperm competition. The next sections review theoretical arguments and empirical evidence for the existence of human sperm competition.

Theoretical Arguments for Human Sperm Competition

SMITH (1984) presented theoretical arguments for the existence of sperm competition in humans. SMITH (1984) explored the possible benefits that females might reap through manipulating the outcome of such competition. Sperm competition in humans requires that a female copulate with more than one male within roughly a five-day period. This can happen in several contexts. Communal sex, including wife-swapping, orgies, and 'swinging', is one such context. Studies conducted over the past 30 years indicate that these behaviors are rare and probably did not represent a recurrent context over the evolutionary history of humans in which sperm competition could act as a selective force (see, e.g., ATHANASIOU 1973; SMITH 1984).

Rape is another context that can place the sperm of different men into competition. BAKER/BELLIS (1995), SMITH (1984), THORNHILL/THORNHILL (1992; see also THORNHILL/PALMER 2000), and others argue that rape of females by males was a recurrent feature of human evolutionary history. Despite cultural institutions that punish and discourage rape, over half a million rapes occur annually in the United States alone (GREEN 1980). There also is a strong association between rape and war, a key feature of our evolutionary past (BROWNMILLER 1975; BUSS 1999; THORNHILL/PALMER 2000). These reports suggest that rape could have provided a recurrent context for sperm competition to act as a selection pressure on human physiology and psychology.

Finally, female infidelity is a context for human sperm competition. Female extramarital sex has been documented in three-fourths of indexed societies (BROUDE/GREENE 1976; BUSS 2000; SHACKELFORD/BUSS 1997). According to SMITH (1984) and BAKER/BELLIS (1995), female infidelity may have been the primary context for sperm competition throughout human evolutionary history. If female infidelity was the primary context for sperm competition, females must have sometimes received substantial benefits for infidelity, because the potential costs to the female of detected infidelity are high. An unfaithful female risks the loss of her long-term partner's investment, and she risks injury and even death at the hands of her jealous partner (BUSS 2000; DALY/WILSON 1988; SHACKELFORD/BUSS/PETERS 2000). If the potential for female infidelity is a universal feature of human psychology, then the benefits of female infidelity must have outweighed these costs, on average. Such benefits include the contribution of superior genes to offspring, material resources, and protection for a woman and her offspring (GREILING/BUSS 2000; SMITH 1984). These are some of the same benefits gained by female birds that mate with extra-pair males (BIRKHEAD/MØLLER 1998). In addition to theoretical arguments for sperm competition being a recurrent feature of human evolutionary history, there is mounting corroborative empirical data.

Empirical Evidence for Human Sperm Competition

The size and structure of genitalia and other reproductive anatomy are key products of an evolutionary history of sperm competition. For example, relative to body size, male humans have testes that are larger than those of gorillas and orangutans, but smaller than those of chimpanzees (SMITH 1984). Humans are more monogamous than chimpanzees, but more promiscuous than gorillas and orangutans (GOMENDIO/HARCOURT/ROLDÁN 1998; SMITH 1984; WRANGHAM/PETERS 1996). Humans, chimpanzees, gorillas, and orangutans thus appear to have a reproductive physiology that is designed to 'expect' a certain level of sperm competition brought about by multiple mating by females.

Another form of evidence that suggests an evolutionary history of human sperm competition is that human sperm are produced in a variety of types or morphs. BAKER/BELLIS (1988, 1995) argue that human sperm is produced in different morphs, each performing a specific, evolved function. The authors classify human sperm into 'egg-getters', those that move most directly to fertilize the egg; 'blockers', those that take up larger areas in the mucus channels and prevent rival sperm from reaching the egg; and 'seek-and-destroyers', those that search for and upon detection kill rival sperm (see BAKER/BELLIS 1995, for a review of sperm polymorphism in other species). In addition to documenting different sperm morphs, BAKER/BELLIS (1989, 1993a, 1993b, 1995) conducted a series of experiments that tested hypotheses about human ejaculate adjustment in relation to the risk of female infidelity and subsequent sperm competition.

A key hypothesis derived from Sperm Competition Theory is that males will adjust the number of sperm they inseminate as a function of the risk that their sperm will encounter competition from the sperm of other males. This hypothesis has been tested and confirmed in numerous studies of various insects (e.g., GAGE 1991; GAGE/BAKER 1992; GAGE/BARNARD 1996; for a review of this work, see SIMMONS/SIVA-JOTHY 1998). Extending this work to humans, BAKER/BELLIS (1989, 1993a, 1995) documented a negative relationship between the proportion of time a couple has spent together since their last copulation and the number of sperm ejaculated at the couple's next copulation. As the proportion of time a couple spends together since their last copulation decreases, there is a predictable increase in the probability that the man's partner has been inseminated by another male (BAKER/BELLIS 1995). Additional regression analyses documented that the proportion of time a couple spent together since their last copulation is a significant predictor of sperm number ejaculated at the couple's next copulation, but not at the male's next masturbation (BAKER/BELLIS 1989, 1995). This class of observations inspired evolutionary psychologists to investigate the psychological mechanisms that might have been designed as solutions to the adaptive problems presented by sperm competition.

The Evolved Psychology of Sperm Competition

A comparative evolutionary psychological perspective predicts that species that recurrently faced similar adaptive problems may have evolved similar psychological mechanisms to solve these problems. The sperm competition that results from female infidelity presents a similar class of adaptive problems for individuals across many species. The fact that humans and birds share similar mating systems suggests additional parallels between the evolved psychologies of humans and birds. Male humans and the males of many bird species are faced with the problems of preventing, correcting, and anticipating their partner's infidelity (SHACKELFORD/LEBLANC 2001). Failure to solve these problems would have had devastating consequences—including failure to pass genes into the next generation and investing finite resources in genetically unrelated offspring. Females are faced with the problems of maximizing the benefits of infidelity relative to the costs of infidelity, and manipulating ejaculates in the presence of sperm competition. Evolutionary psychologists have recently begun to document that these adaptive problems may have designed psychological mechanisms that function to solve them (BAKER/BELLIS 1993a, 1995; SHACKELFORD et al. in press; THORNHILL/GANGESTAD/COMER 1995).

Men who prevented their partner's infidelity would have benefited reproductively over the course of human evolution. Human mate guarding behaviors may be the output of psychological mechanisms designed to prevent sperm competition (BUSS 1988; BUSS/SHACKELFORD 1997; FLINN 1988; SHACKELFORD/LEBLANC 2001). These mate guarding behaviors range from vigilance over a female partner's whereabouts to female-directed violence for a suspected infidelity. Similar mate guarding behaviors have been observed in many species of birds (BIRKHEAD/MØLLER 1992). As predicted by a comparative evolutionary perspective, there are similarities between many classes of mate guarding behaviors in humans and mate guarding behaviors in birds. For example, male humans and male birds adjust the intensity of their mate guarding behaviors according to the reproductive value and fertility of their partners (BUSS/SHACKELFORD 1997; FLINN 1988; MØLLER 1987). Furthermore, just as for male birds, mate guarding is not always possible for male humans. Male humans may be equipped with another mechanism that 'corrects' female infidelity by motivating a man to copulate as soon as possible after a partner's absence, in an apparent effort to compete for paternity in the event of rival insemination (SHACKELFORD et al., in press).

In a study modeled after BAKER and BELLIS' (1993a) research on male ejaculate adjustment as a function of the risk of female infidelity, SHACKELFORD et al. (in press) found psychological evidence suggesting a long evolutionary history of human sperm competition. The ejaculate adjustment documented by BAKER/BELLIS (1993a, 1995) would not be functional if men were not motivated to copulate with their partners sooner rather than later following the possibility of rival insemination. SHACKELFORD et al. (in press) documented a positive relationship between the proportion of time a couple has spent apart since their last copulation and, for example, men's ratings of their partners' sexual attractiveness and men's ratings of their interest in copulating with their partner.

Male humans also face the adaptive problem of anticipating a female partner's infidelity. Men are sensitive to cues to a partner's infidelity (SHACKELFORD/BUSS 1997). One cue to the likelihood of female infidelity is the female's age. Younger, reproductive age women are more likely to promote sperm competition through extra-pair copulations than are older, post-reproductive age women (BAKER/BELLIS 1995; BUSS 1994, KINSEY/POMEROY/MARTIN 1953). Men appear to have psychological mechanisms that are sensitive to this risk. Men mated to younger women copulate more frequently with their partners than do men mated to older women (BAKER/BELLIS 1995). This pattern suggests that men mated to younger women might anticipate the possibility of sperm competition and engage in frequent copulation in an attempt to remain competitive inside their partner's reproductive tract. Anticipating sperm competition also represents an adaptive problem for many male birds. For example, the more time that a female spends in the immediate vicinity of other male birds, the more likely it is that one or more of those males will inseminate her. Male dunnocks appear to be designed to assess and combat this risk by increasing the rate at which they copulate with their partner as a function of the number of rival males close to her (DAVIES 1983).

Less work has been done on the mechanisms of human female psychology that may have been forged by an evolutionary history of sperm competition. Research on birds suggests that there are genetic benefits associated with mating with multiple males and thereby promoting sperm competition. A growing body of empirical work on non-human animals, particularly many species of birds, documents positive relationships between heritable male traits, paternity, and offspring survival (KEMPENAERS/VERHEYEN/DHONDT 1997; MØLLER 1994; SHELDON et al. 1997). There may be similar benefits of multiple mating for human females. BAKER/BELLIS (1993b, 1995) hypothesized that human female orgasm is an adaptation for preferentially retaining the sperm of certain males. BAKER/BELLIS (1993b, 1995) documented that women are more likely to report a 'high retention orgasm' with an extra-pair sexual partner than with their regular, in-pair partner. These researchers also found that, relative to the timing of sexual intercourse with their in-pair partner, women report a greater likelihood of sex with an extra-pair partner during the most fertile phase of their menstrual cycle (BELIS/BAKER 1990, BAKER/BELLIS 1995). These results suggest an active female role in sperm competition, and they are consistent with the argument that female infidelity was the primary context for the evolution of sperm competition mechanisms. THORNHILL/GANGESTAD/COMER (1995) and SHACKELFORD et al. (2000) provide evidence that the attractiveness of a woman's partner is a strong predictor of her orgasm. If male attractiveness is an honest signal of genetic quality, and if female orgasm is an adaptation for retaining preferentially the sperm of favored males, then ancestral women would have benefited by experiencing orgasm with physically attractive males.

Future work from an evolutionary psychological perspective continues to explore the hypothesis that human and non-human minds are equipped with psychological mechanisms designed to solve the problems of sperm competition. One important research area that is starting to be examined in humans is the link between sperm competition and violence perpetrated by a male against his female partner. For example, domestic abuse and marital rape might be predictable from men's perceptions that their partners are currently or have recently been sexually unfaithful, thus placing them at risk of sperm competition (SHACKELFORD/LEBLANC 2001; THORNHILL/PALMER 2000).

Summary and Conclusion

Sperm competition presents individuals of a variety of species with similar adaptive problems. Over the evolutionary history of these species, sperm competition was an integral component of sexual selection. We reviewed the development of Sperm Competition Theory and highlighted some of the evolved mechanisms in insects that were identified in early sperm competition research. This initial work on insects paved the way for a tremendous amount of empirical research on sperm competition in birds. From a comparative evolutionary psychological approach, it is predicted that the similarities in the adaptive problems recurrently faced by socially monogamous birds and humans sometimes led to similarities in the evolved psychologies of these species. Converging research suggests that this is the case. Human females and the females of socially monogamous birds may benefit by mating with multiple males and thereby promoting sperm competition. Socially monogamous male birds and male humans faced similar adaptive problems of preventing, correcting, and anticipating their female partners' infidelities. Analogous evolved solutions to these problems are evidenced by similar classes and patterns of male mate guarding behaviors in birds and in humans.

Early applications of a comparative evolutionary perspective were successful in identifying and explaining the similarities and differences in human and non-human mate selection and parenting behaviors (TRIVERS 1972). A review of the sperm competition literature from a comparative evolutionary perspective suggests the possibility of a similarly successful cross-species analysis and integration. The adaptive problems of sperm competition and their evolved solutions in non-human species, notably insects and birds, provide a model for better understanding human sexual behaviors and psychology.

Authors' address

Todd K. Shackelford and Gregory J. LeBlanc, Florida Atlantic University, Division of Psychology, 2912 College Avenue, Davie, FL 33314, USA. Phone: 954-236-1179; Fax: 954-236-1099. Email: tshackel@fau.edu

Acknowledgements

The authors thank David BUSS, Dave BJORKLUND, Martin DALY, Erika HOFF, Rick MICHALSKI, Viviana WEEKES-SHACKELFORD, and Margo WILSON for comments and suggestions that improved this article.

References

Alatalo, R. V./Gottlander, K./Lundberg, A. (1987) Extra-pair copulations and mate guarding in the polyterritorial pied flycatcher, Ficedula hypoleuca. Behavior 101:139–155.

Ashcroft, R. E. (1976) A function of the pairbond in the common eider. Wildfowl 27:101–105.

Athanasiou. R. (1973) A review of public attitudes on sexual issues. In: Zubin, J./Money, J. (eds) Contemporary Sexual Behavior. Johns Hopkins University Press: Baltimore, pp. 361–390.

Baker, R. R./Bellis, M. A. (1988) 'Kamikaze' sperm in mammals? Animal Behaviour 36:937–980.

Baker, R. R./Bellis, M. A. (1989) Number of sperm in human ejaculates varies in accordance with sperm competition theory. Animal Behaviour 37:867–869.

Baker, R. R./Bellis, M. A. (1990) Do females promote sperm competition? Data for humans. Animal Behaviour 40: 997–999.

Baker, R. R./Bellis, M. A. (1993a) Human sperm competition: Ejaculate adjustment by males and the function of masturbation. Animal Behaviour 46:861–885.

Baker, R. R./Bellis, M. A. (1993b) Human sperm competition: Ejaculate manipulation by females and a function for the female orgasm. Animal Behaviour 46:887–909.

Baker, R. R./Bellis, M. A. (1995) Human sperm competition. Chapman & Hall: London.

Birkhead, T. R./Lessels C. M. (1988) Copulation behavior of the osprey Pandion haliaetus. Animal Behaviour 36:1672–1682.

Birkhead, T. R./Møller, A. P. (1992) Sperm competition in birds. Academic Press: London.

Birkhead, T. R./Møller, A. P. (eds) (1998) Sperm competition and sexual selection. Academic Press: San Diego.

Broude, G. E./Greene, S. J. (1976) Cross-cultural codes on twenty sexual attitudes and practices. Ethnology 15:410–429.

Brownmiller, S. (1975) Against our will. Simon & Schuster: New York.

Buss, D. M. (1988) From vigilance to violence: Tactics of mate retention in American undergraduates. Ethology and Sociobiology 9:291–317.

Buss, D. M. (1994) The evolution of desire. Basic Books: New York.

Buss, D. M. (1999) Evolutionary psychology. Allyn & Bacon: Needham Heights MA.

Buss, D. M. (2000) The dangerous passion. The Free Press: New York.

Buss, D. M./Shackelford, T. K. (1997) From vigilance to violence: Mate retention tactics in married couples. Journal of Personality and Social Psychology 72:346–361.

Daly, M./Wilson, M. (1988) Homicide. Aldine de Gruyter: Hawthorne NY.

Daly, M./Wilson, M. (1995) Evolutionary psychology: Adaptationist, selectionist, and comparative. Psychological Inquiry 6:34–37.

Daly, M./Wilson, M. (1998) Human evolutionary psychology and animal behaviour. Animal Behaviour 57:509–519.

Darwin, C. (1871) The descent of man, and selection in relation to sex. John Murray: London.

Davies, N. B. (1983) Polyandry, cloaca-pecking and sperm competition in dunnocks. Nature 302:334–336.

Drummond, B. A. (1984) Multiple mating and sperm competition in the lepidoptera. In: Smith, R. L. (ed) Sperm competition and the evolution of animal mating systems. Academic Press: San Diego, pp. 547–572.

Eberhard, W. G. (1985) Sexual selection and animal genitalia. Harvard University Press: Cambridge MA.

Flinn, M. V. (1988) Mate guarding in a Caribbean village. Ethology and Sociobiology 9:1–28.

Gage, M. J. G. (1991) Risk of sperm competition directly affects ejaculate size in the Mediterranean fruit fly. Animal Behaviour 42:1036–1037.

Gage, M. J. G. (1992) Removal of rival sperm during copulation in a beetle, Tenebrio molitor. Animal Behaviour 44:587–589.

Gage, M. J. G./Baker, R. R. (1991) Ejaculate size varies with socio-sexual situation in an insect. Ecological Entomology 16:331–337.

Gage, M. J. G./Barnard, C. J. (1996) Male crickets increase sperm number in relation to competition and female size. Behavioral Ecology and Sociobiology 38:349–353.

Gomendio, M./Harcourt, A. H./Roldán, E. R. S. (1998) Sperm competition in mammals. In: Birkhead, T. R./Møller, A. P. (eds) Sperm competition and sexual selection. Academic Press: San Diego, pp. 667–755.

Green, R. (1980) Variant forms of human sexual behavior. In: Austin, C. R./Short, R. V. (eds) Reproduction in mammals, book 8, human sexuality. Cambridge University Press: New York, pp. 68–97.

Greiling, H./Buss, D. M. (2000) Women's sexual strategies: The hidden dimension of extra-pair mating. Personality and Individual Differences 28:929–963.

Harcourt, A. H./Harvey, P. H./Larson, S. G./Short, R. V. (1981) Testes weight, body weight, and breeding systems in primates. Nature 293:55–57.

Kempenaers, B./Verheyen, G. R./Dhondt, A. A. (1997) Extra-pair paternity in the blue tit (Parus caeruleus): Female mate choice, male characteristics, and offspring performance. Behavioral Ecology 8:481–492.

Kinsey, A. C./Pomeroy, W. B./Martin, C. E. (1953) Sexual behavior of the human female. Saunders: Philadelphia.

Møller, A. P. (1987) Copulation behavior in the goshawk Accipiter gentilis. Animal Behaviour 35:755–763.

Møller, A. P. (1994) Male ornament size as a reliable cue to enhanced offspring viability in the barn swallow. Proceedings of the National Academy of Science USA 91:6929–6932.

Møller, A. P. (1998) Sperm competition and sexual selection. In: Birkhead, T. R./Møller, A. P. (eds) Sperm competition and sexual selection. Academic Press: San Diego, pp. 55–90.

Ono, T./Siva-Jothy, M. T./Kato, A. (1989) Removal and subsequent ingestion of rival's semen during copulation in a tree cricket. Physiology and Entomology 14:195–202.

Parker, G. A. (1970a) Sperm competition and its evolutionary consequences in the insects. Biological Review 45:525–567.

Parker, G. A. (1970b) Sexual selection and its evolutionary effect on copula duration in the fly, Scatophaga stercoraria. Journal of Insect Physiology 16:1301–1328.

Parker, G. A. (1982) Why are there so many tiny sperm? Sperm competition and the maintenance of two sexes. Journal of Theoretical Biology 96:281–294.

Parker, G. A. (1984) Sperm competition and the evolution of animal mating strategies. In: Smith, R. L. (ed) Sperm competition and the evolution of animal mating systems. Academic Press: San Diego, pp. 2–60.

Parker, G. A. (1990a) Sperm competition games: Raffles and roles. Proceedings of the Royal Society of London B 242:120–126.

Parker, G. A. (1990b) Sperm competition games: Sneaks and extra-pair copulations. Proceedings of the Royal Society of London B 242:127–133.

Parker, G. A. (1993) Sperm competition games: Sperm size and sperm number under adult control. Proceedings of the Royal Society of London B 253:245–254.

Parker, G. A. (1998) Sperm competition and the evolution of ejaculates: Towards a theory base. In: Birkhead, T. R./ Møller, A. P. (eds) Sperm competition and sexual selection. Academic Press: San Diego, pp. 3–54.

Parker, G. A./Ball, M. A./Stockley, P./Gage, M. J. G. (1996) Sperm competition games: Individual assessment of sperm competition intensity by group spawners. Proceedings of the Royal Society of London B 263:1291–1297.

Parker, G. A./Begon, M. (1993) Sperm competition games: Sperm size and number under gametic control. Proceedings of the Royal Society of London B 253:255–262.

Parker, G. A./Simmons, L. W./Kirk, H. (1990) Analyzing sperm competition data: Simple models for predicting mechanisms. Behavioral Ecology and Sociobiology 27:55–65.

Shackelford, T. K./Buss, D. M. (1997) Cues to infidelity. Personality and Social Psychology Bulletin 23:1034–1045.

Shackelford, T. K./Buss, D. M./Peters, J. (2000) Wife killing: Risk to women as a function of age. Violence and Victims 15:273–282.

Shackelford, T. K./LeBlanc, G. J. (2001) Preventing, correcting, and anticipating female infidelity: Three adaptive problems of sperm competition. Manuscript under editorial review.

Shackelford, T. K./LeBlanc, G. J./Weekes-Shackelford, V. A./ Bleske, A. L./Euler, H. A./Hoier, S. (in press) Psychological adaptation to human sperm competition. Evolution and Human Behavior.

Shackelford, T. K./Weekes-Shackelford, V. A./LeBlanc, G. J./ Bleske, A. L./Euler, H. A./Hoier, S. (2000) Female coital orgasm and male attractiveness. Human Nature 11:299–306.

Sheldon, B. C./Merilä, J. Qvarnström, A./Gustafson, L./ Ellegren, H. (1997) Paternal genetic contribution to offspring condition predicted by the size of male secondary sexual character. Proceedings of the Royal Society of London B 264:297–302.

Short, R. V. (1979) Sexual selection and its component parts, somatic and genital selection, as illustrated by man and the great apes. Advances in the Study of Behavior 9:131–158.

Simmons, L. W./Siva-Jothy, M. T. (1998) Sperm competition in insects: Mechanisms and the potential for selection. In: Birkhead, T. R./Møller, A. P. (eds) Sperm competition and sexual selection. Academic Press: San Diego, pp. 341–434.

Smith, R. L. (1984) Human sperm competition. In: Smith, R. L. (ed) Sperm competition and the evolution of animal mating systems. Academic Press: San Diego, pp. 601–660.

Smith, R. L. (1998) Foreword. In: Birkhead, T. R./Møller, A. P. (eds) Sperm competition and sexual selection. Academic Press: San Diego, pp. xv–xxiii.

Tasker, C. R./Mills, J. A. (1981) A functional analysis of courtship-feeding in the red-billed gull (Larus novaehollandiae). Behaviour 77:222–241.

Thornhill, R. (1976) Sexual selection and nuptial feeding behavior in Bittacus apocalis. American Naturalist 110:529–548.

Thornhill, R./Alcock, J. (1983) The evolution of insect mating systems. Harvard University Press: Cambridge MA.

Thornhill, R./Gangestad, S. W./Comer, R. (1995) Human female orgasm and mate fluctuating asymmetry. Animal Behaviour 50:1601–1615.

Thornhill, R./Palmer, C. T. (2000) A natural history of rape. MIT Press: Cambridge MA.

Thornhill, R./Thornhill, N. W. (1992) The evolutionary psychology of men's coercive sexuality. Behavioral & Brain Sciences 15:363–421.

Trivers, R. L. (1972) Parental investment and sexual selection. In: Campbell, B. (ed) Sexual selection and the descent of man 1871–1971. Aldine: Chicago, pp. 136–179.

Tsubaki, Y./Siva-Jothy, M. T./Ono, T. (1994) Re-copulation and post-copulatory mate guarding increase immediate female reproductive output in the dragonfly, Nannophya pygmaea. Behavioral Ecology and Sociobiology 35:219–225.

Waage, J. K. (1979) Dual function of the damselfly penis: Sperm removal and transfer. Science 203:916–918.

Wilson, M. (1997) Our roots and our colleagues. Human Behavior and Evolution Society Newsletter 6:1–2.

Wrangham, R./Peterson, D. (1996) Demonic males. Houghton Mifflin: Boston.

Zusammenfassungen der Artikel in deutscher Sprache

R. Plutchik

Integration, Differenzierung und Abwandlung von Emotionen

Innerhalb der fast unüberschaubaren Vielfalt von Konzepten im Bereich der Emotionsforschung soll ein integrativer Ansatz folgende Inhalte aufweisen: adaptive Implikationen von Emotionen, zugrundeliegende Gehirnmechanismen, Quellen und Komplexität von emotionalen Zuständen, die Rolle von Kognitionen.

Als zentral erweist sich dabei die Annahme einer evolutionären Kontinuität von Strukturen, Funktionen und der Entwicklung von Emotionen. Darauf basiert die psychoevolutionäre Emotionstheorie, die drei Modelle beinhaltet: ein strukturelles Modell, ein Sequenzmodell und ein Abwandlungsmodell.

Kernpunkt des Strukturmodells ist die Annahme einer begrenzten Anzahl von Primäremotionen, von denen andere Emotionen abgeleitet werden können. Dies steht in enger Beziehung zu Persönlichkeitsmodellen, wobei die Annahme zugrunde liegt, daß typische Persönlichkeitseigenschaften als Ausprägungsformen bestimmter Emotionen beschrieben werden können.

Das Sequenzmodell sieht Emotionen als Teile komplexer zirkulär organisierter Feedbackprozesse, wobei an erster Stelle ein kognitiver Akt bzw. eine kognitive Interpretation steht. Dabei stehen jedoch kognitive Prozesse im Dienste basaler emotionaler, lebenserhaltender Antriebe.

Jesse M. Bering

Sind Schimpansen „reine" Existentialisten?

Eine phylogenetische Annäherung an die Ursprünge von Religion

Die evolutionäre Entwicklung der Fähigkeit zur Bildung mentaler Repräsentationen ist einer der zentralen Inhalte von Forschungen, welche die biologische Basis von Religion untersuchen. In diesem Artikel wird Religion nicht im soziologischen Sinne als kulturelles Konstrukt verwendet, sondern in einem biologischen Sinn. Zentraler Gehalt dabei sind die „nicht natürlichen" Eigenschaften von Elementen bzw. Inhalten religiöser Glaubenssysteme, die allen natürlichen ontologischen Kategorien widersprechen.

Vergleichende psychologische Untersuchungen an Schimpansen und Menschen wurden in vielen Bereichen durchgeführt (Werkzeuggebrauch, Kooperation, Sprache, ...). Was dabei jedoch nie untersucht wurde sind die Ursprünge religiösen Verhaltens mit den dazugehörigen kognitiven Voraussetzungen. Dabei lautet die zentrale Fragestellung nicht ob Schimpansen religiösen Glauben haben, sondern ob die kognitive Architektur von Schimpansen es ihnen erlaubt religiöse Repräsentationen auszubilden.

Als wesentliche kognitive Voraussetzung für die Ausbildung religiösen Verhaltens gelten höhere Formen von Intentionalität welche mit unterschiedlichen Formen von Bedeutungsgebung verbunden sind. Menschliches Verhalten ist dabei dadurch ausgezeichnet, daß die Zuschreibung mentaler Zustände sowohl hinsichtlich des Verhaltens anderer Organismen, wie auch hinsichtlich externer, natürlicher Ereignisse vorgenommen werden kann. Bezüglich der kognitiven Kapazitäten von Schimpansen legt der gegenwärtige Forschungsstand die Annahme nahe, daß diese über eine begrenzte „theory of mind" verfügen. Dabei erweist sich die Unterscheidung zwischen der Fähigkeit bestimmten Ereignissen intentionalen Charakter zuzuschreiben von derjenigen Handlungen intentionale Eigenschaften zuzugestehen als zentral.

Jaak Panksepp

Die neuro-evolutionäre Schleife zwischen Emotion und Kognition

Implikationen zum Verständnis von Bewußtsein und zur Entstehung einer vereinheitlichenden Wissenschaft des Geistes

Ein Integration von Perspektiven der Verhaltenswissenschaften, der Psychologie und den Neurowissenschaften würde zur Etablierung einer „affektiven Neurowissenschaft" beitragen, welche eine zentrale Rolle in wissenschaftlichen und auch gesellschaftlichen Belangen einnehmen kann.

Die Strategien dieser affektiven Neurowissenschaft können dabei zu einem umfassenden Verständnis des menschlichen Geistes beitragen und dabei vor allem auch die subjektive Erfahrungsdimension emotionaler Prozesse berücksichtigen.

Als Voraussetzung gilt dabei die Kenntnis „tiefliegender" affektiver Funktionen, die als Substrat all höherer Gehirn–Geist Funktionen gelten. Dieses „emotionale Gehirn" teilen wir mit allen Säugetieren und dieser Gehirnteil beeinflußt in massiver Art alle darüberliegenden Schichten. Der Neocortex sollte daher besser nicht als Quelle des Bewußtseins konzeptualisiert werden, sondern eher als generalisiertes Multifunktionsorgan welches an die Funktionen der subkortikalen Schichten angepaßt ist.

Matthias Scheutz

Ethologie und Funktionalismus

Der Begriff des „behavioral state" soll als eine Brücke zwischen der funktionellen und der physischen Beschreibungsebene kognitiver Systeme herangezogen werden. Die Beschreibungsebene des „behavioral state" stammt wesentlich aus dem Bereich der Ethologie und bezieht sich dabei vor allem auf ein bestimmtes Aktivierungsniveau und damit verbundene Verhaltensmuster. Diese können sowohl direkt beobachtbar, wie auch „intern" vollzogen werden – wie dies etwa bei Denk- oder Gedächtnisprozessen der Fall ist.

Hinsichtlich der Beziehungen zwischen dem physischen Substrat und dem „behavioral state" (der Implementierung) zeigen sich unterschiedliche Formen, die von der direkten bis zu einer partieller Implementierung reichen.

„Behavioral states" eröffnen dabei, im Gegensatz zu rein funktionellen Beschreibungen, die Möglichkeit die zeitliche Dimension von Interaktionen innerhalb des kognitiven Systems selbst wie auch bei Interaktionen zwischen dem kognitiven System und seinen Umgebungsbedingungen besser darzustellen.

G.C. Geary/M.C. DeSoto

Geschlechtsunterschiede im räumlichen Vorstellungsvermögen zwischen Erwachsenen aus den USA und China

Es handelt sich bei dieser Arbeit um eine empirische Studie welche das räumliche Vorstellungsvermögen bei männlichen und weiblichen Versuchspersonen vergleicht. Getestet wurde diese Fähigkeit mithilfe des mentalen Rotationstests – ein Testverfahren, bei dem die Probanden geometrische Figuren in drei Dimensionen vorstellungsmäßig rotieren lassen.

Es zeigt sich dabei, daß die männlichen Probanden unterschiedlicher Kulturen durchwegs bessere Ergebnisse erzielten.

Die vorliegende Studie versucht dabei den Nachweis zu erbringen, daß diese Differenzen auf einer biologischen Basis beruhen und nicht auf unterschiedliche, geschlechtsspezifische Sozialisierungsformen zurückzuführen sind.

Theresa Schilhab

Mensch und Spiegel

Die anthropozentrische Perspektive bei Spiegelversuchen zum tierischen Selbsterkennen

Das wachsende Interesse an der Frage nach einer „theory of mind" bei nichtmenschlichen Primaten führte zu einer beträchtliche Anzahl von Untersuchungen in diesem Bereich. Dabei geht die „theory of mind" Hypothese davon aus, daß auch nichtmenschliche Primaten ihren Artgenossen und sich selbst innere, mentale Zustände zuschreiben. Die, ausgehend von diesen Fragestellungen durchgeführten empirischen Untersuchungen erbrachten eine Vielzahl von Ergebnissen, die jedoch völlig zu unterschiedlichen Interpretationen führten.

Eine der Absichten dieses Artikels besteht im Aufzeigen der Schwierigkeiten eindeutige Hinweise für die Fähigkeit der Zuschreibung mentaler Zustände bei nichtmenschlichen Primaten zu finden. Vergleichbare Untersuchungen bei Menschenkindern beziehen sich dabei vor allem auf verbale Rückmeldungen, die bei der Untersuchung nichtmenschlicher Primaten nicht in Betracht gezogen werden können. Um anthropomorphistische Fehlschlüsse zu vermeiden wird eine phylogenetische Perspektive als mögliches Korrektiv dargestellt.

Scott Atran

Ein Modul zum Erkennen von Betrügern?

Fragwürdige Interpretationen der Wasonschen Wahlaufgaben

Versuchspersonen haben normalerweise beträchtlich Probleme bei den WASONSCHEN Wahlaufgaben. Es geht dabei um die Prüfung hypothetischer Implikationen auf ihre Korrektheit in der Art „Wenn P, dann Q".

Vor ähnliche Wahlaufgaben finden sich vielfach Tiere (u.a. Vampirfledermäuse), die auf Kooperation angewiesen sind. Normalerweise würde eine Gruppe sozialer Fledermäuse das Eindringen von Betrügern nicht überdauern, wenn diese nicht über die Fähigkeit verfügen würden, Betrüger zu erkennen. Daher setzt jede Form von Kooperation ein kognitives System voraus, welches seinen Träger darüber informiert, ob ein Betrug vorliegt oder nicht.

Werden die WASONSCHEN Wahlaufgaben den Versuchspersonen in konkreter Form sozial relevanter Fragen präsentiert und damit dem Vorwissen der Befragten entsprechend gestaltet, so sind die Ergebnisse wesentlich besser als bei abstrakt formulierten Fragestellungen.

Derartige Ergebnisse gaben zu der Vermutung Anlaß, daß diesem Verhalten ein bereichsspezifisches, evolutionär angepaßtes Modul zugrunde liegt. Als Gegenkonzept dazu steht die plausiblere Annahme universeller, bereichsübergreifender Formen kognitiver Prozesse, die jeweils in Zusammenarbeit mit kontextspezifischen Mechanismen operieren.

T. K. Shackelford/ G. LeBlanc

Spermienkonkurrenz bei Insekten, Vögeln und Menschen

Einsichten aus einer evolutionären Perspektive

Im Rahmen eines vergleichend psychologisch–evolutionären Ansatzes soll die Spermienkonkurrenz bei Vögeln, Insekten und Menschen analysiert werden. Spermienkonkurrenz tritt dann auf, wenn sich Spermien von zwei oder mehr Männchen im Reproduktionstrakt des Weibchens befinden. Dieser stellt damit eine wesentliche Komponente sexueller Selektion dar. Ein vergleichend evolutionär–psychologischer Ansatz vertritt dabei die These, daß sich bei Organismen mit ähnlichen adaptiven Problemen im Bereich der Reproduktion auch ähnliche Strategien hinsichtlich Partnerwahl und Sexualverhalten ausbilden. Dahingehend wird von der Annahme ausgegangen, daß aus den in der Evolution von Vögeln und Insekten entstandenen Strategien hinsichtlich Spermienkonkurrenz ein Modell abgeleitet werden kann, welches dazu beiträgt menschliches Sexualverhalten und damit verbundene psychische Prozesse besser zu verstehen.